

Université de Montréal

Comprendre la fonction cellulaire par l'analyse du transcriptome : une approche bio-informatique

par

Saraï Mola

Département de Biochimie et Médecine moléculaire
Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Bio-informatique

Mai 2021

© Saraï Mola, 2021

Cette thèse intitulée

Comprendre la fonction cellulaire par l'analyse du transcriptome : une approche bio-informatique

Présentée par
Saraï Mola

A été évaluée par un jury composé des personnes suivantes

Sylvie Mader
Président-rapporteur

John Rioux
Directeur de recherche

François Major
Codirecteur

Adrian Serohijos
Membre du jury

Jean-Baptiste Rivière
Examineur externe

RÉSUMÉ

Le système immunitaire permet la défense de l'organisme contre l'intrusion de nombreux pathogènes (ex : microbes, champignons, virus et bactéries). Il est composé de plusieurs populations cellulaires, provenant d'une même cellule souche multipotente, et pouvant identifier la présence de ces pathogènes. La reconnaissance de ces derniers induit deux types de réponses immunitaires afin de les éliminer : les réponses immunitaires innée et adaptative. Ces réponses entraînent le recrutement de nombreuses cellules immunitaires circulantes sur le lieu de détection du pathogène, appelé site d'inflammation. Chaque population de cellule immunitaire présente sur le site, va jouer un rôle spécifique, déterminé par sa fonction cellulaire.

Un individu est défini par son génome composé de gènes hérités et spécifiquement exprimés à travers les populations cellulaires de l'organisme. L'étude de l'expression génique à travers plusieurs tissus humains a montré qu'un sous-groupe de moins de 200 gènes était exclusivement exprimé dans un tissu donné et qu'environ 2 000 gènes sont indispensables pour qu'une cellule humaine se développe et prolifère. De plus, certains gènes ont la capacité d'exprimer plusieurs transcrits qui vont encoder différents isoformes d'une protéine dont certaines ont une fonction qui leur est spécifique.

L'étude du transcriptome dans plusieurs types cellulaires permettrait de mieux comprendre les fonctions spécifiques d'une cellule. Le développement d'approche pour l'étude de l'expression génique dans les cellules immunitaires primaires humaines offrirait un niveau supplémentaire d'information sur l'établissement et le maintien de l'expression des gènes et des transcrits encodant différentes isoformes protéiques, qui sont en partie responsables de la fonction cellulaire. L'identification des éléments et mécanismes régulant l'expression des gènes et de leurs transcrits compléterait l'approche.

La première partie de cette thèse présente une étude de la co-expression génique dans les cellules immunitaires primaires pour mieux comprendre la fonction cellulaire. L'approche développée identifie tout d'abord des modules fonctionnels de gènes. Nos modules de gènes co-exprimés sont obtenus avec l'outil WGNA puis étudiés par trois analyses, une globale d'enrichissement d'annotations géniques et deux ciblées, utilisant respectivement un classement par percentile et une analyse de la première composante principale, pour sélectionner un sous-groupe de gènes d'intérêt. L'approche développée permet aussi l'étude de la régulation de l'expression génique, par les facteurs de transcription au sein des modules de gènes co-exprimés, et dans les décisions de différenciation dans le processus physiologique de production des cellules immunitaires, l'hématopoïèse. Cette approche a permis de démontrer qu'une grande partie des gènes co-exprimés, joue un rôle important dans les fonctions du type cellulaire, auquel ils sont associés. De plus, certains gènes co-exprimés sont impliqués dans la même fonction cellulaire. Pour les gènes dont la fonction et le rôle ne sont pas connus, dans le type cellulaire auquel ils sont associés, nous présumons que leur rôle est important dans ce type cellulaire. Aussi, ces gènes pourraient jouer un rôle dans les fonctions mises en évidence dans leur module.

La seconde partie de la thèse s'intéresse à l'expression par transcrits et aux isoformes de protéines qu'ils encodent afin d'identifier une différence dans le profil d'expression et de la fonction entre les isoformes protéiques provenant d'un même gène. La méthodologie utilisée pour cette approche compare les profils d'expression des transcrits d'un même gène, et identifie ceux qui ont un profil d'expression distinct dans nos types cellulaires et codant pour des isoformes de protéines spécifiques. L'annotation des transcrits a permis de définir quels événements transcriptomiques (ex : sites d'initiation de la transcription distincts et/ou épissage alternatif) sont impliqués dans le choix de l'isoforme protéique exprimée au sein de la cellule. Nous avons obtenu 15 gènes qui ont au moins un transcrit avec un profil d'expression spécifique dans nos données. Les gènes *PTPRC*, qui code pour la protéine CD45, et le gène *CD44* ont été révélés par notre approche et sont connus pour avoir des isoformes de protéines avec des affinités différentes à leur ligand et ces isoformes sont spécifiquement exprimées dans les cellules immunitaires. Cette approche nous a permis d'identifier des isoformes de transcrits spécifiquement exprimés dans nos cellules immunitaires et pouvant avoir une fonction différente et importante dans ces types cellulaires.

Dans cette thèse, nous présentons deux approches d'analyse du transcriptome humain dans les cellules immunitaires primaires. Celles-ci apportent un plus grand niveau de compréhension sur l'expression par gènes et par transcrit, et sa régulation, ainsi que sur l'établissement et le maintien des fonctions cellulaires spécifiques et partagées dans les cellules immunitaires primaires. Ces approches pourront être appliquées dans d'autres conditions de comparaison telles que sain versus malade ou encore traité versus non-traité.

Mots-clés : Séquençage d'ARN; Cellules immunitaires primaires humaines; Modules de gènes co-exprimés; Profils expression différentielle par transcrit; Isoformes protéiques.

ABSTRACT

The human immune system protects the body against the intrusion of many pathogens (bacteria, viruses, fungi, and parasites). It is composed of several cell populations, originating from the same multipotent stem cell, and able to identify the presence of these pathogens. Recognition of these induces two types of immune responses to eliminate them: innate and adaptive immune responses. These responses lead to the recruitment of numerous circulating immune cells to the site of detection of the pathogen, called the site of inflammation. Each immune cell population present on the site will play a specific role, determined by its cellular function.

An individual is defined by his genome composed of genes inherited and specifically expressed through the cell populations of the organism. The study of gene expression across several human tissues showed that a subset of less than 200 genes were exclusively expressed in each tissue and that approximately 2,000 genes are essential for a human cell to develop and proliferates. In addition, some genes can express several transcripts that will encode different isoforms of a protein, some of which have a specific function.

Genome-wide transcriptomic-based analysis in several cell types could give a better understanding to specific cell functions. To increase comprehension of establishment and maintaining gene and transcript expression implicated in cell function, an approach would be developed to study transcriptome of primary human immune cell types. It will be completed with the identification of elements and mechanisms involved in genes and transcript expression in immune cells.

The first part of the these presents a study of gene co-expression in primary immune cells to better understand cell function. This approach identifies functional gene modules. Our co-expressed gene modules are obtained with the WGNA tool and the modules were studied to select a subgroup of genes of interest by: (1) one global analysis of gene annotation enrichment and (2) two targeted analyses, using percentile ranking and first principal component analysis respectively. The approach also evaluates gene expression regulation by transcription factors in co-expressed gene modules and in fate decision of hematopoiesis. This approach demonstrates that a large part of the co-expressed genes play an important role in the functions of the associated cell type. Moreover, some co-expressed genes are involved in the same cellular function. For genes whose function and role are not known, in the associated cell type, we assume that their role is important in this cell type by guilt by association. Also, these genes could play a role in the functions highlighted in their module.

The second part of the thesis focuses on the expression by transcripts and the protein isoforms they encode to identify a difference in the expression profiles and function between protein isoforms come to the same gene. The methodology used for this approach compares the expression profiles of transcripts from the same gene and identifies those with a distinct expression profile in our cell types and encode specific protein isoforms. Transcript annotation

was used to define which transcriptomic events (e.g., distinct transcription initiation sites and/or alternative splicing) are involved in the choice of the protein isoform expressed within the cell. We obtained 15 genes that have at least one transcript with a specific expression profile in our data. The PTPRC genes, which codes for the CD45 protein, and the CD44 gene were revealed by our approach and are known to have protein isoforms with different affinities to their ligand. Their protein isoforms are specifically expressed in immune cells. This approach allowed us to identify isoforms of transcripts specifically expressed in our immune cells and which may have a different and important function in these cell types.

In this thesis, we present two approaches for analyzing the human transcriptome in primary immune cells. These provide a greater level of understanding on gene and transcript expression and its regulation, as well as on the establishment and maintenance of specific and shared cellular functions in primary immune cells. These approaches can be applied in other comparison conditions such as between patients and healthy individuals to identify pathogenic pathways or between treated versus untreated individuals to identify biomarkers.

Keywords: RNA-Seq; Human primary immune cells; Co-expressed gene modules; Differential expression profiles per transcript; Protein Isoforms.

Table des matières

RÉSUMÉ	i
ABSTRACT	iii
Table des matières	v
Liste des tableaux	viii
Liste des figures	ix
LISTE DES SIGLES ET DES ABRÉVIATIONS	xi
REMERCIEMENTS	xvii
1 Chapitre 1 – Introduction	1
1.1 Le génome humain	1
1.1.1 Rôle et fonction du transcriptome.....	1
1.1.2 Mécanisme et éléments régulateurs de la transcription.....	2
1.1.3 Encodage des protéines et expression de multiples isoformes.....	8
1.2 Le système immunitaire	12
1.2.1 Les cellules immunitaires primaires.....	14
1.2.2 La communication intracellulaire : Cytokines et chimiokines.....	15
1.2.3 Identification immunophénotypique : Classe de différenciation.....	24
1.2.4 Fonctions partagées entre les cellules immunitaires primaires : Voies de signalisation et FT	27
1.2.5 Fonctions spécifiques à un type cellulaire : Voies de signalisation et FT.....	35
1.3 Étude du transcriptome humain	44
1.3.1 Analyse de l'expression génique: Outils et méthodes.....	44
1.3.2 Construction d'un réseau de co-expression génique : Outil WGCNA.....	48
1.3.3 Analyse de l'expression génique et sa régulation dans les cellules immunitaires : Projet ImmGen	51
1.3.4 Analyse de la régulation de l'expression génique : Projet ENCODE.....	53
1.4 Mise en contexte et introduction des articles	57
2 Chapitre 2 – A transcriptome-based approach to identify functional modules within and across primary human immune cells	58
2.1 Contribution des auteurs	58
2.2 Article	59
2.3 Abstract	59
2.4 Introduction	60
2.5 Materials and methods	61
2.5.1 Subject selection criteria, ethics committee approval, and blood sample procedures.....	61
2.5.2 Blood collection and cellular isolation.....	61
2.5.3 Macrophage differentiation and activation.....	62

2.5.4	Immunophenotyping	62
2.5.5	RNA extraction and sequencing.....	62
2.5.6	Bioinformatic processing of sequence files	63
2.5.7	Principal component and hierarchical clustering analyses	64
2.5.8	Definition of co-expression modules	65
2.5.9	Global and targeted functional annotation.....	66
2.5.10	Analyses of transcription factors and transcription factor binding sites.....	66
2.5.11	Transcription factors potentially associated with cell lineages and states.....	67
2.6	Results	67
2.6.1	Data generation: isolation of immune cells and RNA-Seq analysis	67
2.6.2	Structure of transcriptomic data is a reflection of the known differentiation scheme for immune cell subsets	68
2.6.3	Defining potential functional groups using a combination of gene co-expression and functional annotation	68
2.6.4	B cell-specific modules; cell activation and BCR engagements	73
2.6.5	A module shared by B cells and monocytes; a gene expression program enabling antigen processing and presentation.....	75
2.6.6	A shared module associated with NK, $\gamma\delta$ T, and CD8+ T cells; a role in cytotoxic granule and non-self-recognition	75
2.6.7	Modules associated with LPS-activated macrophages	77
2.6.8	The neutrophil transcriptome is distinctive from other immune cell types	78
2.6.9	Transcriptional control of genes within B cell modules.....	79
2.6.10	Transcription factors potentially associated with cell lineages and states.....	80
2.7	Discussion.....	80
2.8	Funding.....	82
2.9	Acknowledgments	83
2.10	Competing interests.....	83
2.11	Protocol.....	83
2.12	Supporting information	83
3	<i>Chapitre 3 - Identifying transcript-level differential expression in primary human immune cells</i>	84
3.1	Contribution des auteurs	84
3.2	Article.....	85
3.3	Abstract.....	85
3.4	Introduction.....	86
3.5	Materials and methods	87
3.5.1	Description of dataset	87
3.5.2	Bioinformatic processing of sequence files.....	88
3.5.3	Principal component and hierarchical clustering analyses	88
3.5.4	Transcript annotations and gene information: protein isoform encoded, transcription factor and gene involved in splicing machinery.....	89
3.5.5	RNA isoforms associated with cell lineages and states.	89
3.6	Results.....	90

3.6.1	Alternate promoter usage and differential splicing leads to a large diversity of transcripts in human immune cells	90
3.6.2	Systematic bioinformatic screen identifies genes where two or more transcripts have distinct expression patterns in primary human leukocytes	92
3.6.3	Identifying differentially expressed isoforms of protein-coding genes in monocytes, macrophages and activated macrophages	98
3.7	Discussion	102
3.8	Funding	104
3.9	Competing interests	104
4	Chapitre 4 – Conclusion	105
4.1	Conclusion sur l'étude du transcriptome pour mieux comprendre les fonctions des cellules immunitaires	105
4.1.1	La co-expression génique et la régulation transcriptionnelle permettent de mieux comprendre la fonction cellulaire	106
4.1.2	Les isoformes offrent à la protéine des fonctions spécifiques dans certaines cellules immunitaires primaires	108
4.2	Limites et perspectives des approches d'analyses de la co-expression génique, la régulation par des facteurs de transcription et des isoformes protéiques	110
	Références	122
	<i>Annexe A. Informations supplémentaires pour l'article « A transcriptome-based approach to identify functional modules within and across primary human immune cells »</i>	<i>i</i>
	<i>Annexe B. Informations supplémentaires pour l'article « Identifying transcript-level differential expression in primary human immune cells »</i>	<i>xxi</i>

Liste des tableaux

TABLEAU 1.1.	PRINCIPALES DIFFÉRENCES ENTRE LES RÉPONSES IMMUNITAIRES INNÉES ET ADAPTATIVES	13
TABLEAU 1.2.	FAMILLES DE CYTOKINE	16
TABLEAU 1.3.	ACTIVITÉS SÉLECTIONNÉES POUR LA MAJORITÉ DES CYTOKINES	17
TABLEAU 1.4.	CHIMIOKINES	17
TABLEAU 1.5.	CHIMIOKINES (SUITE)	19
TABLEAU 1.6.	RÉCEPTEURS DES CHIMIOKINES.	20
TABLEAU 1.7.	MARQUEURS CD COMMUNÉMENT UTILISÉS DANS LA BIOLOGIE DES NEUTROPHILES	27
TABLEAU 1.8.	RÉCEPTEURS DES FRAGMENTS DU COMPLÉMENT	33
TABLEAU 1.9.	EXEMPLES D'OUTILS D'ANALYSE DE L'EXPRESSION DIFFÉRENTIELLE PAR EXONS/TRANSCRITS.	45
TABLEAU 1.10.	EXEMPLES D'ANALYSE DE L'EXPRESSION DIFFÉRENTIELLE D'EXONS/TRANSCRITS DANS PLUSIEURS TYPES CELLULAIRES.	46
TABLEAU 1.11.	RÉCAPITULATIF D'ANALYSE D'ÉPISSAGES ALTERNATIFS, D'ISOFORMES DE TRANSCRITS DIFFÉRENTIELLEMENT EXPRIMÉS DANS DIFFÉRENTS STADES CELLULAIRES ET MALADIES ASSOCIÉES AUX ERREURS D'ÉPISSAGE OU MUTATIONS.	53
TABLEAU 1.12.	PROJET ENCYCLOPÉDIE D'ENCODE - EXPÉRIENCES RÉALISÉES DANS LES CELLULES IMMUNITAIRES PRIMAIRES	56
TABLE 2.1.	SUMMARY OF KEY FUNCTIONS IDENTIFIED FOR MODULES OF CO-EXPRESSED GENES.	70
TABLE 3.1	GENES WITH AT LEAST TWO TRANSCRIPTS WITH DISTINCT EXPRESSION PATTERNS BETWEEN CELL TYPES OR CELL LINEAGES.	92
TABLE 3.2	GENES WITH AT LEAST TWO TRANSCRIPTS THAT ARE DIFFERENTIALLY EXPRESSED BETWEEN MONOCYTES, MACROPHAGES AND LPS-ACTIVATED MACROPHAGES	93
S2.1 TABLE.	LIST OF ANTIBODIES USED FOR LYMPHOCYTE/MONOCYTE IMMUNOPHENOTYPING	XIII
S2.2 TABLE.	LIST OF ANTIBODIES USED FOR MACROPHAGE IMMUNOPHENOTYPING	XIII
S2.3 TABLE.	SUMMARY STATISTICS OF RNA-SEQ DATA FROM RAW READS THROUGH QUALITY CONTROL STEPS	XIV
S2.4 TABLE.	SUMMARY OF GENE ANNOTATION ENRICHMENTS FROM DAVID TOOL	XV
S2.5 TABLE.	DIFFERENTIAL GENE EXPRESSION AND RATIOS OF HUMAN TRANSCRIPTION FACTORS	XV
S2.6 TABLE.	PERCENTILE, MEAN, STANDARD DEVIATION, MEDIAN, AND IQR OF GENE EXPRESSION READ COUNTS	XV
S2.7 TABLE.	SUMMARY OF TRANSCRIPTION FACTOR BINDING SITE OR TFBS ENRICHMENTS FROM THE ENCODE PROJECT	XV
S2.8 TABLE.	LITERATURE REVIEW OF KEY TRANSCRIPTION FACTORS INVOLVED IN B-CELL DIFFERENTIATION AND MATURATION.	XVI
S3.1. TABLE	TRANSCRIPT EXPRESSION AND ANNOTATIONS IN OUR RNA-SEQ DATA.	XXXIV
S3.2. TABLE	RESULTS OF SELECTED GENES IN LYMPHOID CELL/MONOCYTE ANALYSIS	XXXIV
S3.3. TABLE	SUMMARIES OF GENE AND TRANSCRIPT NUMBER FOR GENES WITH AT LEAST TWO TRANSCRIPTS.	XXXIV
S3.4. TABLE	RESULTS OF SELECTED GENES IN MONOCYTE/MACROPHAGE LINEAGE ANALYSIS	XXXIV

Liste des figures

FIGURE 1.1.	TYPES DE RÉGIONS RÉGULATRICES DE L'EXPRESSION GÉNIQUE	2
FIGURE 1.2.	MÉCANISME DE LA TRANSCRIPTION CHEZ LES EUCARYOTES.....	3
FIGURE 1.3.	SCHÉMA DE L'ORGANISATION D'UN NUCLÉOSOME	5
FIGURE 1.4.	EXEMPLE DE MATRICE DE POIDS	5
FIGURE 1.5.	SPLICEOSOME ET FORMATION DE L'ARNM	7
FIGURE 1.6.	CONTRÔLE ÉPIGÉNIQUE DE L'ÉPISSAGE ALTERNATIF	7
FIGURE 1.7.	SCHÉMATISATION DE L'IMPACT DU TEMPS D'ÉLONGATION SUR L'ÉPISSAGE ALTERNATIF.....	8
FIGURE 1.8.	SCHÉMATISATION DU MÉCANISME DE TRADUCTION DES PROTÉINES.	9
FIGURE 1.9.	SCHÉMATISATION DE PLUSIEURS MÉCANISMES D'ÉPISSAGE ALTERNATIF ET DES VARIANTS D'ARNM OBTENUS À PARTIR D'UN UNIQUE GÈNE	11
FIGURE 1.10.	QUATRE CATÉGORIES DES RÉGIONS SPÉCIFIQUES D'ISOFORMES PERMETTANT DES EFFETS SUS LES INTERACTIONS 11	11
FIGURE 1.11.	SCHÉMA DE DIFFÉRENCIATION DES CELLULES IMMUNITAIRES.....	13
FIGURE 1.12.	DIFFÉRENCIATION DES MONOCYTES EN SOUS-GROUPE DE POPULATION DE MACROPHAGES).	19
FIGURE 1.13.	SCHÉMATISATION DE L'EXPRESSION ET DE LA PRODUCTION DES CHIMIOKINES PAR LES NEUTROPHILES IMPACTANT LES CELLULES IMMUNITAIRES DE LA RÉPONSE INNÉE ET ADAPTATIVE	22
FIGURE 1.14.	SCHÉMA DES VOIES DE SIGNALISATION INTRACELLULAIRE DE LA TRANSDUCTION DU SIGNAL INDUIT PAR IL4 ET IL12. T	24
FIGURE 1.15.	IDENTIFICATION DES SOUS-TYPES DE CELLULES IMMUNITAIRES PAR HUIT COULEURS DE COLORATION D'ANTICORPS. 26	26
FIGURE 1.16.	VOIES DU CMH DE CLASSES I ET II	29
FIGURE 1.17.	SCHÉMA DU SYSTÈME DU COMPLÉMENT ET DES PROTÉINES RÉGULATRICES	31
FIGURE 1.19.	UN MODÈLE DESCRIPTIF DE LA BIOGÉNÈSE ET DE L'EXOCYTOSE DES GRANULES CYTOTOXIQUES.....	35
FIGURE 1.21.	VOIE DE SIGNALISATION DES CELLULES B EN RÉPONSE À UN Ag	37
FIGURE 1.22.	ANTAGONISATION DU SIGNAL DU BCR	37
FIGURE 1.23.	SIGNAL INTRACELLULAIRE INDUIT PAR LE RÉCEPTEUR BAFF	38
FIGURE 1.24.	ILLUSTRATION SCHÉMATIQUE DE LA GRANULOPOÏÈSE.....	39
FIGURE 1.25.	CASCADE CLASSIQUE DE RECRUTEMENT DES NEUTROPHILES.....	40
FIGURE 1.26.	MODÈLE DE FORMATION DES NETS	41
FIGURE 1.27.	SCHÉMATISATION DE LA PROTÉINE NOX2 ACTIVÉE À LA SURFACE DE LA MEMBRANE D'UN PHAGOSOME AYANT INGÉRÉ UNE BACTÉRIE, FUSIONNÉ À UNE GRANULE	42
FIGURE 1.28.	RÔLE ET HOMÉOSTASIE DU ZINC DANS LES FONCTIONS DES MACROPHAGES	43
FIGURE 1.29.	SYNERGIE DES RÉCEPTEURS DURANT LA PHAGOCYTOSE.....	43
FIGURE 1.30.	DÉFINITION DE LA VALEUR B EN TENANT COMPTE DE LA TOPOLOGIE DE TYPE SCALE-FREE ET LA CONNECTIVITÉ MOYENNE	49
FIGURE 1.31.	HEATMAP DE MODULES DE GÈNES SIGNIFICATIVEMENT ASSOCIÉS À DES TRAITS CLINIQUES CHEZ LES SOURIS FEMELLES ET MALES	50
FIGURE 1.32.	REPRÉSENTATION DE L'ABONDANCE RELATIVE DES EXONS DANS LES CELLULES T CD4+ ET B	52
FIGURE 1.33.	SCHÉMATISATION DES CARACTÉRISTIQUES ÉTUDIÉES PAR LE PROJET D'ANNOTATION GÉNOMIQUE ET TRANSCRIPTIONNELLE	54
FIGURE 1.34.	REPRÉSENTATION GRAPHIQUE DES ÉTUDES DU PROJET ENCYCLOPÉDIE D'ENCODE CHEZ L'HUMAIN	55
FIGURE 2.1.	TRANSCRIPTOMIC DATA DISTINGUISH IMMUNE CELL TYPES.....	64
FIGURE 2.2.	UNSUPERVISED HIERARCHICAL CLUSTERING OF RNA-SEQ DATA REFLECTS HEMATOPOIETIC DIFFERENTIATION	65
FIGURE 2.3.	HEATMAP OF THE CORRELATION VALUES OF WGCNA MODULES WITH PRIMARY IMMUNE CELL TYPES .	72
FIGURE 2.4.	BIOLOGICAL FUNCTIONS ASCERTAINED FROM THE ANALYSIS OF GENES WITHIN TWO B CELL-ASSOCIATED MODULES	74
FIGURE 2.5.	TARGETED ANALYSES OF GENES WITHIN MODULE 26, ASSOCIATED WITH NK, $\gamma\delta$ T, AND CD8+ T CELLS.	76

FIGURE 3.1	TRANSCRIPT DISTRIBUTION IN OUR DATA AND UNSUPERVISED HIERARCHICAL CLUSTERING OF TRANSCRIPT RNA-SEQ DATA	91
FIGURE 3.2	EXPRESSION PATTERN AND ENCODED PROTEIN STRUCTURES OF <i>PTPRC</i> TRANSCRIPTS ACROSS IMMUNE CELL POPULATIONS	96
FIGURE 3.3	EXPRESSION PATTERN AND ENCODED PROTEIN STRUCTURES OF <i>SEPTIN9</i> TRANSCRIPTS ACROSS IMMUNE CELL POPULATIONS	98
FIGURE 3.4	EXPRESSION PATTERN AND ENCODED PROTEIN STRUCTURES OF <i>CD44</i> TRANSCRIPTS ACROSS IMMUNE CELL POPULATIONS	100
FIGURE 3.5	EXPRESSION PATTERN AND ENCODED PROTEIN STRUCTURES OF <i>KLF6</i> TRANSCRIPTS ACROSS IMMUNE CELL POPULATIONS	101
FIGURE 4.1.	PRÉSENTATION DE TROIS MÉTHODES DE SÉQUENÇAGE D'ARN	111
FIGURE 4.2	SCHÉMATISATION DES ÉTAPES D'ANALYSE DE DONNÉES DE SÉQUENÇAGE D'ARN POUR ÉTUDIER L'EXPRESSION DIFFÉRENTIELLE DES GÈNES.	114
FIGURE 4.3	SCHÉMATISATION COMPLÈTE DES ÉTAPES DE PRÉPARATIONS DES BIBLIOTHÈQUES POUR LE SÉQUENÇAGE D'ARN PAR CELLULE UNIQUE.	116
FIGURE 4.4	CLASSIFICATION DES FONCTIONS DES LNCRNA	118
FIGURE 4.5	SCHÉMA REPRÉSENTATIF DU LIEN ENTRE LA TRADUCTION ET LES TROIS MODES DE SURVEILLANCE POUR LA DÉGRADATION DE L'ARNM	119
S2.1 FIG.	IMMUNOPHENOTYPING OF IMMUNE CELL POPULATIONS FROM HUMAN PERIPHERAL BLOOD	V
S2.2 FIG.	BOXPLOT OF GENE EXPRESSION DISTRIBUTION AFTER NORMALIZATION BEFORE AND AFTER NORMALIZATION..	VII
S2.3 FIG.	PLOT OF SCALE INDEPENDENCE AND MEAN CONNECTIVITY – PEARSON CORRELATION	VIII
S2.4 FIG.	HIERARCHICAL TREES OF GENE MODULES BEFORE AND AFTER CUT TREE	IX
S2.5 FIG.	HEMATOPOIETIC DIFFERENTIATION SCHEME AND ASSOCIATED TRANSCRIPTION FACTORS FROM DIFFERENTIAL GENE EXPRESSION	X
S2.6 FIG.	HEATMAP OF THE CORRELATION VALUES OF <i>WGCNA</i> MODULES WITH PRIMARY IMMUNE CELL TYPES	X
S2.7 FIG.	HEATMAP OF MEAN NORMALIZED EXPRESSION FOR A SUBSET OF GENES	XI
S2.8 FIG.	GLOBAL AND TARGETED ANALYSES OF GENES WITHIN MODULE 41, ASSOCIATED WITH B CELLS AND MONOCYTES, DESCRIBE MHC CLASS II AND ANTIGEN PROCESSING AND PRESENTATION FUNCTIONS	XII
S3.1. FIG	BOXPLOTS OF TRANSCRIPT EXPRESSION DISTRIBUTION BEFORE AND AFTER NORMALIZATION.	XXXV
S3.2. FIG	PC ANALYSIS OF TRANSCRIPTOMIC DATA DISTINGUISHING IMMUNE CELL TYPES	XXXVII
S3.3. FIG	UNSUPERVISED HIERARCHICAL CLUSTERING OF TRANSCRIPT RNA-SEQ DATA REFLECTS HEMATOPOIETIC DIFFERENTIATION	XXXVIII
S3.4. FIG	DISTRIBUTION OF THE HIGHEST GEOMETRIC MEAN OBSERVED ACROSS CELL TYPE	XL
S3.5. FIG	TRANSCRIPT ISOFORM DISTRIBUTION IN OUR DATA BY CELL TYPES.	XLI

LISTE DES SIGLES ET DES ABRÉVIATIONS

a.a. : Acide aminé
AAM : Alternatively activated macrophage
Ac : Anticorps
AD : Activation Domain
ADN : Acide désoxyribonucléique
Ag : Antigène
AG : Adenosine, Guanine
AG : Granule Azurophile
AML disease : Acute myelogenous leukemia
ApC : Apoptotic cell
APC : Amplification en Chaîne par Polymérase
APN : Aminopeptidase N
ARN : Acide ribonucléique
ARNm : ARN messenger
ARNt : ARN de transfert
AS : Alternative Splicing
ATGC : Arginine, Thymine, Guanine, Cytosine
AUGC : Arginine, Uracil, Guanine, Cytosine
B : B cell
Ba : Basophil
BC : Band Cells
BCR : Récepteur des cellules B / B cell receptor
BD : Becton Dickinson
Ca²⁺ : Calcium
CD : Classe de différenciation / Cluster of differentiation
CD4⁺/8⁺ : Cluster of differentiation antigen 4⁺/8⁺
CHAK : CC-Chemokine-Activated Killers
ChIP-Seq : Chromatin Immunoprecipitation Sequencing
CMH : Complexe majeur d'histocompatibilité
CNTF : Ciliary neutrophil factor
CO₂ : Dioxyde de carbone
CPM : Counts Per Million
CPT : Cell Preparation Tube
DC : Dendritic cell
DGE : Differential Gene Expression
DNA : Deoxyribonucleic Acid
EDG : Expression Différentielle des Gènes
ENCODE : The Encyclopedia of DNA Elements
EST : Expressed Sequence Tags
eQTL : Expression quantitative trait loci

FPKM : Fragments Per Kilobase Million
FT : Facteur de transcription
GALT : Gut-associated lymphoid tissue
GDP : Guanosine diphosphate
G-CSF : Granulocyte CSF
GG : granule gélatinase
GGR : The Genomics of Gene Regulation
GM-CSF : Granulocyte macrophage CSF
Gn : Granulocyte
GNLY : Granulysine / Granulysin
GO : Gene Ontology
GTP : Guanosine triphosphate
GU : Guanine, Uracile
Gzm : Granzyme
H1/2A/2B/3/4 : Histone
HCDM : Human Cell Differentiation Molecules
HLDA : Human Leukocyte Differentiation Antigens
IC : Immune complexes
iDC : immature DC
ILC : Innate lymphoid cell
ImmGen : The immunological Genome
IR : Intron retention
ISR : Isoform-Specific Region
KD : Knockdown
LEC : Lymphatic endothelium
LN : Lymph node
lncRNA : long non-coding RNA/ ARN long non codant
LPS : Lipopolysaccharide
MΦ : Macrophage
M1/2 : Macrophage de type ½
MAPK : MAP kinase
MB: Myeloblasts
MB : méga bases
MC: Myelocytes
M-CSF : Monocyte CSF
MF : Macrophage
MG : Microglia
MHC : Major histocompatibility complex
MHI : Montreal Heart Institute
MISO : Mixture-of-isoforms
MM: Metamyelocytes
Mo : Monocyte
MRG: Module Representative Genes
Mtb : Mycobacterium tuberculosis

MTM : Moyenne Tronquée des valeurs M
N : Neutrophil
NCBI : National Centerfor Biotechnology Information
NCBI GEO : NCBI Gene Expression Omnibus
NCBI SRA : NCBI Sequence Read Archive
NET : Neutrophil extracellular trap
NGD : No-Go Decay
NHGRI : National Human Genome Research Institute
NIH : National Institutes of Health
NK : Natural killer
NKT : Natural killer T
NMD : Nonsense-mediated mRNA decay
NSD : Nonstop mediated mRNA Decay
ONT : Oxford Nanopore
PacBio : Pacific Biosciences
pb : Paires de bases
PC : Plasma cell
PCR : Polymerase Chain Reaction
pDC : plasmacytoid DC
PFN : Perforine / Perforin
P_i : Phosphate
PIC : Preinitiation complex
PM: Promyelocytes
PMN: Polymorphonuclear leukocytes
Pol II : Polymérase de type II
Poly(A) : polyadenylation
pré-ARNm : ARN pré-messager
PRR : Pattern recognition receptor / Récepteur de reconnaissance de motif moléculaire
PSI : Percent spliced in index
RBC : Red blood cell
RE : Réticulum endoplasmique
rMATS : Replicate multivariate analysis of transcript splicing
RNA : Ribonucleic Acid
RNA-Seq : RNA-Sequencing / Séquençage d'ARNm
Roadmap : The Roadmap Epigenomics Mapping Consortium
ROS : Reactive Oxygen Species
RPKM : Reads Per Kilobase Million
RSEM : RNA-Seq by Expectation Maximization
SC : Segmented neutrophilic cells
scRNA-Seq : single cell RNA-Sequencing / Séquençage d'ARNs sur cellule unique
SC : Stem cell
SG : Granule Spécifique
SIT : Site d'initiation de la transcription
SL : Site de liaison

snRNA : small nuclear RNA / petit ARN nucléaire
sQTL : Single quantitative trait loci
SV : Vésicule Sécrétrice
T_{ab} : $\alpha\beta$ T cell
T_{gd} : $\gamma\delta$ T cell
T_{CM} : Central memory T cell
TCR : Récepteur des cellules T / T cell receptor
T_{EF} : Effector T cell
TEG : Top Expressing Gene
TF : Transcription factor
T_{FH} : Follicular helper T
T_{FR} : Follicular regulatory T
TMM : Trimmed Mean of M-values
TOM : Topology overlap matrix
TPM : Transcripts Per Million
TSS : Transcription start site
TXI : TXIImport
U2AF1 : U2 small nuclear RNA auxiliary factor 1
UCSC : Université de Californie à Santa Cruz
UMI : Unique Molecular Identifiers
UTR : Untranslated region
VSV : Vesicular stomatitis virus
WBC : White blood cell
WGCNA : Weighted gene co-expression network analysis
WTA : whole-transcriptome analysis
Zn : Zinc

*« Celui qui ne peut pas instruire sa famille n'est pas
celui qui peut instruire les autres. »*

Proverbe Mandchou

REMERCIEMENTS

Je tiens tout d'abord à remercier mon directeur de recherche le Dr John Rioux pour son accueil au sein de son laboratoire et pour la confiance qu'il m'a accordée. Mais aussi pour sa disponibilité, ses conseils, sa patience, son soutien moral et financier et pour sa précieuse aide dans la rédaction des articles et la relecture de ma thèse. Je suis très honorée d'avoir eu pour directeur de recherche le Dr John Rioux.

Je remercie mon co-directeur, le Dr François Major, pour ces précieux conseils pour mon avenir et ses encouragements tout au long de ma thèse. Je souhaite aussi remercier Dre Sylvie Mader et Dr Guillaume Bourque pour avoir accepté de faire partie de mon comité de suivi de thèse et pour leurs idées pertinentes.

Je remercie également l'ensemble des membres du laboratoire du Dr John Rioux que j'ai eu la chance de côtoyer tout au long de ma thèse et avec qui j'ai partagé de très bon moments. Je remercie particulièrement Sylvain et Frédéric, les deux bio-informaticiens du laboratoire, qui ont toujours été là pour répondre à mes questions et ont été de très bons conseils. J'ai une pensée particulière pour Gabrielle, qui a été très patiente et compréhensive avec moi dans mon apprentissage d'outils bio-statistiques et qui a su m'encourager quand je me sentais dépassée. Je remercie aussi Philippe, Claudine et Mohamad qui ont toujours accepté de répondre à mes questions sur la biologie, même pour les plus basiques. Je remercie tout particulièrement Marie-Ève qui a généreusement accepté d'être la relectrice attitrée de ma thèse.

Je souhaite également remercier mes proches qui m'ont été d'un grand soutien moral et à l'écoute quand j'en avais besoin d'eux. Je remercie tout d'abord mes amis de long date, Madely, Mahéva et Edwin, qui malgré la distance, ont été présents pour moi. Je remercie mes nouvelles amies que Montréal m'a permis de rencontrer, Lyvia, Judy, Bianca, Cassandre, Yasmine, Sandra, et surtout Vanessa et Flora.

Je remercie aussi mon père que j'ai rejoints à Montréal et qui m'a épaulé tout au long de ma thèse et ma mère sans qui je n'aurai peut-être pas pu voir toutes les merveilleuses choses qu'offrent un doctorat. Je remercie aussi mes frères et ma soeur qui savent toujours comment me redonner le sourire.

1 Chapitre 1 – Introduction

1.1 Le génome humain

1.1.1 Rôle et fonction du transcriptome

Le système biologique humain s'appuie sur son génome pour construire ses différents transcriptomes et permettre ainsi l'établissement de la complexité de son organisme. L'organisme humain étant composé de plusieurs types cellulaires ayant chacun des fonctions spécifiques, l'étude des différents transcriptomes a permis de conclure que chaque type cellulaire est constitué d'un ensemble unique de gènes exprimés. Ainsi la comparaison et l'analyse des transcriptomes de plusieurs types cellulaires et tissus humains ont montré la formation de clusters de gènes enrichis pour des caractéristiques et fonctions spécifiques aux différents types cellulaires et tissus étudiés (1, 2). De plus, une étude de M. Melé et al. (3) a comparé l'expression génique dans plusieurs tissus et lignées cellulaires humains (43 organes, sous-régions du cerveau, sang et lignées cellulaires). Ils ont prouvé que moins de 200 gènes étaient exclusivement exprimés dans un tissu donné et que la variation d'expression génique entre les tissus était de 43%, beaucoup plus important qu'entre les individus, qui est de 4% dans leur étude.

Certains gènes ont la particularité de pouvoir exprimer plus d'un transcrite qui vont encoder une ou plusieurs isoformes de la protéine. Cette caractéristique offre un niveau de complexité supplémentaire pour la compréhension de l'impact du transcriptome, et indirectement du protéome, sur la fonction cellulaire. L'étude de M. Melé (3) a aussi analysé, dans les tissus, l'impact de la variation de l'épissage, un mécanisme permettant la maturation des transcrits et qui est donc, directement, impliqué dans la décision du choix des isoformes protéiques exprimées au sein de la cellule. Cette étude a montré qu'une large proportion de cette variation correspond à une variation résiduelle inexplicée alors que les valeurs de variation de l'épissage à travers les individus et à travers les tissus sont similaires.

Le projet GENCODE, qui a pour objectif d'annoter les génomes d'humain et de souris, décompte au total 60 660 gènes chez l'humain dont 19 962 codant pour une protéine, pour sa version 36. Ces 19 962 gènes regroupent 85 269 transcrits qui vont pouvoir être traduits en protéines. Trois études de V.A. Blomen et al., T. Wang et al. et T. Hart et al. (4-6) ont démontré, dans des lignées cellulaires humaines, qu'environ 2 000 gènes étaient essentiels pour le développement et la prolifération des cellules humaines. Ces gènes codent pour des protéines impliquées dans l'épissage, la translation et le trafic de protéines mais pas dans la signalisation cellulaire. Pour identifier ce sous-groupe de gènes essentiels pour l'établissement et le maintien de la cellule, ces trois études ont utilisé des approches permettant d'inactiver près de 18 000 gènes en inhibant leur expression avec deux méthodes : vecteur de piégeage (soit *gene-trap* en anglais) et *CRISPR*. Dans l'étude de T. Hart et al. (6), ils ont démontré que 1 580 gènes constituaient le corps essentiel des cellules humaines dont 829 (soit 52%) étaient communs aux 5 lignées cellulaires utilisées dans cette étude (HTC116, HeLa, DDL1, GBM et RPE1). Pour les gènes ne composant pas le corps essentiel, ils ont pu prouver qu'ils marquaient les différences biologiques entre ces lignées

cellulaires. Cela supposerait que chaque type cellulaire possède un ensemble de processus biologiques. Cette observation est aussi validée par les deux autres études (4). Par exemple, dans l'étude de T. Wang et al. (5), trois gènes (*EBF1*, *POU2AF1* et *PAX5*) se trouvent dans la liste des gènes essentiels de la lignée cellulaire représentant les cellules B, placés au rang 3, 6 et 8 respectivement. Ces gènes codent pour des protéines qui sont des facteurs de transcription (FT) et sont indispensables dans la différenciation et la fonction des cellules B (7-13). *MEF2B* fait aussi partie des gènes essentiels et est impliqué dans l'activation de *BCL6* dans les cellules B (14). Ces études montrent que parmi les gènes essentiels à une cellule, un certain nombre d'entre eux sont uniques à un type cellulaire et semblent être impliqués dans les fonctions spécifiques de ce type cellulaire.

1.1.2 Mécanisme et éléments régulateurs de la transcription

Pour établir un transcriptome au détriment d'un autre au sein d'une cellule, de nombreux mécanismes biologiques sont impliqués, allant de l'ouverture d'une région génomique à la liaison entre un élément régulateur et l'ADN pour finir par un remodelage de transcrits (exemple : région promotrice, FT, épissage, etc.). Ces mécanismes sont essentiels pour la cellule et permettent ainsi la présence d'un ensemble unique de transcrits matures appelés ARN messager (ARNm). Ces ARNm vont pouvoir par la suite être traduits en protéines et ainsi fournir à la cellule un protéome qui lui est spécifique.

1.1.2.1 Régions génomiques régulatrices

Les régions génomiques régulatrices se situent dans les régions non codantes de l'ADN et ciblent un gène ou un cluster de gènes. Ces régions peuvent activer l'expression génique (promoteur proximal et amplificateur) ou la réprimer (silenceur et *insulator*) (**Figure 1.1**).

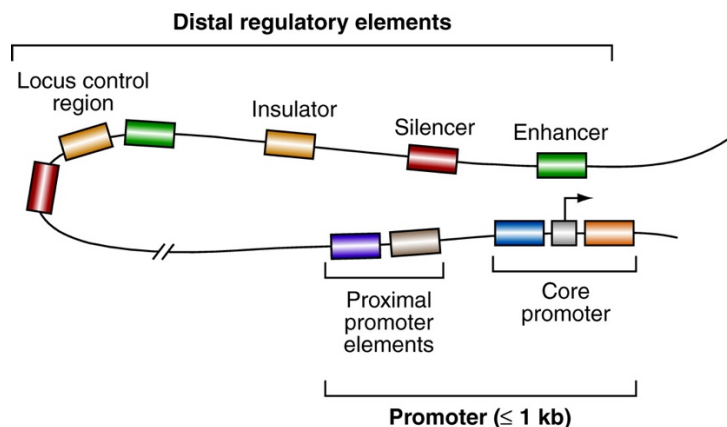


Figure 1.1. Types de régions régulatrices de l'expression génique. Tiré de (15).

Le promoteur est la combinaison du cœur du promoteur et du promoteur proximal. Le cœur du promoteur se situe en amont d'un gène et couvre environ 100 paires de bases (pb) autour du site d'initiation de la transcription (SIT) (16). Cette région permet la liaison de la polymérase de type II (Pol II) à l'ADN par l'intermédiaire de facteurs généraux de la transcription (exemple : TFIIA, TFIIB, TFIID, TFIIE, TFIIH), ce qui va initier la transcription du gène à proximité (**Figure 1.2**). Cette liaison permet un faible niveau d'expression du gène appelé expression basale. Le promoteur proximal, quant à lui, commence à une centaine de pb en amont du SIT. Il contient les

sites de liaison des FT qui, seul ou associés à un ou plusieurs cofacteurs, se lient à la Pol II et ainsi, augmentent le niveau d'expression du gène cible (15, 16). Généralement, le promoteur est défini par 1000 pb en amont du SIT et 500 pb en aval de celui-ci. Un gène peut avoir plus d'un SIT et ainsi permettre la transcription de plusieurs transcrits.

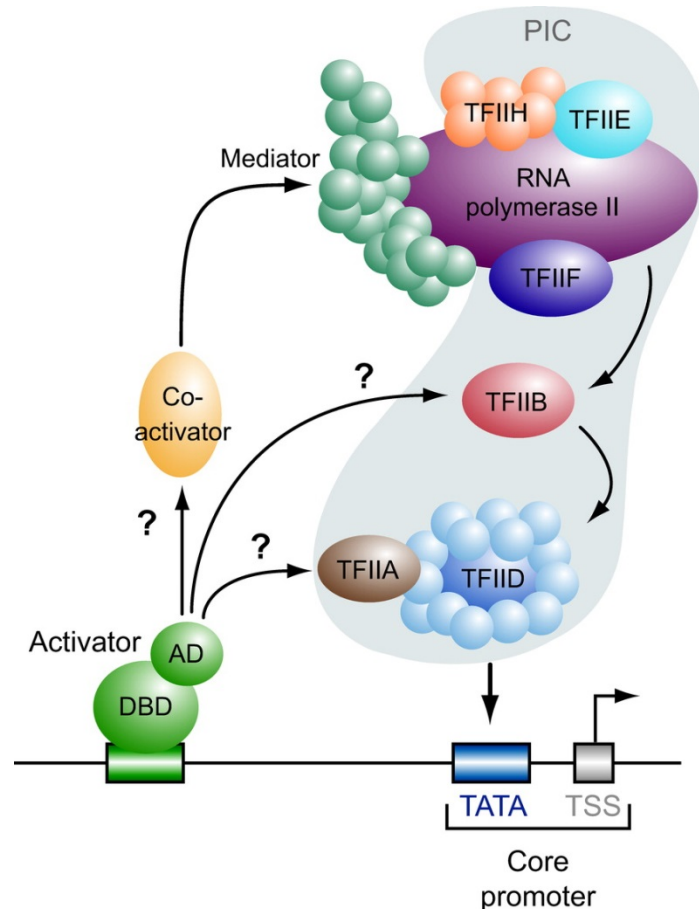


Figure 1.2. Mécanisme de la transcription chez les eucaryotes. Les facteurs impliqués dans l'initiation de la transcription sont classifiés en trois groupes distincts : les FT généraux, les activateurs et les co-activateurs. Le cœur du promoteur (appelé en anglais *core promoter*) est constitué des FT généraux (Pol II, TFIIA, TFIIB, TFIID, TFIIE, TFIIF et TFIIH) qui forment le complexe pré-initiatique (appelé en anglais *preinitiation complex (PIC)*) et permet ainsi la liaison directe entre la Pol II et le SIT (appelé en anglais *transcription start site (TSS)*). Les facteurs activateurs vont être responsables de la stimulation de l'activité transcriptionnelle en liant les éléments régulateurs en amont du SIT et en formant un complexe composé d'un domaine liant l'ADN (appelé en anglais *DNA-binding domain (DBD)*) et un domaine d'activation (appelé en anglais *activation domain (AD)*). Les cibles directes des facteurs activateurs sont inconnues. Tiré de (15)

Certaines régions, situées à plusieurs centaines de pb du SIT, peuvent aussi réguler l'activité de la Pol II, soient les amplificateurs, silenceurs, *insulators* et régions contrôles de locus (**Figure 1.1**). Les amplificateurs activent l'expression d'un gène cible par l'intermédiaire d'un FT qui lie la Pol II, favorisé par le repliement de l'ADN, formant ainsi une boucle. Les silenceurs suivent le même

mécanisme de liaison à la Pol II que les amplificateurs mais ont un rôle opposé, en diminuant ou bloquant l'expression du gène cible. Les *insulators*, eux, vont lier le FT d'un amplificateur et ainsi empêcher sa liaison avec la Pol II ou vont entraver la formation de la boucle d'ADN (15-17).

Les régions contrôles de locus sont composées d'un groupe de régions régulatrices de types amplificateurs, silenceurs et *insulators* et régulent un locus ou un cluster de gènes. Ces régions sont généralement spécifiques à un tissu et ont des propriétés spatiales et temporelles permettant la régulation de l'expression de gènes cibles. Les régions contrôles de locus sont appelées super-amplificateurs lorsqu'elles sont constituées principalement d'éléments amplificateurs (15, 18-21).

La présence de ces régions dans le génome ne suffit pas pour réguler l'expression génique. Ces régions doivent être combinées à deux types de mécanismes pour être fonctionnelles. Tout d'abord, l'accessibilité à ces régions est régulée par des éléments impliqués dans le mécanisme d'ouverture et de fermeture de l'ADN (histones et îlots CpG) (22-25). Par la suite, lorsque l'ADN est en conformation ouverte, d'autres éléments de régulation appelés FT, vont pouvoir s'y fixer et initier ou réprimer l'expression du ou des gènes cibles dépendamment du type de la région où se trouve leurs sites de liaison (26).

1.1.2.2 Histones : Éléments et mécanismes de régulation des régions génomiques

L'ADN a la capacité de réguler l'accessibilité aux régions régulatrices et aux gènes en modifiant sa condensation. On parle d'euchromatine lorsqu'elle est faible et hétérochromatine quand elle est forte. La chromatine est composée de nucléosomes, qui sont des complexes composés d'un segment d'ADN d'environ 150 paires de nucléotides, enroulé autour d'un cœur de protéines d'histone. Le cœur du nucléosome est composé de huit protéines d'histone dont deux exemplaires des classes H2A, H2B, H3 et H4 (**Figure 1.3**). Une histone de la classe H1 se lie à l'ADN et lui permet ainsi de s'enrouler autour du cœur du nucléosome. Les modifications post-traductionnelles des histones influent sur la condensation de l'ADN et impactent ainsi plusieurs processus nucléaires (exemple : la transcription, la réplication ou la réparation de l'ADN). Elles régulent aussi l'accès aux gènes et aux régions impliquées dans la régulation de l'expression génique, soit l'épigénétique. Ces modifications s'appliquent sur l'extrémité N-terminale des histones, comme des phosphorylations, acétylations, méthylations, ubiquitinations et sumoylations, seule ou en combinaison, et permettent ainsi l'activation ou la répression de l'expression de gènes cibles. L'étude de ces histones a démontré qu'un grand nombre de modifications post-traductionnelles régulent des processus biologiques, mais il reste encore de nombreuses relations à établir (27, 28).

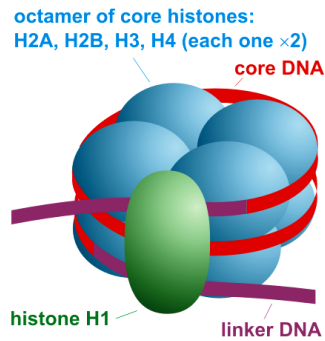


Figure 1.3. Schéma de l'organisation d'un nucléosome. Tiré de (29).

1.1.2.3 Facteurs de transcription : Éléments de régulation de l'expression génique

Les régions régulatrices ne peuvent être fonctionnelles que si un élément de régulation spécifique s'y lie, appelé facteur de transcription. La liaison entre la région régulatrice et le FT détermine le type de régulation à appliquer sur le gène cible, soit activer ou réprimer son niveau d'expression.

Le FT est une protéine qui contient deux domaines : un premier domaine lui permettant de lier l'ADN sur son site de liaison et un second, qui va réguler directement ou indirectement l'activité de la Pol II. Le site de liaison (SL) d'un FT est composé de 6 à 12 acides nucléiques et leur étude a montré que celui-ci a une haute complexité. Une matrice de poids permet de définir l'ensemble des motifs possibles pour le SL d'un FT donné en calculant pour chacune des positions du SL la probabilité de rencontrer chacun des 4 acides nucléiques qui composent l'ADN (**Figure 1.4**) (30). Certains FT forment un complexe protéique avec un ou plusieurs autres FT ce qui leur permet de lier l'ADN. Cette liaison se fait soit par l'intermédiaire du SL d'un ou de plusieurs des FT du complexe et doit respecter des propriétés spatio-temporelles, ce qui peut ajouter un niveau de complexité au SL associé au complexe de FT (31, 32).

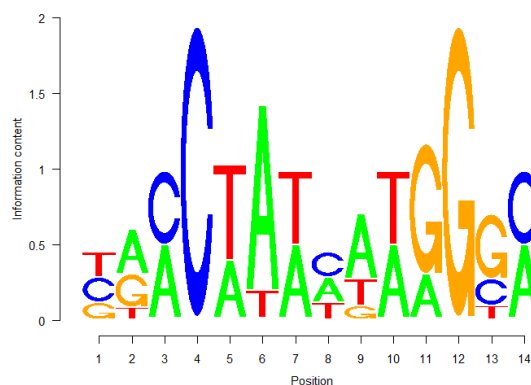


Figure 1.4. Exemple de matrice de poids. Tiré de [Dave Tang's blog](#).

L'étude des protéines cellulaires a permis d'avoir aujourd'hui de nombreuses informations sur elles, telles que leur structure ou leur localisation subcellulaire ainsi que leur fonction dans de

nombreux types cellulaires. À l'aide de ces connaissances, Lambert et al. (26) ont pu cataloguer l'ensemble des FT présents chez l'humain et ont compilé dans des fiches les connaissances existantes pour chacun de ces FT. Ils ont pu associer un motif de site de liaison aux deux tiers des FT humains. Ces données sont accessibles sur leur site web : <http://humantfs.cabr.utoronto.ca>. Les critères utilisés pour identifier leur liste de 1 639 FT sont les suivants : (1) être capable de lier l'ADN sur une séquence spécifique et (2) pouvoir réguler la transcription génique.

Les FT sont essentiels dans la vie d'une cellule. Leur capacité à réguler l'expression génique leur offre un rôle important dans le développement et la différenciation cellulaire, la survie et la mort cellulaire ainsi que la fonction cellulaire. Par exemple, PAX5 et EBF1 sont primordiales pour l'établissement et le maintien de l'identité des cellules B tout au long de leur développement (7, 11).

1.1.2.4 Épissage : Mécanisme de régulation de l'ARN messager

Un gène est défini par sa composition en exons et introns. La transcription d'un gène est conduite par la Pol II qui produit un ARN pré-messager (pré-ARNm) qui, après application de l'épissage alternatif, va donner un ARNm prêt pour la traduction en protéine. Ce pré-ARNm correspond à la région génomique complète, soit les exons et les introns. L'épissage alternatif a pour première fonction de retirer les régions introniques. Pour cela, le spliceosome, un complexe de ribonucléoprotéines composé de cinq petits ARN nucléaires (U1, U2, U4, U5 et U6) ainsi que de plus d'une centaine de protéines spliceosomales, va reconnaître les dinucléotides GU à l'extrémité 5' et AG à l'extrémité 3' des introns ainsi que la séquence point de branchement (appelé en anglais *branch point*) intronique, à proximité de l'extrémité 3' (**Figure 1.5**). Ces 3 régions vont permettre au spliceosome de se lier à l'intron, le catalyser et le relâcher dans le noyau pour qu'il soit dégradé (33).

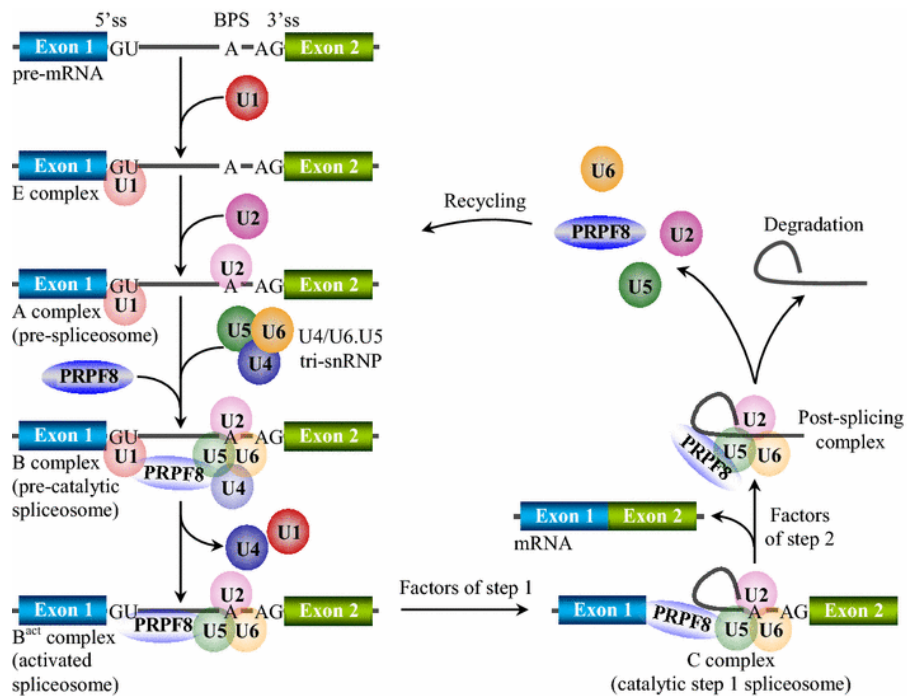


Figure 1.5. Spliceosome et formation de l'ARNm. Le facteur d'épissage PRPF8 interagit avec les petits ARN nucléaires (U1, U2, U4, U5 et U6) et plusieurs protéines du spliceosome (telles que BRR2 et SNU114). Ces interactions sont suivies de plusieurs évènements dont l'élimination de l'intron (34). Tiré de (33).

Les évènements d'épissage alternatif sont régulés par plusieurs autres éléments. Tout d'abord, les quatre types d'éléments régulateurs *cis-acting* (amplificateurs et silenceurs exoniques d'épissage, amplificateurs et silenceurs introniques d'épissage) sont liés spécifiquement par des protéines régulatrices non-spliceosomales, et ainsi, activent ou inhibent l'activité du spliceosome. L'impact de ces éléments sur l'activité du spliceosome est dépendant de leur position dans l'ARN et des facteurs qui s'y lient (33). Par exemple, la protéine SETD2 sélectionne spécifiquement l'exon IIIc du gène *FGFR2* à la place de l'exon IIIb (**Figure 1.6**) et le facteur NOVA1 permet la conservation de l'exon 9 du gène *GABRG2* lorsqu'il lie l'amplificateur intronique d'épissage, mais exclut l'exon 4 pour ce même gène quand il se lie à la région silenseur exonique d'épissage (33, 35).

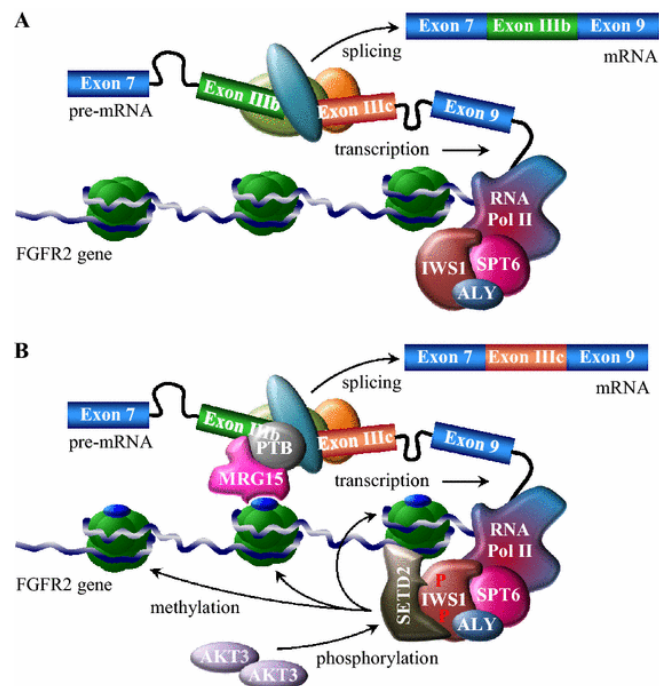


Figure 1.6. Contrôle épigénétique de l'épissage alternatif. **A.** Formation de l'ARNm contenant l'exon IIIb. **B.** Formation de l'ARNm avec l'exon alternatif IIIc. La protéine SETD2, composant du complexe formé autour de la Pol II, est responsable de la méthylation des histone H3-K36 du gène *FGFR2* et ainsi, peut prédéterminer la sélection de l'exon IIIc ou IIIb lors de l'épissage alternatif. Le mécanisme de sélection de l'exon IIIc résulte de la liaison entre la protéine MRC15 et l'histone H3-K36me3 du gène *FGFR2* puis le recrutement du facteur d'épissage PTBP1. Tiré de (33).

Les histones peuvent aussi intervenir dans la régulation de la machinerie du spliceosome en recrutant une protéine spliceosomale qui va faire le pont entre l'histone et le spliceosome (**Figure 1.6**) (33). Cette liaison va favoriser l'exclusion d'un exon au détriment d'un autre (36, 37).

Pour finir, la vitesse d'élongation va aussi impacter l'épissage alternatif (**Figure 1.7**). La transcription et l'épissage alternatif sont deux mécanismes qui peuvent être actifs en même temps sur un même brin de pré-ARNm. Si l'élongation est lente, lorsque le spliceosome sera actif, certains sites impliqués dans l'épissage alternatif seront inaccessibles car non transcrits. Ce mécanisme va empêcher l'exclusion ou la conservation de certains exons et donner un ARNm différent (33, 38).

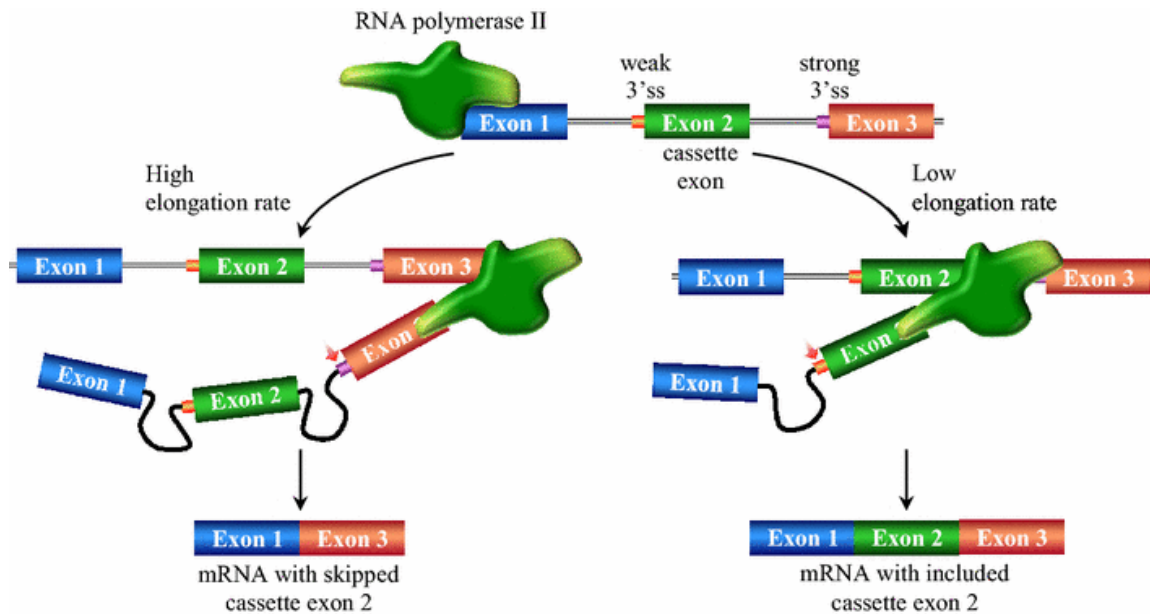


Figure 1.7. Schématisation de l'impact du temps d'élongation sur l'épissage alternatif. L'élongation lente permet la formation du spliceosome sur un faible site d'épissage en amont, alors que l'élongation rapide met en compétition le faible site d'épissage en amont, avec un site d'épissage fort placé en aval de l'ARNm. Cela amène à l'élimination de l'exon 2. Tiré de (33).

De nombreux autres mécanismes sont impliqués dans la régulation de l'activité du spliceosome comme la structure secondaire du pré-ARNm. Tous ces mécanismes travaillent en synergie pour déterminer la proportion d'expression des transcrits d'un gène (33).

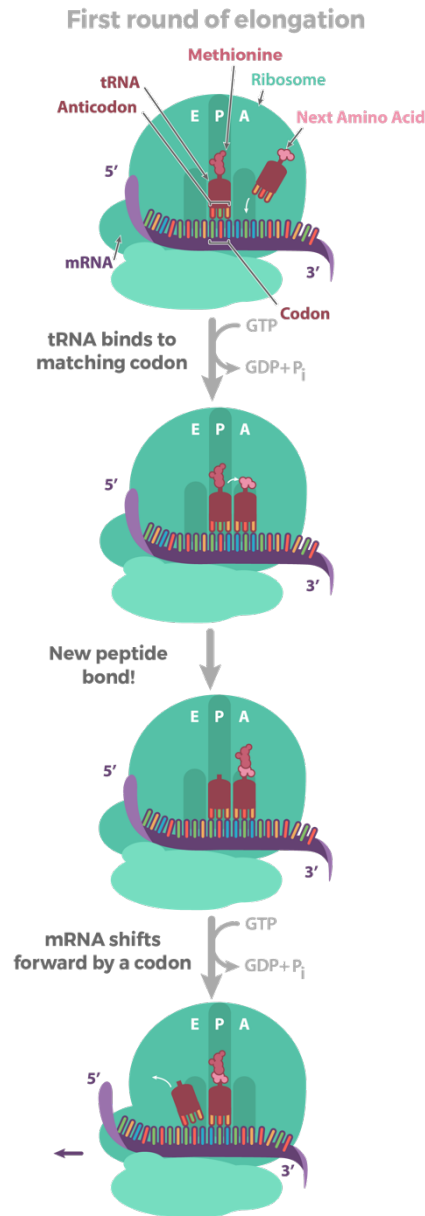
1.1.3 Encodage des protéines et expression de multiples isoformes

L'ARNm est l'élément le plus mature qui appartient au transcriptome. La prochaine étape est la traduction en protéine de cet élément puis, après un certain nombre de lectures, ce dernier va être dégradé pour permettre la production de nouveau pré-ARNm. Ainsi, après la transcription et l'épissage, la traduction permet d'obtenir une isoforme prédéfinie de la protéine. L'ensemble des protéines d'une cellule constitue son protéome et elles permettent à la cellule d'être fonctionnelle et d'interagir avec son environnement.

1.1.3.1 Traduction : Mécanisme de production de protéines

Après l'épissage du pré-ARNm, l'ARNm est transporté dans le cytoplasme où aura lieu sa synthèse en protéine par le ribosome. Le rôle du ribosome est de lier chaque triplet de nucléotides soit un codon, à son anticodon appelé ARN de transfert (ARNt) et de le traduire en acide aminé (a.a.). Pour cela, le ribosome est composé de deux sous-unités dont la petite sous-unité contient 3 sites

de liaison (A, P et E) aux ARNt et lie l'ARNm. Le site A permet la liaison des ARNt aminoacylés correspondants après reconnaissance spécifique du codon par la sous-unité du ribosome. Le génome humain possède au total 64 codons qui codent pour 22 a.a. ainsi que les trois codons stop. Lorsque l'ARNt arrive sur le site P, son a.a. se lie à la chaîne polypeptidique naissante catalysée par la grande sous-unité du ribosome. Pour finir, l'ARNt déacylé transite sur le site E



avant d'être éjecté du ribosome (**Figure 1.8**) (39-41).

Figure 1.8. Schématisation du mécanisme de traduction des protéines. La traduction en protéine de l'ARNm est guidée par le ribosome (bleu) qui est composé d'une petite et une grosse sous-unité contenant trois sites de liaison de l'ARNt : A, P et E. Le ribosome lie un ARNm et recrute un complexe ternaire constitué d'un ARNt, d'un GTP et de l'acide aminé correspondant au

codon présent sur le site A du ribosome. La liaison entre le site A et le complexe ternaire correspondant déclenche l'hydrolyse du GTP et sa dissociation du complexe (Accommodation). Par la suite, une liaison peptidique est catalysée entre l'acide aminé de l'ARNt situé sur le site A et la protéine en formation portée par l'ARNt situé sur le site P (Transpeptidation). Se suit un déplacement du ribosome et les ARNt présents sur les sites A et P se retrouvent respectivement sur les sites P et E. Cela va permettre la libération de l'ARNt du site E (Translocation) et la traduction du nouveau codon présent sur le site A (Accommodation). Tiré de (42).

Ainsi pour mener à bien la synthèse des protéines, ce mécanisme se découpe en trois parties : l'initiation, l'élongation et la terminaison. Lors de l'initiation, la petite sous-unité du ribosome reconnaît le codon *start* de l'ARNm et permet à l'ARNt initiateur de s'y lier. Par la suite la grande sous-unité du ribosome se lie au complexe et traduit l'ARNt initiateur en l'acide aminé méthionine. L'élongation suit le même procédé que l'initiation et associe à chaque codon suivant l'ARNt correspondant. La grande sous-unité du ribosome va permettre ainsi la catalysation des liaisons peptidiques entre les acides aminés. La terminaison débute lorsque le ribosome rencontre un codon stop, ce qui enclenche la dissociation entre la grande sous-unité du ribosome et la chaîne d'acides aminés.

1.1.3.2 Isoformes protéiques

Certains gènes ont la particularité de pouvoir exprimer plus d'un transcrit, obtenus par différents SIT et/ou par l'épissage alternatif (**Figure 1.9**). L'étude de certains de ces gènes a démontré que la différence d'expression de leurs transcrits offrait un niveau d'information et de spécificité supplémentaires au transcriptome. Ces transcrits sont par la suite, traduits en protéine et deux transcrits d'un même gène peuvent encoder deux isoformes de la protéine. Une isoforme alternative d'une protéine est une isoforme ayant une séquence peptidique différente de la séquence de l'isoforme de référence de la protéine. Les isoformes d'une même protéine proviennent de l'expression d'un unique gène et se distinguent par la ou les délétions/insertions de régions protéiques (exemples : extrémités 3' ou 5') ou la substitution d'une ou plusieurs séquences peptidiques. De plus, les isoformes d'une protéine peuvent diverger par leurs fonctions, leurs localisations cellulaires, leurs activités enzymatiques ou leurs interactions physiques avec une autre protéine (43).

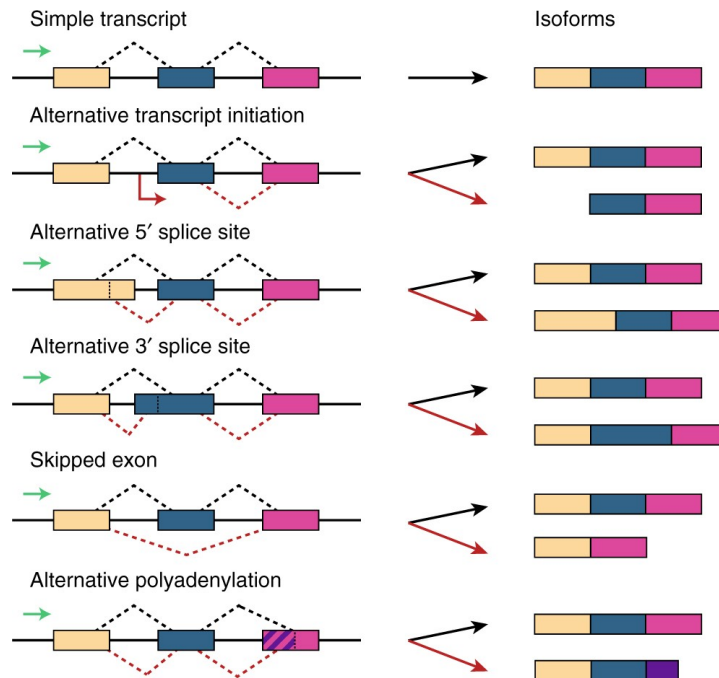


Figure 1.9. Schématisation de plusieurs mécanismes d'épissage alternatif et des variants d'ARNm obtenus à partir d'un unique gène. Dans cette figure, les isoformes (appelé en anglais *isoforms*) correspondent à des variants d'ARNm. Tiré de (44).

Une étude de X. Yang et al. (45) a montré que sur 1 492 gènes humains étudiés seul 506 gènes expriment des transcrits qui codent pour au moins deux isoformes de la protéine du gène. Cette étude a été réalisée dans 5 tissus différents (le cerveau, le foie, le cœur, le placenta et les testicules). X. Yang et al. ont étudié l'interaction protéine-protéine de 398 gènes des 506 car leurs transcrits codent pour au moins 2 différents isoformes de la protéine. Pour compléter leur analyse, ils ont identifié les régions spécifiques des isoformes d'une longueur minimum de 40 acides aminés (a.a.). La combinaison de ces deux résultats a permis de caractériser 4 types d'interaction avec une autre protéine, appelée partenaire : (1) la « *Promoting* », (2) la « *Inhibiting* », (3) la « *Promoting and inhibiting* » et pour finir (4) la « *Complex* » (**Figure 1.10**). Au total, 133 gènes codent pour des isoformes de protéines ayant le type d'interaction « *complex* » ce qui suggère que l'interaction protéine-protéine est contrôlée par la combinaison de multiples régions spécifiques d'isoformes dans une grande partie des cas.

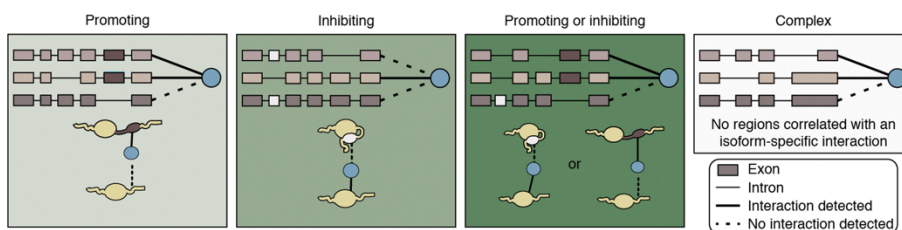


Figure 1.10. Quatre catégories des régions spécifiques d'isoformes (appelé en anglais isoform-specific regions (ISRs)) permettant des effets sus les interactions. *Promoting* : interaction du partenaire uniquement avec les isoformes de la protéine qui possède une ISR. *Inhibiting* :

l'interaction contraire car le partenaire se lie aux isoformes qui n'ont pas de ISR. *Promoting and inhibiting* : interaction entre le partenaire et l'isoforme qui a une ISR et a perdu une ISR. Complex : qui ne montre pas un profil spécifique de gain ou de perte de ISR pour pouvoir interagir avec son partenaire. Tiré de (45).

Ils ont aussi étudié l'interaction domaine-domaine et les résultats suggèrent que l'épissage alternatif de domaines protéiques pourrait expliquer les nombreuses différences d'interaction des isoformes d'une protéine avec ses partenaires. Pour finir, ils ont comparé la liste des partenaires des isoformes d'une protéine. Ils ont observé que les protéines avec des isoformes à un ou plusieurs partenaires spécifiques avaient comme particularité : leurs isoformes jouaient un rôle important dans la spécificité des tissus où ils étaient exprimées et ils avaient des fonctions distinctes des autres isoformes de la protéine dans ces tissus. De plus, X. Yang et al. ont analysé l'expression de ces gènes dans les 5 tissus et ont découvert que pour 235 gènes parmi les 506, dans un tissu donné, le transcrit qui code pour une isoforme alternative de la protéine est plus fortement exprimé que le transcrit qui encode l'isoforme de référence de la protéine. Cela suggère qu'une isoforme, et indirectement le transcrit qui l'encode, pourrait être spécifique à un tissu et avoir une fonction spécifique dans ce dit tissu.

De nombreuses études ont démontré que les isoformes d'une protéine pouvaient avoir des fonctions qui leurs sont spécifiques (43, 46). Par exemple, le gène *BCL2* code pour deux isoformes de la protéine BCL2L1 : Bcl-xL qui est impliqué dans la survie de la cellule et Bcl-xS qui lui est pro-apoptotique (47). De plus, la fonction spécifique des isoformes peut être couplée à une expression tissulaire ou cellulaire spécifique. Le gène *PTPRE* est connu pour être exprimé dans les macrophages, mais ses deux isoformes sont spécifiquement exprimées dans les sous populations M1 et M2 de macrophages et ont des fonctions distinctes dans ces populations. Dans les M1, l'isoforme de PTPRE est une protéine cytoplasmique inhibitrice de l'apoptose et de la différenciation en monocyte induite par IL6 et LIF (48). L'isoforme PTPRE exprimée dans les M2 est membranaire et est impliquée dans la distribution des podosomes, l'activation de ROCK2 et la migration cellulaire en réponse à LTD4 (49).

1.2 Le système immunitaire

L'hématopoïèse est le processus biologique qui produit l'ensemble des cellules sanguines. Ce processus implique la différenciation de cellules souches hématopoïétiques multipotentes en différentes cellules matures regroupées en trois catégories de lignées : les globules rouges, les cellules lymphoïdes et la lignée myéloïde. Les globules rouges permettent le transport du dioxygène (O₂) et du dioxyde de carbone (CO₂), respectivement, du poumon aux muscles et inversement. Les lignées lymphoïdes et myéloïdes constituent les cellules du système immunitaire qui ont pour rôle de protéger l'organisme contre les microbes (ex. : bactéries, virus, champignons et parasites) et les cellules stressées ; ils modulent également l'homéostasie du système immunitaire (**Figure 1.11**). Des organes composent aussi le système immunitaire et participent à la protection de l'organisme (ex. : moelle osseuse (50), organes secondaires (51-53), peau et muqueuses (51, 52)). Le système immunitaire a deux modes de réponse pour se défendre : réponses innée et adaptative. Les actions de ces réponses sont distinctes mais complémentaires et impliquent l'intervention de cellules immunitaires spécifiques (**Tableau 1.2**).

Le système immunitaire applique deux types de réponse à la défense de l'organisme, la réponse innée qui est la première, la plus rapide et la moins spécifique. Cette réponse est suivie par la réponse adaptative, 4 à 7 jours après le début de la défense immunitaire, qui est plus lente et plus spécifique car elle ne cible que les cellules présentant un antigène donné à leur surface (**Tableau 1.2**). Ces deux réponses impliquent les cellules myéloïdes et NK pour la réponse innée, et les cellules B et T pour la réponse adaptative.

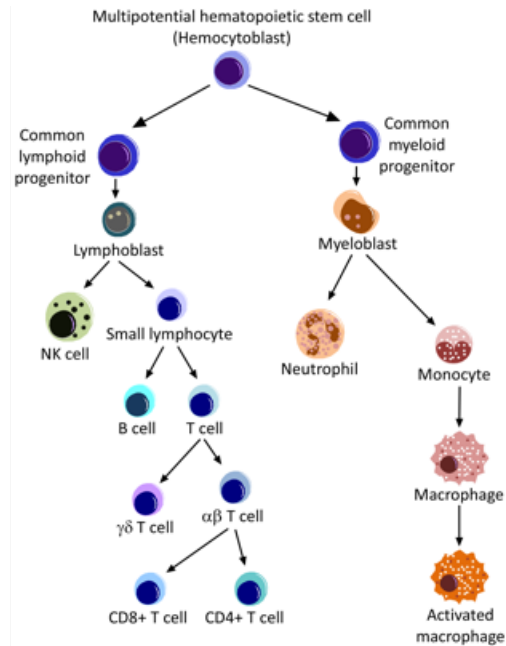


Figure 1.11. Schéma de différenciation des cellules immunitaires. Tiré de (54).

Tableau 1.1. Principales différences entre les réponses immunitaires innées et adaptatives. Traduit de (55).

	Inné	Adaptative
Échelle de temps	Heures	Jours
Type de reconnaissance	Modèle moléculaire associé à un pathogène	Antigènes
Spécificité	Rapide	Lente
Réponse à une seconde exposition	Identique	Induite et rapide
Types cellulaires	Cellules NK, monocytes, macrophages, neutrophiles, éosinophiles, basophiles et mastocytes	Lymphocytes T, lymphocytes B et cellules présentatrices

		d'antigènes (incluant les cellules dendritiques)
--	--	--

1.2.1 Les cellules immunitaires primaires

Les cellules lymphoïdes sont composées de trois grands groupes de population : les cellules B, les cellules NK et les cellules T qui représentent, respectivement, 1-7%, 1-6% et 7-24% des cellules immunitaires primaires présentes dans le sang. Les cellules T se composent aussi de trois sous-groupes de populations, les cellules T $\gamma\delta$, T CD4+ et T CD8+ (**Figure 1.11**). Ces six types cellulaires sont la résultante de la différenciation des progéniteurs lymphoïdes communs dans la moelle osseuse. Les cellules NK mature sont libérées dans le sang alors que les cellules B et T, qui sont immatures, vont finir leur développement dans les zones folliculaires ou marginales et dans le thymus, respectivement. Parfois les cellules NK progénitrices vont être libérées dans le sang et finir leur maturation dans les organes secondaires tels que les ganglions, la rate, le foie et le thymus (56). Après maturation, les cellules lymphoïdes B et T se retrouveront également dans la circulation. Contrairement aux autres cellules lymphoïdes, les cellules T vont majoritairement se trouver dans les nœuds lymphatiques et passer par le sang ou la lymphe pour circuler entre les nœuds lymphatiques. Parmi les cellules lymphoïdes, les cellules NK sont les seules impliquées dans la réponse immunitaire innée. Grâce à leurs récepteurs de surface, elles peuvent identifier les cellules du non-soi ou cancéreuses et favoriser leur élimination très rapide par lyse cellulaire. Pour les cellules lymphoïdes impliquées dans la réponse adaptative, les cellules T CD8+, dites cytotoxiques, vont aussi permettre la lyse cellulaire mais seulement après avoir été activées par les cellules T CD4+. Les cellules T CD4+ sont dites auxiliaires et, lorsqu'elles reconnaissent une cellule présentatrice d'antigène, elles prolifèrent et enclenchent l'activation des cellules T CD8+ et des cellules B. Les cellules B, comme les cellules T CD4+, ont la capacité de reconnaître les antigènes présents à la surface d'une cellule. Cette identification va activer les cellules B qui vont se différencier en plasmocytes et ainsi libérer dans le sang l'anticorps spécifique de l'antigène rencontré ou devenir une cellule B mémoire qui possède majoritairement, à sa surface, l'anticorps spécifique de l'antigène. Ces cellules B mémoires vont permettre une réponse plus rapide lors d'une seconde exposition à cet antigène. Ce mécanisme est un processus spécifique de la réponse immunitaire adaptative. Il permet de rendre la réponse plus rapide, en passant de 7 à 4 jours, mais reste plus lente que celle de la réponse innée qui contrôle l'infection durant cette période (57). Pour finir, les cellules T $\gamma\delta$ sont peu présentes dans le sang (3-5% des cellules T circulantes) et se trouvent majoritairement dans les organes secondaires tel que le poumon (moins de 10% des cellules T). Elles sont très tôt impliquées dans la réponse immunitaire innée en mobilisant les neutrophiles (58, 59). Les cellules T $\gamma\delta$ sont aussi responsables de la modulation de l'homéostasie du système immunitaire et peuvent exercer des fonctions immunosuppressives sur les cellules immunitaires primaires (59-61).

Les cellules myéloïdes sont composées de nombreuses populations de cellules provenant de la différenciation des progéniteurs myéloïdes communs (**Figure 1.11**). Parmi ces populations se trouvent les neutrophiles et les monocytes, deux types cellulaires qui jouent un rôle primordial dans la réponse immunitaire innée. Contrairement aux cellules lymphoïdes, les cellules myéloïdes atteignent leur maturation uniquement dans la moelle osseuse puis sont libérées dans le sang. Les neutrophiles sont impliqués dans les réponses pro-inflammatoires et antivirales de

l'organisme (62). Ils représentent 40 à 70% des cellules immunitaires circulantes, ce qui leur permet de faire partie de la première ligne de défense contre les pathogènes et les dommages cellulaires. Les neutrophiles s'activent lorsqu'ils rencontrent une cellule du non-soi identifiée grâce à ses récepteurs de surface. Afin d'éliminer un corps étranger, les neutrophiles le phagocyte pour le dégrader dans son cytoplasme et la mort cellulaire régulée des neutrophiles va produire des débris cellulaires pro-inflammatoires. Ces derniers vont être éliminés par les macrophages. Les macrophages, quant à eux, sont des cellules résultant de la différenciation des monocytes, qui sont eux même des cellules myéloïdes. Les monocytes représentent 2 à 12% des cellules immunitaires du sang et leurs rôles sont multiples dans le système immunitaire. Les monocytes vont phagocyter de petites quantités de déchets cellulaires, sécréter des cytokines pour mobiliser les cellules immunitaires sur les sites d'inflammation et être présentateurs d'antigènes pour les lymphocytes T. Lorsque les monocytes migrent dans un tissu, ils se différencient en macrophages et peuvent ainsi reconnaître les cellules du non-soi et les agents pathogènes, ce qui va permettre leur activation, menant à la phagocytose de ces corps étrangers. Contrairement aux neutrophiles, les macrophages digèrent la cellule cible mais cela n'entraîne pas leur mort. Les macrophages sont aussi impliqués dans la réparation et la cicatrisation cellulaire. Ces macrophages sont appelés macrophages anti-inflammatoires. Les monocytes et macrophages participent à la réponse immunitaire innée mais ils ont aussi la capacité de présenter les antigènes aux lymphocytes T CD4+, ce qui les associe à la réponse immunitaire adaptative.

1.2.2 La communication intracellulaire : Cytokines et chimiokines

Pour que les cellules immunitaires puissent trouver le site d'inflammation et déclencher les deux types de réponses immunitaires, elles doivent communiquer entre elles. Pour cela, chaque cellule immunitaire a la capacité de sécréter des petites molécules appelées cytokines qui vont lier les récepteurs à la surface de la cellule qui les a sécrétés (autocrine), les récepteurs des cellules proches de la cellule sécrétrice (paracrine) ou des messagers endocrines qui se lient aux cellules

Family	Members
Hematopoietic	
Common γ chain	IL-2, IL-4, IL-7, IL-9, IL-15, IL-21
Shared β chain (CD131)	IL-3, IL-5, GM-CSF
Shared	IL-2, IL-15
IL-2 β chain (CD122)	
Other hematopoietic	IFN- γ , IL-7, IL-13, IL-21, IL-31, TSLP
IL-1 family	IL-1 α , IL-1 β , IL-1ra, IL-18, IL-33
gp130-utilizing	IL-6, IL-11, IL-27, IL-31, ciliary neurotrophic factor (CNTF), cardiotrophin 1 (CT-1), leukemia inhibitory factor (LIF), oncostatin M (OSM), osteopontin
IL-12	IL-12, IL-23, IL-35
IL-10 superfamily	IL-10, IL-19, IL-20, IL-22, IL-24, IL-26, IL-28, IL-29
IL-17	IL-17A-F, IL-25 (IL-17E)
Interferons	
Type I interferons	IFN- α , IFN- β , IFN- ω
Type II interferon	IFN- γ (also a hematopoietic cytokine)
Type III interferons	IFN- λ 1 (IL-29), IFN- λ 2 (IL-28A), IFN- λ 3 (IL-28B)
TNF superfamily	TNF- α , TNF- β , BAFF, APRIL

distantes de la cellule sécrétrice (51, 63). Ces petites molécules se divisent en trois sous-catégories : (1) les facteurs de croissance hématopoïétique qui vont enclencher la différenciation et la prolifération des cellules souches, (2) les interférons qui vont perturber la réplication virale et (3) les chimiokines qui ont une activité chimio-attractante (**Tableaux 1.2, 1.3, 1.4 et 1.5**). Ces trois types de cytokines jouent un rôle essentiel dans le bon fonctionnement du système immunitaire et la sécrétion de chimiokines va généralement être stimulée par les deux autres catégories de cytokines et ainsi enclencher la migration cellulaire vers le site d'inflammation. Les deux premières catégories de cytokine sont induites par l'activation de la cellule en réponse à la reconnaissance d'un antigène ou à la stimulation par un chimio-attracteur. Les cytokines sont aussi impliquées dans de nombreux processus cellulaires (prolifération, différenciation, maturation, et apoptose cellulaire) telle que la migration des cellules B et T vers les organes secondaires où ils finiront leur différenciation cellulaire. Par exemple, les cytokines vont permettre la différenciation des monocytes en macrophages M1 ou M2 qui sont respectivement pro- et anti-inflammatoires (**Figure 1.12**).

Tableau 1.2. Familles de cytokine. Tiré de (63).

Tableau 1.3. Activités sélectionnées pour la majorité des cytokines. Tiré de (64).

Cytokine	Abbreviation	Target cells	Cytokine activity	Reference
Interferon- α/β	IFN- α/β	All NK cells	Antiviral, antiproliferative Activation	8
Interferon- γ	IFN- γ	Mononuclear phagocytes, endothelial cells, NK cells	Activation, up-regulation of MHC I and II expression	10
Interferon- ω	IFN- ω	Trophoblast	Activation, conceptus maintenance	33
Tumour necrosis factor- α/β	TNF- α/β	Hypothalamus Liver cells Neutrophils, endothelial cells, thymocytes Muscle, fat	Fever Acute-phase protein release Activation, proliferation, co-factor immune stimulation	72
Interleukin-1- α/β	IL-1 α/β	Liver cells Endothelial cells, lymphocytes Muscle, fat	Catabolism (cachexia) Acute-phase protein release Activation, proliferation, co-factor immune stimulation	18
Interleukin-2	IL-2	T cells B cells, NK cells	Growth factor, cytokine production Growth, activation	67
Interleukin-3	IL-3	Immature progenitor cells	Multicolony-stimulating factor, haematopoietic growth factor	
Interleukin-4	IL-4	B cells Mononuclear phagocytes	Th2 cytokine stimulant, promotion of IgG and IgE secretion Activation inhibition	23, 29
Interleukin-5	IL-5	Eosinophils B cells	Activation Growth, differentiation, IgM secretion, activation for IgG secretion	58
Interleukin-6	IL-6	Thymocyte Liver cells Mature B cells Haematopoietic progenitors	Co-factor immune stimulation Acute-phase protein synthesis Growth factor Proliferation and differentiation	66, 69
Interleukin-7	IL-7	Immature progenitor cells	Growth and differentiation to mature B cells or T cells	
Interleukin-8	IL-8	Monocytes, lymphocytes Progenitor cells	Chemotactic factor Growth modulation	46
Interleukin-9	IL-9	Th2 T cells Erythroid cells, mast cells	Growth support Co-stimulatory signal with other cytokines	
Interleukin-10	IL-10	Th1 cells Monocytes, neutrophils Mast cells, eosinophils	Th1 cytokine synthesis inhibition Cytokine synthesis inhibition Inhibition of differentiation	38, 53
Interleukin-11	IL-11	Progenitor cells	Stimulation of megakaryopoiesis, enhancement of macrophage development	
Interleukin-12	IL-12	Macrophages NK cells T cells	Induction of IFN- γ production Induction of IFN- γ production, activation Stimulation of type 1 commitment	69
Interleukin-13	IL-13	Premyeloid cells Monocytes, B cells Activated monocytes	Induction of proliferation Induction of activation and proliferation Inhibition of inflammatory cytokine production	
Interleukin-14	IL-14	B cells	Growth and expansion, inhibition of Ig secretion	
Interleukin-15	IL-15	T cells Multiple cell targets	Stimulation of proliferation Proliferation	
Interleukin-16	IL-16	CD4 ⁺ T cells CD4 ⁺ monocytes CD4 ⁺ eosinophils	Chemotaxis, activation Chemotaxis, activation Potent chemotaxis, activation	16
Interleukin-17	IL-17	Stromal cells Cartilage Fibroblast	Induction of IL-6, IL-8 secretion Augmentation of nitric oxide production Enhancement of ICAM-1 expression	61
Interleukin-18	IL-18	T cells NK cells	Induction of IFN- γ production Induction of IFN- γ production and NK activity	71
Transforming growth factors	TGF- β 1 - TGF- β 5	Macrophages, endothelial cells B cells, T cells, NK cells Angiogenesis	Activation Inhibition of activity Stimulation of extracellular matrix	55
Granulocyte-macrophage CSF	GM-CSF	Bone marrow progenitor cells	Growth and differentiation to granulocytes and macrophages	63
Monocyte CSF	M-CSF	Bone marrow progenitor cells	Growth and differentiation to macrophages	51
Granulocyte CSF	G-CSF	Bone marrow progenitor cells	Growth and differentiation to granulocytes	51
c-kit ligand		Bone marrow stem cells	Activation, increase in responsiveness of stem cells to CSFs	

a) Summarised from Abbas *et al.* (1), Aggrawal and Puri (4), Paul (47), Sigal and Ron (60), and sources noted in the Table

NK : natural killer

MHC : major histocompatibility complex

Th1 : T helper 1

Th2 : T helper 2

Ig : immunoglobulin

CD4⁺ : cluster of differentiation antigen 4⁺

ICAM : intercellular adhesion molecule

CSF : colony-stimulating factor

Tableau 1.4. Chimiokines. Tiré de (65).

Chemokine	Other names	Receptor	Key/main immune function
CXCL1	GRO- α , MGSA, mouse KC	CXCR2	Neutrophil trafficking
CXCL2	GRO- β , MIP-2 α , mouse MIP2	CXCR2	Neutrophil trafficking
CXCL3	GRO- γ , MIP-2 β	CXCR2	Neutrophil trafficking
CXCL4	PF4	?	Procoagulant
CXCL5	ENA-78, mouse LIX	CXCR2	Neutrophil trafficking
CXCL6	GCP-2	CXCR1, CXCR2	Neutrophil trafficking
CXCL7	NAP-2	CXCR2	Neutrophil trafficking
CXCL8	IL-8 (no mouse)	CXCR1, CXCR2	Neutrophil trafficking
CXCL9	Mig	CXCR3	Th1 response; Th1, CD8, NK trafficking
CXCL10	IP-10	CXCR3	Th1 response; Th1, CD8, NK trafficking
CXCL11	I-TAC	CXCR3	Th1 response; Th1, CD8, NK trafficking
CXCL12	SDF-1	CXCR4	Bone-marrow homing
CXCL13	BLC, BCA-1	CXCR5	B-cell and T _{FH} -positioning LN
CXCL14	BRAK	?	Macrophage skin homing (human)
Cxcl15	Lungkine (mouse only)	?	?
CXCL16		CXCR6	NKT and ILC migration and survival
CXCL17		?	Macrophage and DC chemotaxis
CCL1	I-309, mouse TCA3	CCR8	Th2 cell and Treg trafficking
CCL2	MCP-1, mouse JE	CCR2	Inflammatory monocyte trafficking
CCL3	MIP-1 α	CCR1, CCR5	Macrophage and NK-cell migration; T-cell–DC interactions
CCL4	MIP-1 β	CCR5	Macrophage and NK-cell migration; T-cell–DC interactions
CCL5	RANTES	CCR1, CCR3, CCR5	Macrophage and NK-cell migration; T-cell–DC interactions
Ccl6	C-10, MRP-1 (mouse only)	Unknown	?
CCL7	MCP-3, mouse Fic or MARC	CCR2, CCR3	Monocyte mobilization
CCL8	MCP-2	CCR1, CCR2, CCR3, CCR5 (human); CCR8 (mouse)	Th2 response, skin homing mouse
Ccl9/10	MIP-1 γ , MRP-2 (mouse only)	Unknown	?
CCL11	Eotaxin-1	CCR3	Eosinophil and basophil migration
Ccl12	MCP-5 (mouse only)	CCR2	Inflammatory monocyte trafficking
CCL13	MCP-4	CCR2, CCR3, CCR5	Th2 responses
CCL14	HCC-1	CCR1	?
CCL15	Leukotactin-1, HCC-2, MIP-5	CCR1, CCR3	?
CCL16	HCC-4, NCC-4, LEC	CCR1, CCR2, CCR5	?
CCL17	TARC	CCR4	Th2 responses, Th2-cell migration, Treg, lung, and skin homing
CCL18	PARC, DC-CK1	CCR8	Th2 response, marker AAM, skin homing
CCL19	ELC, MIP-3 β	CCR7	T-cell and DC homing to LN
CCL20	MIP-3 α , LARC	CCR6	Th17 responses, B-cell, and DC homing to gut-associated lymphoid tissue
CCL21	SLC, 6CKine	CCR6, CCR7	T-cell and DC homing to LN
CCL22	MDC	CCR4	Th2 response, Th2-cell migration, Treg migration
CCL23	MPIF-1, MIP-3	Unknown	?
CCL24	Eotaxin-2, MPIF-2	CCR3	Eosinophil and basophil migration
CCL25	TECK	CCR9	T-cell homing to gut, thymocyte migration

Tableau 1.5. Chimiokines (suite). Tiré de (65).

Chemokine	Other names	Receptor	Key/main immune function
CCL26	Eotaxin-3	CCR3, CX3CR1	Eosinophil and basophil migration
CCL27	CTAK	CCR10	T-cell homing to skin
CCL28	MEC	CCR3, CCR10	T-cell and IgA plasma–cell homing to mucosa
XCL1	Lymphotactin, SCM-1 α	XCR1	Cross presentation by CD8 ⁺ DCs
XCL2	SCM-1 β	XCR1	Cross presentation by CD8 ⁺ DCs
CX3CL1	Fractalkine	CX3CR1	NK, monocyte, and T-cell migration

This table was created from data adapted from Bachelerie et al. 2013.

AAM, alternatively activated macrophage; ILC, innate lymphoid cell; LN, lymph node; NK, natural killer; NKT, natural killer T; T_{FH}, follicular helper T cell; Treg, regulatory T cell.

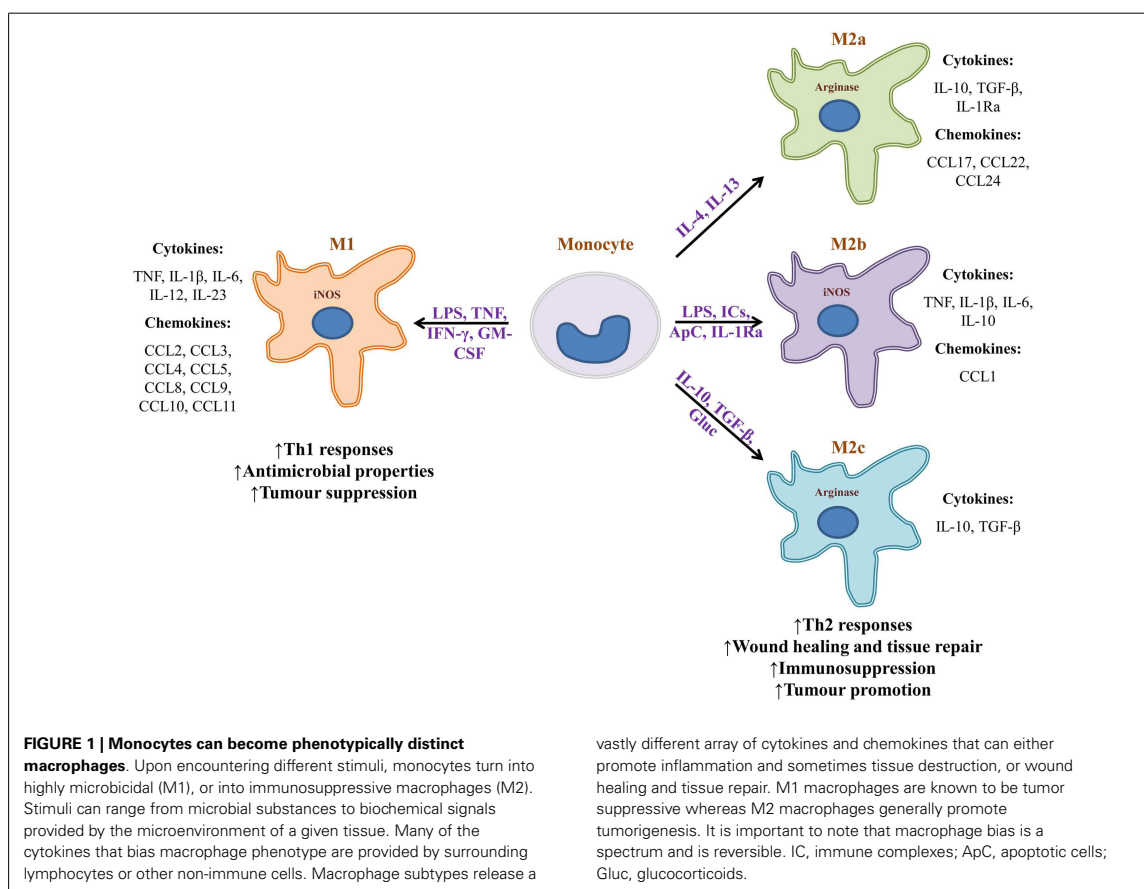


Figure 1.12. Différenciation des monocytes en sous-groupe de population de macrophages. Tiré de (66).

Tableau 1.6. Récepteurs des chimiokines. Tiré de (65).

Receptor	Immune cell expression	Key immune function
G-protein-coupled chemokine receptors		
CXCR1	N>Mo, NK, MC, Ba, CD8 T _{EFF}	Neutrophil trafficking
CXCR2	N>Mo, NK, MC, Ba, CD8 T	B-cell lymphopoiesis Neutrophil egress from bone marrow Neutrophil trafficking
CXCR3	Th1, CD8 T _{CM} and T _{EM} , NK, NKT, pDC, B, Treg, T _{FH}	Th1-type adaptive immunity
CXCR4	Most (if not all) leukocytes	Hematopoiesis Organogenesis Bone marrow homing
CXCR5	B, T _{FH} , T _{FR} , CD8 T _{EM}	B- and T-cell trafficking in lymphoid tissue to B-cell zone/follicles
CXCR6	Th1, Th17, $\gamma\delta$ T, iLC, NKT, NK, PC	Innate lymphoid cell function Adaptive immunity
CCR1	Mo, M Φ , N, Th1, Ba, DC	Innate immunity Adaptive immunity
CCR2	Mo, M Φ , Th1, iDC, Ba, NK	Monocyte trafficking Th1-type adaptive immunity
CCR3	Eo>Ba, MC	Th2-type adaptive immunity Eosinophil distribution and trafficking
CCR4	Th2, skin- and lung-homing T, Treg>Th17, CD8 T, Mo, B, iDC	Homing of T cells to skin and lung Th2-type immune response
CCR5	Mo, M Φ , Th1, NK, Treg, CD8 T, DC, N	Type-1 adaptive immunity
CCR6	Th17>iDC, $\gamma\delta$ T, NKT, NK, Treg, T _{FH}	iDC trafficking, GALT development Th17 adaptive immune responses
CCR7	T _N , T _{CM} , T _{RCM} , mDC, B	mDC, and B- and T-cell trafficking in lymphoid tissue to T-cell zone Egress of DC and T cells from tissue
CCR8	Th2, Treg, skin T _{RM} , $\gamma\delta$ T, Mo, M Φ	Immune surveillance in skin type-2 adaptive immunity, thymopoiesis
CCR9	Gut homing T, thymocytes, B, DC, pDC	Homing of T cells to gut GALT development and function, thymopoiesis
CCR10	Skin-homing T, IgA plasmablasts	Humoral immunity at mucosal sites Immune surveillance in skin
XCR1	Cross-presenting CD8 ⁺ DC, thymic DC	Ag cross-presentation by CD8 ⁺ DCs
CX3CR1	Resident Mo, M Φ , MG, Th1, CD8 T _{EM} , NK, $\gamma\delta$ T, DC	Patrolling monocytes in innate immunity Microglial-cell and NK-cell migration type-1 adaptive immunity
Atypical chemokine receptors (non-G-protein-coupled signaling)		
ACKR1 (DARC; Duffy)	RBC, LEC	Chemokine transcytosis Chemokine scavenging
ACKR2 (D6)	LEC, DC, B	Chemokine scavenging
ACKR3 (CXCR7)	Stromal cells, B	Shaping chemokine gradients for CXCR4
ACKR4 (CCRL1; CCX-CKR)	Thymic epithelium	Chemokine scavenging

Data in table is modified from Bachelierie et al. 2013.

B, B cell; Ba, basophil; Eo, eosinophils; GALT, gut-associated lymphoid tissue; iDC, immature DC; LEC, lymphatic endothelium; M Φ , macrophage; MG, microglia; Mo, monocyte; N, neutrophil; pDC, plasmacytoid DC; PC, plasma cell; RBC, red blood cell; T_{CM}, central memory T cell; T_{EFF}, effector T cell; T_{FH}, follicular helper T cell; T_{FR}, follicular regulatory T cell; T_N, naïve T cell; T_{RCM}, recirculating memory T cell.

Les chimiokines se divisent en quatre sous-groupes définis par la position de leurs résidus cystéiques : les chimiokines α (CXC-), les chimiokines β (CC-), les chimiokines δ (CX3C-) et les

chimiokines γ (-C-) (**Tableaux 1.4 et 1.5**) (65). Certaines chimiokines ont la capacité de se lier à plusieurs récepteurs présents à la surface des cellules immunitaires (**Tableau 1.6**). Pour permettre aux cellules d'identifier la direction à prendre pour rejoindre le site d'inflammation, les chimiokines sécrétées forment un gradient où la concentration la plus élevée de chimiokines se trouve au niveau du site d'inflammation. De plus, ces chimiokines vont aussi augmenter l'expression des molécules d'adhérence cellulaire à la surface des cellules immunitaires qu'elles auront stimulé ainsi qu'à la surface des cellules épithéliales présentes tout le long de ce gradient. Cela va favoriser le déplacement des cellules immunitaires le long de la paroi des capillaires sanguins.

Les macrophages sont les premières cellules à répondre à la reconnaissance d'un antigène ce qui leur permet d'être responsables de l'initiation de la réponse immunitaire innée. Ils sécrètent de nombreuses cytokines afin de recruter un maximum de cellules immunitaires sur le site d'inflammation (66). Ces cytokines sécrétées par les macrophages se divisent en trois catégories distinctes : les cytokines pro-inflammatoires, les cytokines anti-inflammatoires et les chimiokines. Les cytokines pro-inflammatoires (TNF α , IL1 β , IL1 α , IL6, IL12, IL18, IL23 et IL27) sont aussi sécrétées par les monocytes, les lymphocytes actifs, les cellules endothéliales et les fibroblastes, et un grand nombre d'entre elles, vont initier la sécrétion de chimiokines par les macrophages. Elles vont aussi permettre le recrutement de nombreuses cellules immunitaires tels que les neutrophiles et monocytes et provoquer la différenciation des cellules T ainsi que l'activation des cellules T et NK. Par exemple, lorsque TNF α est sécrété, en concert avec IL17, cela déclenche l'expression par les macrophages de trois chimiokines-attractrices (CXCL1, CXCL2 et CXCL5) qui vont stimuler les neutrophiles ayant à leur surface le récepteur CXCR2. Les cytokines anti-inflammatoires (IL10 et TGF- β) sont aussi sécrétées par les cellules B et T. Elles vont réduire l'expression par les macrophages des cytokines pro-inflammatoires (IL10 agit sur TNF α , IL1 β , IL6, IL8, IL12 et GM-CSF; TGF- β agit sur TNF, IL1 β , IL2 et IL12) dont TGF- β qui joue un rôle majeur dans la régénération tissulaire, la prolifération et la différenciation cellulaire. IL10 inhibe la présentation d'antigènes et supprime l'expression du complexe majeur d'histocompatibilité (CMH) de classe II, ainsi que la production de IFN- γ par les cellules T et NK et la capacité des macrophages à répondre à sa stimulation. En parallèle, elle induit la croissance, la différenciation et la sécrétion des immunoglobulines G (IgG) par les cellules B. L'étude de la cytokine IL6 dans les macrophages a montré qu'elle peut être pro-inflammatoire ou anti-inflammatoire par ses voies trans ou classiques, respectivement (66). Pour finir, le groupe de chimiokines sécrétées par les macrophages (CXCL1, CXCL2 (alias MIP-2 α), CCL5 (alias RANTES), CXCL8 (alias IL8), CXCL9 (alias MIG), CXCL10 (alias IP-10) et CXCL11 (alias IP-9)). CXCL1 et CXCL2 permettent le recrutement des neutrophiles, des cellules T et NK ainsi que l'activation des cellules NK parmi les CHAK (*CC-chemokine-activated killers*), qui sont des cellules cytolytiques (67). La chimiokine CXCL8 induit la dégranulation et des changements morphologiques chez les neutrophiles et les macrophages sont les premières cellules à sécréter cette chimiokine. CXCL10 est la seule chimiokine induite par l'interféron- γ et elle recrute les cellules T et NK.

Les neutrophiles sont les premières cellules immunitaires à arriver sur le site d'inflammation en réponse à un chimio-attracteur. Leur principal rôle est de phagocyter les corps étrangers pour permettre leur dégradation. Les neutrophiles expriment à leur surface les récepteurs de

chimiokines CCR1, CXCR1, CXCR2 et CCR5 (**Tableau 1.6**). La sécrétion de chimiokines par les macrophages entraîne leur migration vers le site d'inflammation. Tout le long de leur migration, les neutrophiles libèrent plusieurs chimiokines (CCL3, CCL4, CCL5, CCL20, CXCL1, CXCL8, CXCL9, and CXCL10) qui vont alimenter le gradient de migration et ainsi permettre le recrutement de neutrophiles, macrophages et monocytes (**Tableaux 1.4 et 1.5**) (65, 68-70). Une fois sur le site d'inflammation, les neutrophiles vont libérer CXCL2 et IL1 β qui vont stimuler les macrophages et maintenir leur production de chimiokines. Les neutrophiles sécrètent des chimio-attracteurs qui permettent le recrutement de l'ensemble des cellules immunitaires impliquées dans les réponses innée et adaptative (**Figure 1.13**).

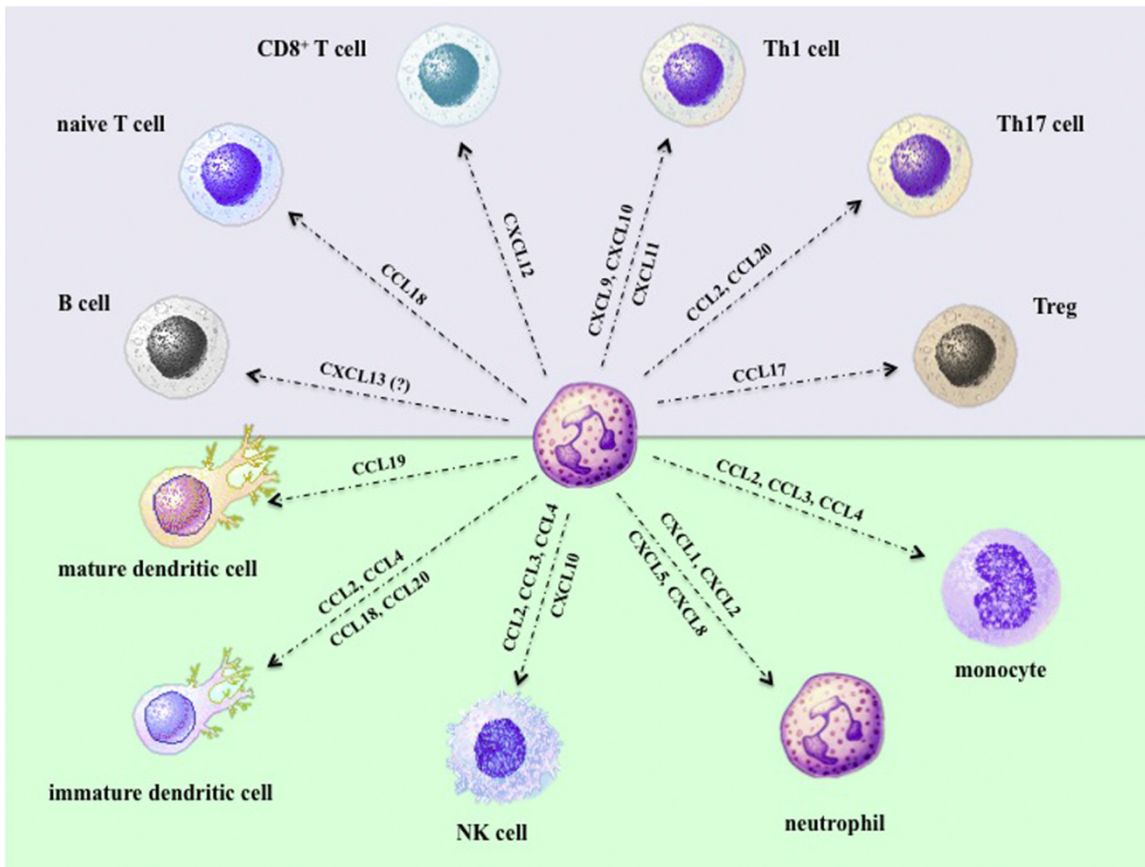


Figure 1.13. Schématisation de l'expression et de la production des chimiokines par les neutrophiles impactant les cellules immunitaires de la réponse innée (fond vert) et adaptative (fond violet-gris). Tiré de (69).

Les monocytes ont des propriétés similaires aux macrophages dans la migration cellulaire. Après avoir été stimulés par les chimio-attracteurs CCL2 et CX3CL1, les monocytes vont atteindre le site d'inflammation situé dans un tissu ou un vaisseau sanguin. Une fois sur le site d'inflammation, les monocytes sécrètent de nombreuses cytokines et chimiokines pour recruter un maximum de cellules immunitaires. Parmi les cytokines exprimées se trouve IL1, IL6, IL8, IL10, IL12, IL15, IL18 et TNF α qui sont des cytokines pro et anti-inflammatoires (65, 68, 70, 71). Certaines chimiokines sont sécrétées par les monocytes tels que CCL5 qui est induite par IFN γ et TNF α (71). Les

monocytes expriment de nombreux récepteurs des cytokines et chimiokines à leur surface (CCR2, CXCR2, CD47, CCR1, CCR5, CCR6 et CX3CR1) dont CXCR2 qui a pour ligand CXCL8, mais sa stimulation ne déclenche pas de migration chez les monocytes, contrairement aux neutrophiles (72).

Les cellules lymphoïdes sont impliquées dans la réponse immunitaire adaptative. De ce fait, elles répondent majoritairement aux chimio-attracteurs et sécrètent peu de cytokines et chimiokines. Les deux principaux types cellulaires qui produisent et libèrent des cytokines sont les cellules T CD4+ et $\gamma\delta$. Les cellules T CD4+ expriment CCL3, CCL4 et CCL17 après avoir reconnues un antigène afin de stimuler les cellules T CD8+ et les cellules B (72). Les cellules T $\gamma\delta$, quant à elles, sécrètent IL4, IL5, IFN γ , TNF α , CCL5, CCL3 et CCL4 (59, 73). CCL3 et CCL4 sont connues pour recruter et activer respectivement les cellules T CD4+ et CD8+, mais il n'a pas été prouvé que les cellules T $\gamma\delta$ agissent sur ces cellules par la sécrétion de ces deux chimiokines. En contrepartie, les cellules lymphoïdes expriment de nombreux récepteurs de cytokines et chimiokines à leur surface (65, 70, 74). Ces récepteurs permettent la stimulation des cellules lymphoïdes, ce qui va enclencher la différenciation et l'activation cellulaire en plus de leur migration. Les cellules T $\gamma\delta$ expriment IL12R et les cellules B expriment CXCR5, CCR7 et CXCR4 à leur surface (59, 70, 73). Les cellules lymphoïdes qui possèdent le plus de récepteurs à leur surface sont les cellules T CD4+ et CD8+ soit environ une dizaine chacun (**Tableau 1.6**) (70).

La transduction du signal par les récepteurs des cytokines est induite malgré l'absence de domaines cytoplasmiques, responsable de l'activité tyrosine kinase intrinsèque de la cellule. L'une des voies de signalisation déclenchée par la liaison d'une cytokine sur son récepteur implique deux familles de protéines, JAK et STAT (74, 75). Les membres de la famille JAK sont des protéines tyrosines kinases qui lient spécifiquement chacune des deux branches du récepteur de cytokine ce qui va induire la phosphorylation de ces deux protéines ainsi que des deux branches du récepteur (**Figure 1.14**). Par exemple, les deux sous-unités du récepteur de IL12 lient JAK2 et TYK2 respectivement et ceux du récepteur d'IL4, JAK1 et JAK3 (**Figure 1.14**). JAK2 est connue pour être impliquée dans la réponse au GM-CSF, G-CSF, IL12, IL6 et IL3. JAK1 et JAK3 sont présentes dans la voie de signalisation induite par IL2, IL4 et toutes les cytokines qui lient un récepteur de la famille γc . La phosphorylation des membres de la famille JAK confère au complexe récepteur-JAK une activité de catalyseur qui va activer un membre spécifique de la famille STAT. Ces protéines sont des FT et leur phosphorylation déclenche leur migration vers le noyau puis leur liaison sur les régions régulatrices des gènes répondant à la cytokine de départ (**Figure 1.14**). Pour les interférons- α et - γ , la protéine JAK phosphoryle un complexe de protéines composé de STAT1 α ou STAT1 β , STAT2 et p48, qui est une protéine non-STAT. STAT1 α est impliquée dans les cascades induites par EGF, PGDF, M-CSF, IL6, IL11 et les interférons. IL6, IL12 et IL4 induisent respectivement la réponse par STAT3, STAT4 et STAT6. La cascade biochimique déclenchée par IL4 est dépendante de STAT6. *SOCS1* est un inhibiteur de *STAT6* qui est aussi un gène cible induit par la liaison de IL4 sur son récepteur (76). Ce mécanisme permet l'autorégulation de la stimulation par IL4.

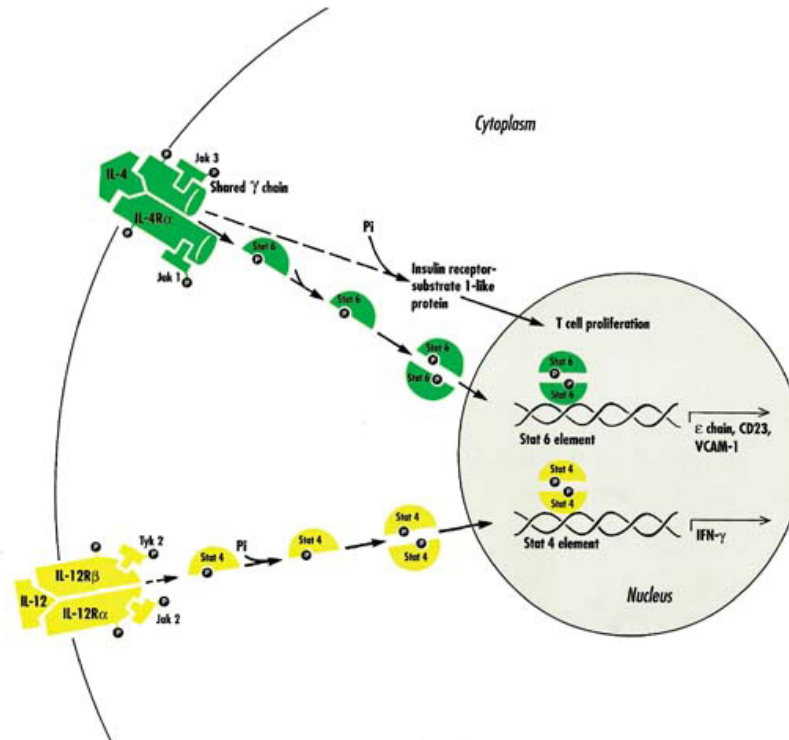


Figure 1.14. Schéma des voies de signalisation intracellulaire de la transduction du signal induit par IL4 et IL12. Tiré de (74)

1.2.3 Identification immunophénotypique : Classe de différenciation

Le système immunitaire est constitué d'un grand nombre de types cellulaires pouvant être identifiés selon leurs morphologies ou leurs fonctions. Mais ces approches d'identification demandent la mise en place de protocoles complexes et fastidieux qui, dans certain cas, ne permettent pas d'identifier les sous types cellulaires d'une population. De plus, ces protocoles peuvent altérer l'intégrité de la cellule qui ne pourra donc pas être utilisée dans une étude.

La surface des cellules immunitaires est constituée de plusieurs molécules dont certaines sont utilisées pour identifier les différents types cellulaires. En 1982, la commission *Human Leukocyte Differentiation Antigens* (HLDA) regroupa des laboratoires pour identifier ces molécules de surface et définir leur fonction. Pour nommer ces molécules de surfaces, la HLDA utilise la nomenclature suivante pour construire le symbole de ces molécules: classe de différenciation (CD) suivi d'un nombre (77, 78). La numérotation dans le symbole représente le groupe ou cluster de molécules de surface qui sont reconnues par un unique anticorps monoclonal.

Des éléments peuvent être ajoutés à la racine du symbole et apporter des informations supplémentaires sur le groupe de molécules de surface. Lorsque l'annotation est considérée insuffisante par le groupe HLDA, un « w » est ajouté entre le CD et nombre. Les lettres majuscules après le nombre du symbole informe que ce groupe de molécules de surface est composé de sous-groupes qui se différencient par la variation de leurs domaines extracellulaires, obtenu par l'épissage (ex : le groupe CD45 possède deux variants: CD45RA et CD45RO). En opposition, les lettres minuscules après le nombre démontre qu'il existe plusieurs molécules de surface qui ont en commun une chaîne protéique pour former un dimère (ex : CD1a/b/c/d/e avec la microglobuline $\beta 2$ ou CD11a/b/c avec CD18). La lettre minuscule permet aussi d'identifier les

membres d'une même famille de gènes (ex : *CD66a-f*) ou une modification dans la structure de la séquence carbohydrate (ex : CD15s ou CD60b).

Un autre type d'information peut être ajouté au symbole, représentant le niveau d'expression du groupe de molécules à la surface de la cellule. Les symboles « + » et « - » déterminent respectivement la présence et l'absence du cluster à la surface cellulaire alors que les exposants « low » et « high » représentent la quantité relative à la surface de la cellule.

Le groupe HLDA a pour objectif d'identifier la majorité des classes de différenciation à la surface des cellules immunitaires et ainsi, mieux comprendre les propriétés fonctionnelles, la différenciation, la maturation et l'activation des cellules immunitaires. Actuellement, cet organisme a identifié plus de 400 classes de différenciation et ont créé un site web contenant la liste de CD et des annotations telles que la structure de la molécule, ses ligands, son expression et sa fonction, quand ils sont connus (<http://www.hcdm.org/>).

La bonne connaissance de la présence des CD à la surface des cellules immunitaires est un atout pour leur étude (79-83). Comme certains de ces groupes de molécules sont présents à la surface de plusieurs types cellulaires immunitaires, la combinaison de présence, absence et niveau d'expression de ces classes de différenciation est utilisée pour définir un type cellulaire immunitaire donné (**Figure 1.15 et Tableau 1.**). Grâce à la cytométrie de flux, le tri cellulaire est très efficace et dans certains cas, il est possible d'obtenir une pureté de plus de 99% (84). La cytométrie de flux permet l'analyse de cellule unique et transmet des informations sur la taille et la granularité intracellulaire. Cet outil permet également d'identifier les cellules en utilisant des anticorps monoclonaux des CD présents à la surface cellulaire (85). La combinaison de ces deux techniques est utilisée pour réaliser le tri cellulaire des cellules immunitaires. Par exemple, il est difficile de différencier les monocytes des cellules B et T car ils ont une morphologie semblable, mais les monocytes possèdent à leur surface le CD14, contrairement aux cellules B et T (**Figure 1.15**). Pour les cellules B, il suffit d'identifier les cellules qui sont CD19+CD3-. Pour les neutrophiles, leur morphologie joue un rôle important dans leur identification mais malgré qu'il possède plus de 140 CD, tous sous-types confondus, le marqueur de surface CD15 permet de bien caractériser cette population ainsi que CD11b, CD14, CD16 ou CD62L (82). Ainsi la combinaison de CD+/- et CD^{low/high} permettent de réaliser le tri cellulaire des populations de cellules immunitaires mais aussi des sous-types cellulaires au sein d'une population car la différenciation, la maturation et l'activation cellulaire sont accompagnés d'une distribution unique de molécules de surface.

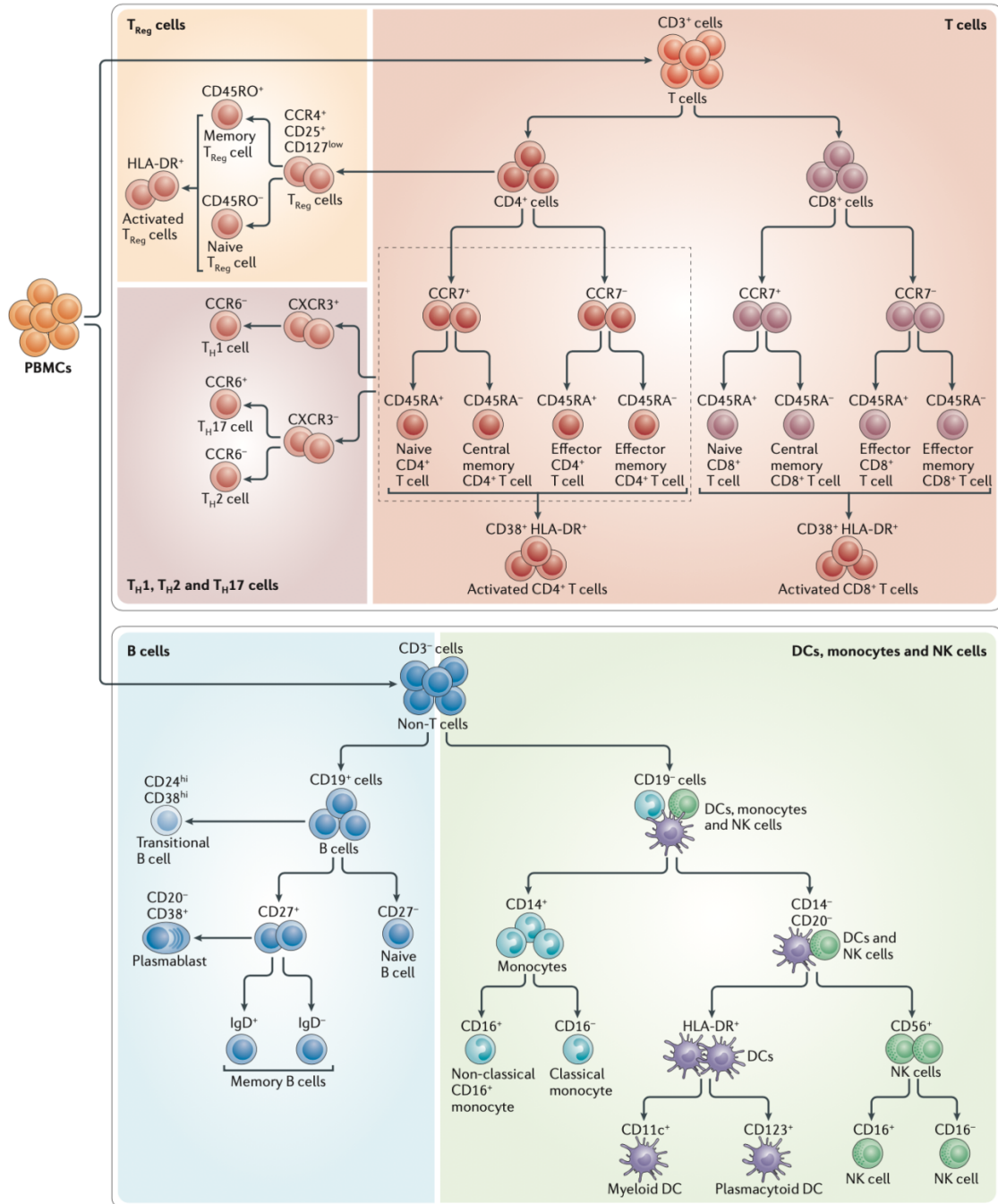


Figure 1.15. Identification des sous-types de cellules immunitaires par huit couleurs de coloration d'anticorps. Tiré de (80).

Tableau 1.7. Marqueurs CD communément utilisés dans la biologie des neutrophiles. Tiré de (82).

Antibody	Other names	Distribution
CD10	Common acute lymphoblastic leukemia antigen (CALLA), Enkephalinase, Neutral endopeptidase, Nephrylsin	B cell precursors, T cell precursors, neutrophils
CD11a CD11B	Integrin alpha L chain αM integrin, Mac-1, CR3, and C3biR	All leukocytes Granulocytes, monocytes/macrophages, dendritic cells, NK cells, subset of T cells, subset of B cells
CD11C	Integrin α X subunit, CR4, p150	Myeloid, dendritic cells, NK cells, B and T cell subsets
CD13	Aminopeptidase N, APN, gp150	Granulocytes, monocytes, myeloid progenitors, endothelial and epithelial cells, granular lymphocyte subset
CD14	LPS receptor	Monocytes, macrophages, granulocytes (low) Some dendritic cell populations such as interfollicular dendritic cells, reticular dendritic cells, and Langerhans cells have also been reported to express CD14.
CD15	Lewis X, 3-FAL, 3-FL, LNFP III, Lex, SSEA-1, X-hapten	Neutrophils, eosinophils, monocytes
CD16	Fc γ R III/II, Ly-17	B cells, monocyte/macrophages, NK cells, neutrophils, mast cells, dendritic cells
CD16b	glycosyl phosphatidyl inositol-anchored (GPI) protein that acts as a receptor for the Fc region of immunoglobulin gamma (Fc gamma RIII)	Neutrophils and activated eosinophils
CD18		All leukocytes
CD32	Fc γ R III/II, Ly-17	B cells, monocyte/macrophages, NK cells, neutrophils, mast cells, dendritic cells
CD31	PECAM-1, EndoCAM	Monocytes, platelets, granulocytes, endothelial cells, lymphocyte subset
CD33	Siglec-3, gp67, p67	Myeloid progenitor, monocytes, granulocytes, dendritic cells, mast cells
CD55	Decay accelerating factor	Hematopoietic lineage
CD62L	L-selectin, LECAM-1, LAM-1, Leu-8, MEL-14, TQ-1	Majority of B cells, naïve T cells, subset of memory T and NK cells, monocytes, granulocytes, thymocytes
CD64	Fc γ RI, FcR I	Monocytes, macrophages, dendritic cells, activated granulocytes
CD66b	CD67, CGM6, NCA-95	Granulocytes
CD66c	Carcinoembryonic antigen family	Granulocytes
CD88	C5a ligand, C5a anaphylatoxin receptor, C5AR, C5R1, C5AR1	Monocytes, neutrophils and eosinophils. There has been reports of CD88 expression in non-immune cells such as glial cells, cerebellar granule cells, cardiomyocytes and vascular endothelial cells
CD114	CSF3 R, HG-CSFR, Colony Stimulating Factor 3 receptor	Granulocytes (in all stages of differentiation), monocytes, platelets, endothelial cells, placenta and trophoblasts.
CD170	Siglec-5	Granulocytes, monocytes/macrophages, subsets of lymphocytes, and a subset of activated dendritic cells
CD172		Monocytes/macrophages, granulocytes, dendritic cells and neurons
CD182	CD182, CXCR2, CD128b, IL-8RB, Type II IL-8 receptor	Granulocytes, NK cells, subset of T cells, monocytes, endothelial cells, oligodendrocytes
CD181	CD181, CXCR1, CDw128, CD128a, IL-8RA, Type I IL-8 receptor	Granulocytes, NK cells, subset of T cells, monocytes, endothelial cells, oligodendrocytes
CD282	Toll like receptor 2, TIL4, CD282	Monocytes, granulocytes, endothelial and epithelial cells

1.2.4 Fonctions partagées entre les cellules immunitaires primaires : Voies de signalisation et FT

1.2.4.1 Présentation d'antigènes

Le complexe majeur d'histocompatibilité (CMH) a pour fonction principale la présentation d'antigènes (Ag) aux cellules du système immunitaire, qui est essentielle pour la régulation de la réponse immunitaire. Il existe deux types de CMH, classe I et classe II présentant des Ag endogènes et exogènes respectivement.

L'ensemble des gènes encodant les molécules formant les protéines présentatrices d'Ag se trouvent dans le bras court du chromosome 6 (6p21) s'étendant sur environ 4 kilobases. Ces gènes se regroupent en trois régions distinctes et s'organisent de la façon suivante sur le brin d'ADN : la région de classe I, la région de classe III et la région de classe II. La région de classe I contient trois gènes *HLA* (*HLA-A*, *HLA-B* et *HLA-C*) qui encodent tous la chaîne lourde, appelée aussi chaîne α , de la molécule du CMH de classe I. La chaîne lourde polymorphique de type A, B ou C se lie à la microglobuline β_2 qui correspond à la chaîne légère ou chaîne β de la molécule et son gène se trouve sur le chromosome 15. Les molécules du CMH de classe I sont exprimées à la surface de toutes les cellules nucléaires et sont reconnues par le récepteur des cellules T (TCR) des lymphocytes T CD8+ qui seront activés en lymphocytes T cytotoxiques après contact cellulaire. Les lymphocytes NK, quant à eux, détectent la diminution d'expression de cette

protéine à la surface cellulaire car certains récepteurs des cellules NK sont des ligands des molécules du CMH de classe I (86-88). La région de classe II est constituée de trois familles de gènes *HLA* (*HLA-DR*, *HLA-DQ* et *HLA-DP*) dont chaque famille possède deux gènes, un codant pour la chaîne α et un codant pour la chaîne β de la molécule du CMH de classe II (*DRA* et *DRB1-9*; *DQA* et *DQB*; *DPA* et *DPB*). Il existe donc trois types de molécule du CMH de classe II (*HLA-DR*, *HLA-DQ* et *HLA-DP*). Ces molécules sont exprimées à la surface des lymphocytes B, des cellules présentatrices d'antigènes (monocytes, macrophages et cellules dendritiques) et des cellules de Langerhans, et sont reconnues par le TCR des lymphocytes T CD4+ qui, à leur contact, se différencient en lymphocytes T CD4+ *helpers* ou auxiliaires (86-88). La transcription des gènes des régions de la classe I et de la classe II est régulée respectivement par *NLRC5* et *CIITA*, deux co-activateurs membres de la famille *NLR*. Les facteurs de transcription *CREB*, *NF-Y* et un membre de la famille *RFX* forment un complexe dans la région promotrice des gènes *HLA* et la liaison de *NLRC5* ou *CIITA* à ce complexe démarre l'initiation et l'élongation de la transcription (89-91) (92). La région de classe III est constituée de gènes qui ne codent pas pour des molécules formant la protéine présentatrice d'Ag mais pour des protéines impliquées dans le CMH de classe II, dont le gène *TNF*, l'hydroxylase 21 et des gènes du complément (*C2*, *C4*, *facteur B*) (86, 88, 91). Les gènes *HLA* sont très polymorphiques chez l'humain et, par hérédité mendélienne, chaque individu possède un haplotype de sa mère et de son père. Chaque haplotype possède un exemplaire de chaque gène *HLA*, excepté pour le gène *HLA-DR* qui peut en présenter plusieurs. Cela permet à chaque individu d'avoir au moins deux exemplaires, identiques ou non, de chaque gène *HLA*. En 2016, plus de 10 000 allèles différents de la molécule du CMH de classe I dans l'espèce humaine ont été référés et plus de 20 000, en 2021 ([Statistiques EMBL-EBI](#)) (88). Ces molécules n'ont pas d'interactions spécifiques avec les peptides, ce qui leur permet de présenter près de $6 \times 10^{6-7}$ et 12×10^{10} différents peptides pour les CMH de classe I et de classe II, respectivement. Par exemple, à la surface des cellules B et T se trouvent environ 200 000 molécules du CMH de classe I et 20 000, du CMH de classe II et cela ne permet pas de présenter tous les différents peptides connus chez l'humain. En règle générale, une cellule présente environ 10 000 peptides différents à sa surface (88).

Le CMH se divise en deux classes distinctes par le type de peptides qu'il présente à la surface de la cellule. La seconde distinction est que les molécules du CMH de classe I se trouvent à la surface de toutes les cellules nucléées de l'organisme alors que celle du CMH de classe II sont à la surface de certaines cellules immunitaires dont les cellules présentatrices d'Ag. La molécule du CMH de classe I ne lie que les peptides endogènes produits dans le cytoplasme provenant soit d'Ag du soi ou de protéines tumorales, soit de protéines virales synthétisées dans la cellule. Ces protéines endogènes vont être dégradées dans le cytoplasme par le protéasome en peptides de 8 à 11 a.a. qui seront transportés dans la lumière du réticulum endoplasmique (RE) par le transporteur TAP, constitué des molécules TAP-1 et TAP-2 (**Figure 1.16**) (87, 88). Dans le RE, les chaînes α et β de la molécule du CMH de classe I sont associées ensemble par des protéines chaperonnes telles que la calréticuline, la calnexine et la tapasine. Ces dernières maintiennent la conformation de la molécule du CMH de classe I jusqu'à ce qu'un peptide, présent dans le RE, s'y lie. Cette liaison libère la molécule du CMH de classe I des protéines chaperonnes et initie sa migration vers la membrane plasmique en passant par l'appareil de Golgi (**Figure 1.16**) (87, 88). La molécule de CMH de classe II, elle, interagit avec des peptides exogènes provenant d'agents pathogènes

extracellulaires ou de corps apoptotiques internalisés par endocytose. Le système endo-lysosomal dégrade la protéine en peptides de 13 à 18 a.a. (**Figure 1.16**) (87, 88). Comme pour la molécule du CMH de classe I, la molécule du CMH de classe II est construite dans le RE dont la conformation est maintenue par la chaîne invariante li encodé par le gène *CD74*. Cette chaîne invariante a une région appelée CLIP qui va se lier à la place du peptide sur la molécule du CMH de classe II. Ce complexe migre vers l'appareil de Golgi pour former une vésicule endocytique de classe II. Dans la vésicule, des cathepsines L et S vont dégrader la chaîne invariante sans toucher le fragment CLIP qui bloque le site de liaison peptidique. La molécule HLA-DM, encodée par le gène du même nom et présent dans la région de classe II, va remplacer le fragment CLIP par un peptide exogène présent dans la vésicule et ainsi former le complexe *MIIC*. Ce complexe va par la suite être transporté jusqu'à la membrane plasmique par la vésicule endocytique (**Figure 1.16**) (87, 88). Dans les cellules dendritiques, une voie croisée de présentation a été observée mais n'est pas encore totalement comprise. Les peptides exogènes dégradés dans le système endo-lysosomal se retrouvent dans la lumière du RE et se lie à une molécule de CMH de classe I, au lieu d'une molécule du CMH de classe II. Cette molécule du CMH de classe II associée à un peptide exogène va ensuite migrer jusqu'à la membrane plasmique en passant par l'appareil de Golgi (**Figure 1.16**) (87).

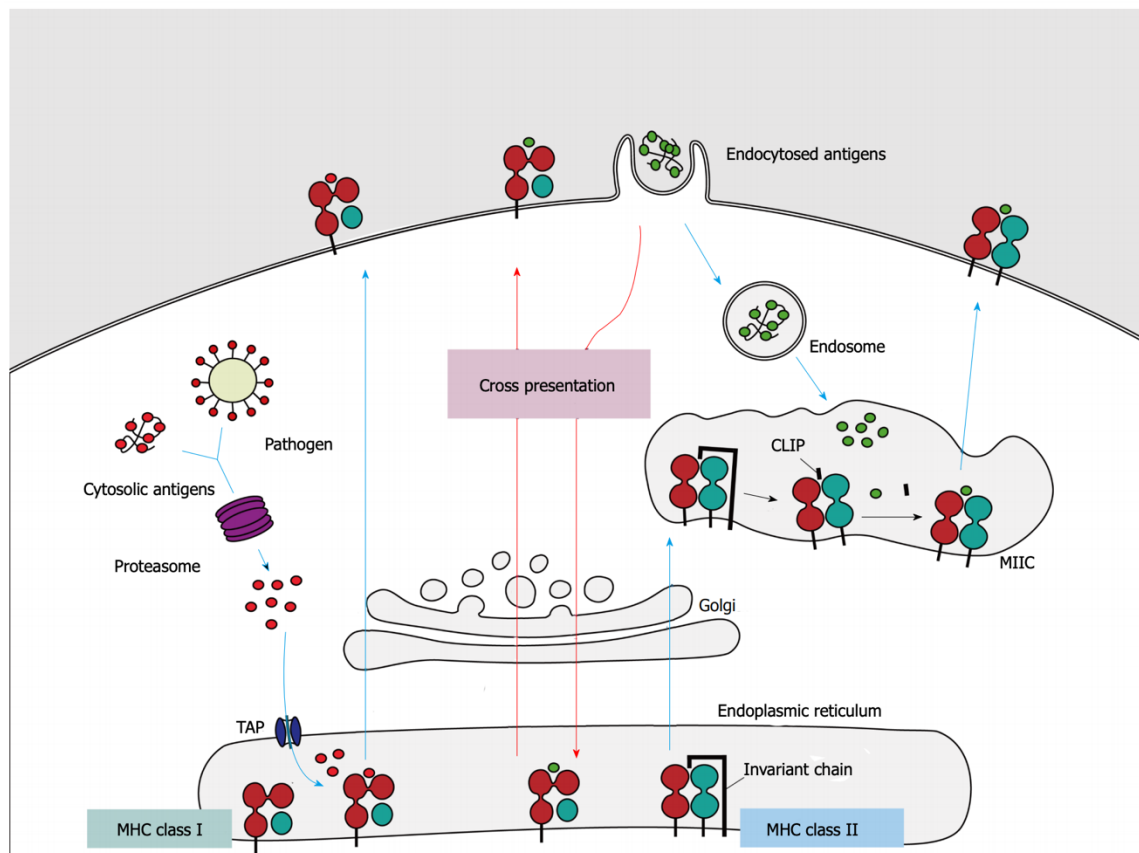


Figure 1.16. Voies du CMH de classes I et II. Tiré de (87).

Des protéines de type non-HLA sont impliquées dans les CMH de classe I et classe II telles que les protéines de la famille MIC (MICA et MICB), CD1 et le complément. Ces protéines ne peuvent pas

remplacer les molécules du CMH mais sont complémentaires. Les gènes *MICA* et *MICB* sont les deux seuls membres de la famille des *MIC* qui codent pour une protéine fonctionnelle et ils se situent sur le chromosome 6, proche du gène *HLA-B* (87, 93). Les protéines MICA et MICB sont présentes à la surface de la membrane plasmique en cas de stress cellulaire dû à une infection virale, des dommages cellulaires ou une transformation cellulaire de type oncogénique, par exemples. Ces protéines sont reconnues par le récepteur NKG2D, encodé par le gène *KLRK1*, présent à la surface des lymphocytes NK, T CD8+ et T $\gamma\delta$ et déclenche une réponse cytotoxique (87, 94). Les gènes *MICA* et *MICB* ont un très grand polymorphisme allélique au niveau de leurs extrémités 5'UTR et 3'UTR (95, 96). D'après les [statistiques EMBL-EBI](#), *MICA* compte 224 allèles pour 104 protéines et *MICB*, 225 allèles pour 38 protéines, en avril 2021. L'expression des gènes *MICA* et *MICB* est régulée par le facteur de transcription BCL11A par ARN compétitif endogène (97). La protéine CD1 a une structure et une fonction similaire aux molécules du CMH de classe I. Elle est composée d'une chaîne lourde encodée par un gène de la famille des CD1 associée à la microglobuline $\beta 2$, formant la chaîne légère de la protéine (87, 98). Les gènes qui encodent pour la chaîne lourde de la protéine CD1 se trouvent sur le chromosome 1 (1q23). Cette molécule ne lie pas de peptides mais des fragments lipidiques endogènes et exogènes ou des Ag glycolipides. Chez l'humain, CD1 possède 5 isotypes (CD1A-E) et est non-polymorphique (98, 99). Chaque protéine CD1 a un site de liaison aux lipides qui lui sont spécifiques, ce qui a pour conséquence que chaque isotype lie spécifiquement un ensemble unique de lipides (100). La protéine CD1 est transmembranaire et présente à la surface des cellules présentatrices d'antigènes comme les molécules du CMH de classe II et elle est reconnue par le TCR des lymphocytes T CD8+ et T $\gamma\delta$ ainsi que par les cellules NK et NKT (98, 99). Les 5 isotypes de la protéine CD1 sont divisés en trois groupes selon l'homologie des séquences : le groupe 1 contient les protéines CD1a, CD1b et CD1c, la protéine CD1d compose le groupe 2 et la protéine CD1e, le groupe 3 (100, 101). Contrairement aux autres isoformes, la protéine CD1e est soluble et intracellulaire et a pour fonction de lier les lipides et de les transférer aux autre isoformes CD1 transmembranaires par le système endo-lysosomal. La protéine CD1d est la seule protéine CD1 exprimée dans les monocytes et macrophages alors que dans les lymphocytes B, les protéines CD1d et CD1c sont exprimées, par exemple (99, 101, 102).

Le système du complément a trois gènes qui codent pour des composants du complément (*C2*, *C4* et le *facteur B*) dans la région de classe III. Cette région, comme précisé précédemment, se trouve entre les régions de classe I et II sur le bras court du chromosome 6. Hauptmann et Bahram (103) se sont demandés pourquoi ces gènes du complément se trouvent dans cette région du génome qui contient majoritairement des gènes du CMH. L'étude de ces trois gènes et des protéines qu'ils encodent ont permis d'avoir des pistes de réflexion. Tout d'abord, le gène *C4* peut être sous deux formes, longue et courte qui vont encoder respectivement les protéines C4A et C4B. De plus, dans le génome humain, le gène *C4* peut être présent entre 2 et 6 exemplaires, et dans certains cas, en 8 exemplaires. Ces locus peuvent autant coder pour la protéine C4A que C4B. Pour compléter ses caractéristiques, le gène *C4* se trouve dans une région à haut polymorphisme appelée RCCX représentant les 4 gènes qui la composent. Les gènes présents dans cette région (*RP*, *C4*, *CYP21* et *TNX*) partagent des propriétés similaires à *C4*. *CYP21* et *TNX* ont tous deux, une forme longue et une forme courte de leur gène qui encodent respectivement les protéines CYP21A/CYP21B et TNXA/TNXB. Le gène *RP* a deux versions, *RP1* et *RP2* où *RP2* est

un pseudogène. Hauptmann et Bahram ont pu observer une forte activité d'échange et de recombinaison génique, ce qui offre aux haplotypes RCCX de nombreuses combinaisons. Mais l'hétérogénéité de variation de longueur de RCCX provoque des inégalités dans les échanges et recombinaison entre les haplotypes. Par exemple, l'hyperplasie surrénale congénitale est expliquée par la déficience de *CYP21* et la prédisposition de certaines maladies leucocytaires sont associées aux haplotypes RCCX (103, 104). De plus, il a été démontré que la maturation et la différenciation des monocytes et macrophages sont impactées par la production extra-hépatique des protéines C2, C4 et facteur B (104). Les protéines C2 et le facteur B ont près de 40% d'a.a. résiduels identiques et une structure similaire, dont leur domaine A. Ces deux protéines ont la même fonction mais dans des voies différentes du complément. Le complément est activé par trois voies distinctes, la voie classique, la voie lectine et la voie alternative. Le fragment actif de C2, C2a, lie C4 dans la voie classique et lectine alors que le fragment actif du facteur B, Bb, lie C3 dans la voie alternative du complément et ces complexes forment la convertase C3 (**Figure 1.17**). Les gènes codant pour les protéines C2 et le facteur B sont polymorphiques et certains des variants du gène du facteur B codent pour des protéines dysfonctionnelles (103).

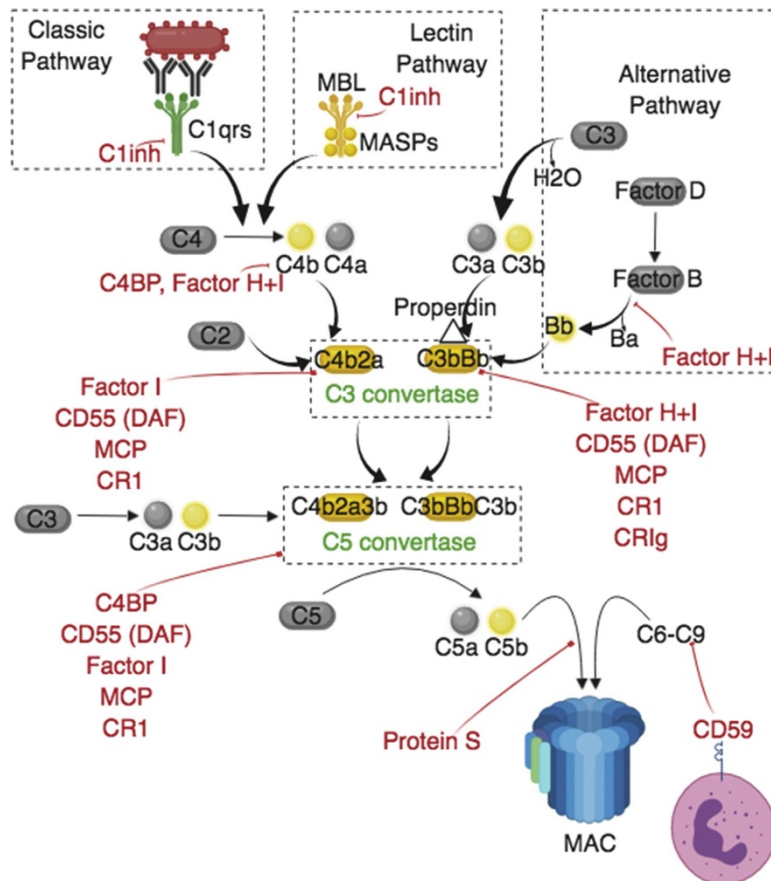


Figure 1.17. Schéma du système du complément et des protéines régulatrices. Tiré de (105).

Le système du complément a pour principales fonctions de défendre l'organisme contre les pathogènes, éliminer les débris cellulaires et les cellules apoptotiques, augmenter l'inflammation et agir comme co-stimulateur de l'activation des cellules B, qui vont induire la production d'Ac et donc, indirectement, augmenter les réponses immunes innée et adaptative (105). Pour cela, la

cascade de formation du complément est déclenchée par trois voies distinctes présentées dans la **Figure 1.17** : (1) les complexes Ag-Ac (la voie classique), (2) la liaison à une lectine (la voie lectine) et (3) l'augmentation de protéine C3 (la voie alternative). L'initiation de ces trois voies leur est spécifique, mais elles aboutissent toutes à la première étape commune, la formation de la convertase C3. La voie classique est initiée par la reconnaissance par C1q de molécules cibles telles que des complexes immuns ou des pentraxines puis, C1q va se lier à C1s et C1r pour former C1qr₂s₂, qui est responsable du clivage de C4 et C2 et ainsi, former la convertase C3 (**Figure 1.17**). La voie lectine est enclenchée par la reconnaissance de carbohydrates, tels que la collectine MBL ou les *ficolins*, à la surface microbienne par les récepteurs reconnaisseurs de formes (PPR) circulants. Le complexe MBL-MAPS1/2 va également cliver C4 et C2 pour former la convertase C3. La dernière voie est, quant à elle, déclenchée par une augmentation de protéines C3 circulantes. L'hydrolyse de la protéine C3 permet d'obtenir le fragment C3b et, en parallèle, le facteur B est clivé par le facteur D, et les fragments C3b et Bb vont pouvoir former la convertase C3. Pour former la convertase C5, un fragment C3b lie la convertase C3. La convertase C5, associé à un fragment C5a, va par la suite se lier aux protéines C6, C7, C8 et C9 pour former le complexe cytolytique MAC (**Figure 1.17**). Les fragments C3a et C4a, non utilisés dans la cascade, vont induire l'activation des mastocytes et le chimio-attracteur C5a va recruter les neutrophiles et monocytes. Les protéines solubles du système du complément sont synthétisées par le foie et la forme non clivée de ces protéines est inactive, ce qui leur permet de circuler librement dans le sang. Le clivage de ces protéines est réalisé par des protéases. La production de certaines protéines du complément peut être augmentée par les cytokines inflammatoires IFN α , IL1 et IL6 en réponse à une infection ou une inflammation. La cascade de formation du complément est régulée par deux ensembles de type protéiques, les régulateurs majeurs de la phase fluide (C1INH, C4BP, la protéine S, le facteur H, le facteur I et AI), qui ont pour rôle d'éviter l'activation inappropriée du complément, et les régulateurs liant la membrane (CD55, CD59, MCP, CR1 et CR1g), qui, quant à eux, permettent l'interruption de la cascade en séparant les fragments composant les convertases ou MAC. Certains fragments provenant du système du complément ont des récepteurs, principalement à la surface des cellules immunitaires. Ces récepteurs sont couplés à des protéines G et la liaison de leur ligand initie leur activation. Ils se retrouvent à la surface de certains types cellulaires et ont une fonction qui leur est propre (**Tableau 1.8**). À la fin de la réponse immune, le complément a pour fonction de maintenir l'homéostasie immunitaire en nettoyant l'organisme des complexes immuns par phagocytose afin d'éviter le développement d'auto-immunité (105).

Tableau 1.8. Récepteurs des fragments du complément. Tiré de (105).

Table 1 Complement receptors			
Ligand	Receptor	Cells of Distribution	Function
C3a ^a	C3aR	Mast cells, basophils	Anaphylatoxin
C3b ^a	CR1 (CD35)	RBCs, WBCs, monocytes, DCs, podocytes	Immune complex clearance, phagocytosis
C3d ^a	CR2 (CD21)	B cells	Costimulation
iC3b ^a	CR3 (CD11b/CD18)	Phagocytic cells	Phagocytosis
C3dg ^a	CR4 (CD11c/CD18)	?NK cells, monocytes, DCs	?Phagocytosis
C5a	C5aR (CD88)	WBCs, monocytes, DCs, certain tissues	Anaphylatoxin
C5a	C5L2 (intracellular, not functionally a GPCR)	WBCs, monocytes, DCs, certain tissues	?Decoy receptor

Abbreviations: DC, dendritic cell; GPCR, G protein-coupled receptor; NK, natural killer; RBC, red blood cell; WBC, white blood cell.

^a Fragments of C3.

1.2.4.2 Granules cytotoxiques

Le système immunitaire peut enclencher deux réponses immunitaires distinctes, adaptative et innée. Parmi les cellules impliquées dans la réponse immunitaire innée se trouve les cellules « tueuses » (cellules T CD4+ et CD8+, cellules NK, cellules NKT et T $\gamma\delta$) appelées ainsi car elles libèrent le contenu de leurs granules cytotoxiques, ce qui va enclencher la lyse de la cellule ou microbe cible (**Figure 1.18**). La dégranulation peut être enclenchée par les récepteurs de la famille KIR et NRC dont NKG2D, dans les cellules NK et T, et par le TCR dans les cellules T. Ces récepteurs lient les molécules du CMH de classe I (106-108). Cette liaison va déclencher la dégranulation (**Figure 1.19**)

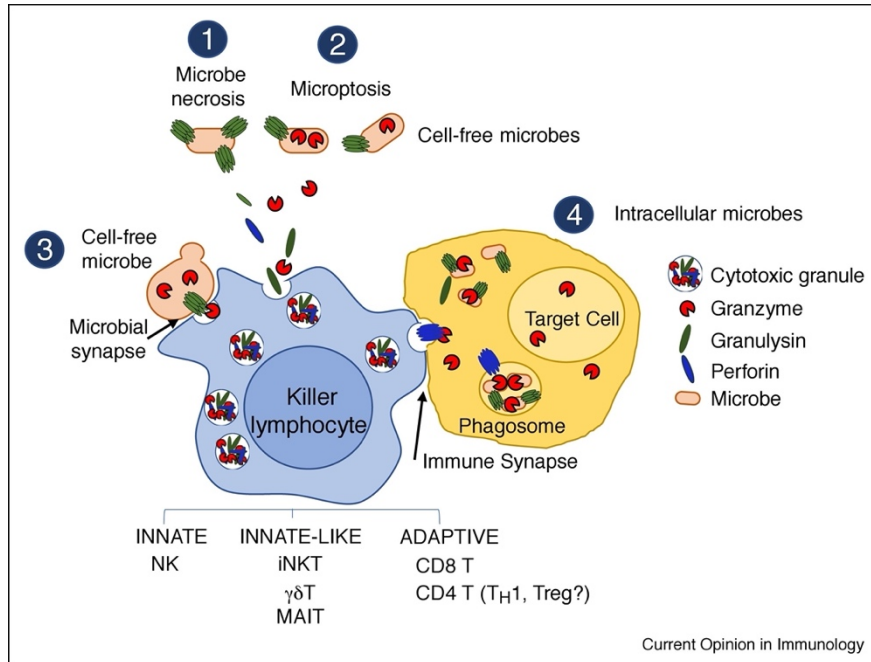


Figure 1.18. Mécanismes des lymphocytes « tueuses » pour tuer des microbes. Tiré de (109).

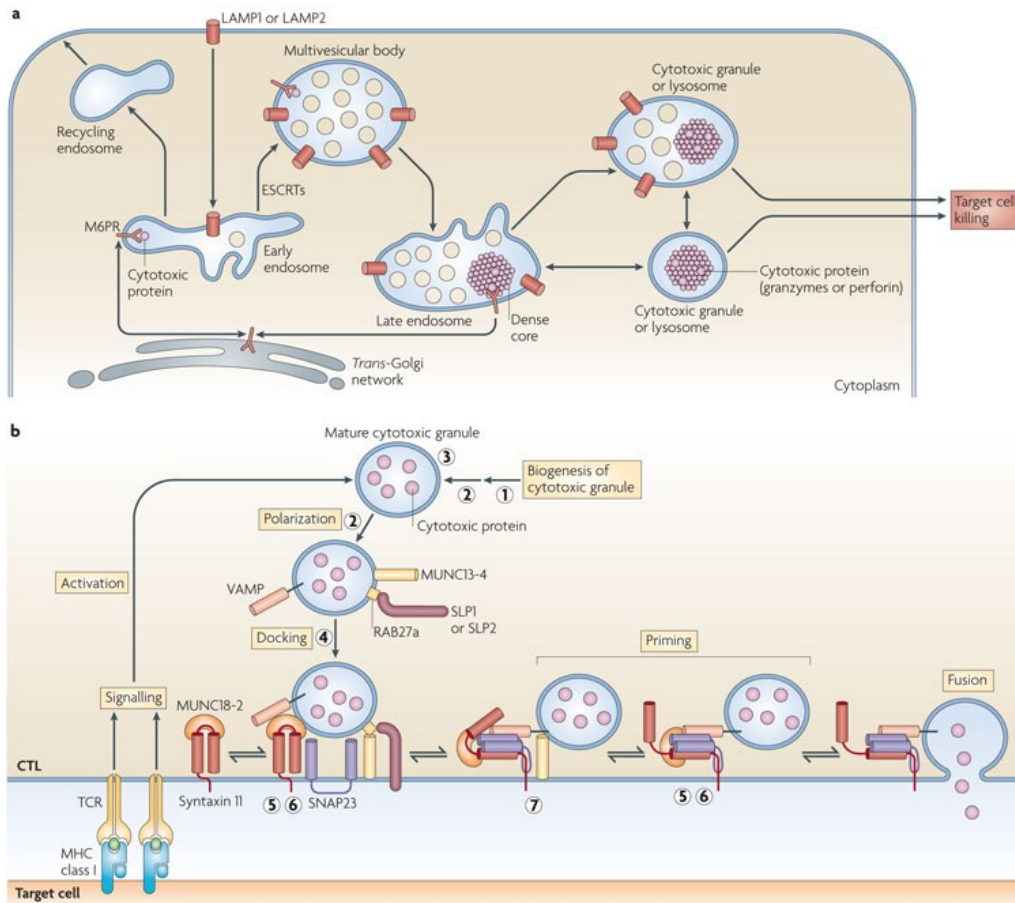


Figure 1.19. Un modèle descriptif de la biogénèse et de l'exocytose des granules cytotoxiques. a. Schéma de la biogénèse des granules cytotoxiques. b. Schéma de l'exocytose des granules cytotoxiques. Tiré de (110).

Les granules cytotoxiques sont composées de trois principaux types de protéines. La perforine (PFN), une glycoprotéine qui perfore la membrane plasmique en présence de calcium (106, 111). La calréticuline, une protéine chaperonne qui protège la PFN de l'activation en présence de calcium. La granulysine (GNLY), qui permet le transport des granzymes à travers le canal créé par la PFN dans la membrane plasmique (106, 109). Cette protéine est un peptide antimicrobien qui peut être présent sous sa forme précurseur de 15 kDa ou active de 9 kDa. La forme active de la GNLY ne peut être activée que dans un milieu à pH neutre. Le dernier type de protéine présente dans les granules cytotoxiques sont les membres de la famille des granzymes (Gzm), des protéases sérines composées de cinq membres chez l'humain : Gzm A, B, H, K et M (106, 109). La Gzm A et B sont les principales granzymes présentes dans les granules cytotoxiques et elles sont impliquées dans le clivage de nombreuses protéines intra et extracellulaires (106). La Gzm A clive, par exemple, les protéines du complexe ETC I et la Gzm B, quant à elle, enclenche la dégradation de plusieurs protéines ribosomales. Elle peut aussi activer les voies de signalisation impliquées dans l'apoptose des bactéries (**Figure 1.20**).

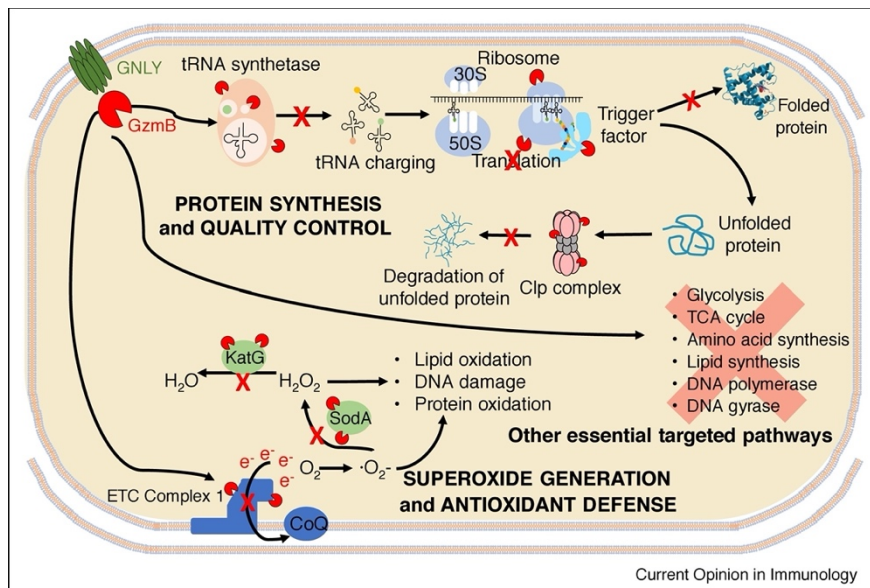


Figure 1.20. Mécanismes de lyse dans les bactéries. Tiré de (109).

1.2.5 Fonctions spécifiques à un type cellulaire : Voies de signalisation et FT

1.2.5.1 Protéines impliquées dans les fonctions spécifiques des cellules B

Les cellules B sont présentatrices d'Ag et leur activation par cette voie va induire leur différenciation en plasmocyte pour la production d'Ac ou en cellule B mémoire pour une reconnaissance plus rapide lors d'une seconde rencontre avec l'Ag. Pour activer cette voie, la cellule doit être activée par un Ag reconnu par le récepteur des cellules B (BCR) (**Figure 1.21**)

(112). Une fois le BCR stimulé, les voies de signalisation permettant l'activation des cellules B, vont être enclenchées. Ce récepteur est le point central des cellules B primaire. Il est composé de deux paires de chaîne d'immunoglobuline (Ig) lourde et légère et l'extrémité extracellulaire est la seule partie variable de la protéine (112). Ce récepteur se lie à la protéine CD79 formée par l'hétérodimère CD79A et CD79B. Cette liaison permet au BCR d'être enchevêtré dans la membrane plasmique des cellules B (**Figure 1.21**). Lorsque le BCR reconnaît un Ag, la liaison avec CD79 est rompue et le BCR est internalisé par la cellule. Ce mécanisme enclenche la voie de signalisation du BCR (**Figure 1.21**) (112). La partie intracellulaire de CD79 est appelée ITAM et le signal du BCR commence par une liaison entre la région ITAM et une kinase SRC. Cette kinase se lie à LYN, FYN ou BLK puis elle phosphoryle la région ITAM du BCR. Cette phosphorylation va entraîner l'activation de la protéine SYK, ce qui marque le début de la signalisation du BCR. La phosphorylation de PLC γ_2 associé à PIP $_2$ produit les messagers secondaires, DAG et IP $_3$. IP $_3$ va, par la suite, se lier à son récepteur présent à la membrane du RE et ainsi permette de relâcher le Ca $^{2+}$ présent dans le RE. Le signal du Ca $^{2+}$ induit par le BCR impose une fenêtre de temps pour que la cellule B soit maintenue en vie ou non et ainsi permettre sa différenciation en plasmocyte ou en cellule B mémoire (**Figure 1.21**) (112). Le messager DAG va aussi lier la protéine RasGRP3 pour activer la voie de signalisation des MAP kinases (MAPK). Dans cette voie, MEK phosphoryle ERK et sa dimérisation va lui permettre de transloquer du cytoplasme vers le noyau et d'activer les FT Fos, Jun, et Ets pour qu'ils régulent l'expression de leur gènes cibles (**Figure 1.21**) (112). En parallèle, PLC γ_2 déclenche la voie de signalisation passant par PKC pour permettre l'ubiquitination et la dégradation de I κ B et ainsi libérer NF- κ B qui va transloquer du cytoplasme vers le noyau et réguler ses gènes cibles (**Figure 1.21**) (112). Pour finir, la stimulation du BCR enclenche la phosphorylation du corécepteur CD19 par LYN permettant ainsi la liaison avec la sous-unité p85 à PI $_3$ K. La liaison de la protéine NCK à CD79 permet le recrutement de BCAP qui lie p85 pour induire la voie de signalisation PI $_3$ K. Dans cette voie, PDK1 va phosphoryler AKT avec l'aide de mTORC2 (**Figure 1.21**). Par la suite, la phosphorylation de GSK3 par AKT permet l'activation de molécules régulatrices de la croissance cellulaire (c-Myc et cycline D3) ce qui enclenche les mécanismes de survie cellulaire (**Figure 1.21**) (112).

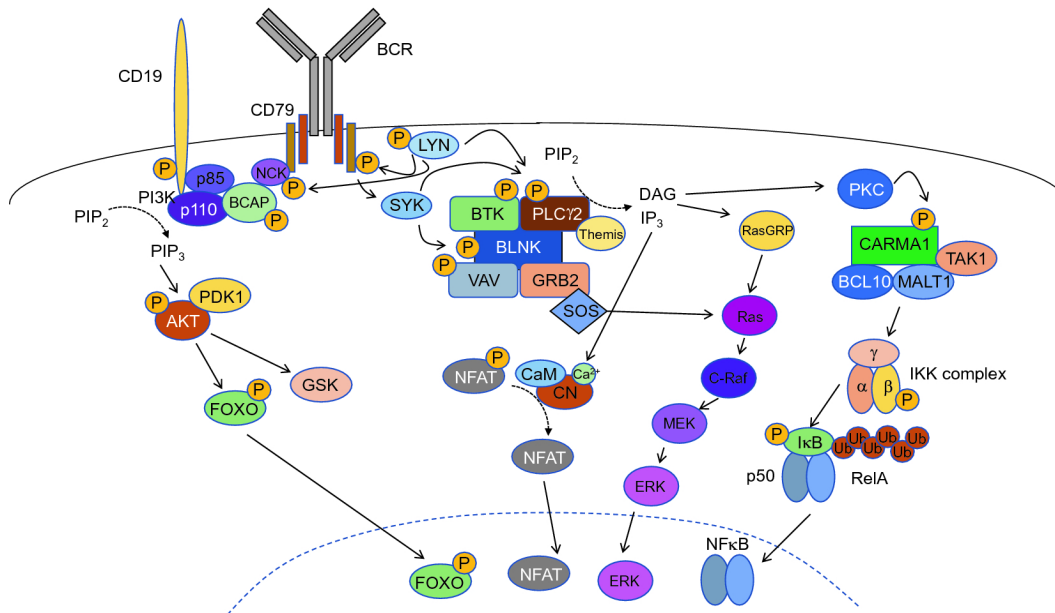


Figure 1.21. Voie de signalisation des cellules B en réponse à un Ag. Tiré de (112).

Des voies de signalisation sont présentes dans les cellules B pour stabiliser et localiser le BCR. Ces voies sont présentées dans la **Figure 1.22**. LYN, qui est essentielle dans la signalisation induite par le BCR, est aussi impliquée dans la voie de régulation négative du BCR. Il peut phosphoryler les récepteurs FcγRIIB et CD22, qui sont des inhibiteurs du BCR. Ces phosphorylations vont permettre de recruter la protéine SHIP-1 qui va déphosphoryler les protéines CD79, SYK, BLNK et VAV et ainsi arrêter la voie de signalisation du BCR (**Figure 1.22**) (112). En parallèle, le récepteur CD300d lie DAP12 et SHP-1, ce qui va suspendre l'activation des cellules B (**Figure 1.22**) (112). La protéine SHIP-2 est recrutée par PD-1 ou FcγRIIB et, tout comme SHIP-1, SHIP-2 va déphosphoryler CD79, SYK, PLCγ et GRAB1 (**Figure 1.22**) (112). La protéine adaptatrice DOK-3, quant à elle, va supprimer la fonction de LYN et former un complexe avec GRB2 pour lier PLCγ₂ and SYK et inhiber leurs activités (**Figure 1.22**) (112). A20 est une protéine cytoplasmique qui, en parallèle à CYLD et TANK, diminue la régulation du signal de NF-κB (**Figure 1.22**) (112).

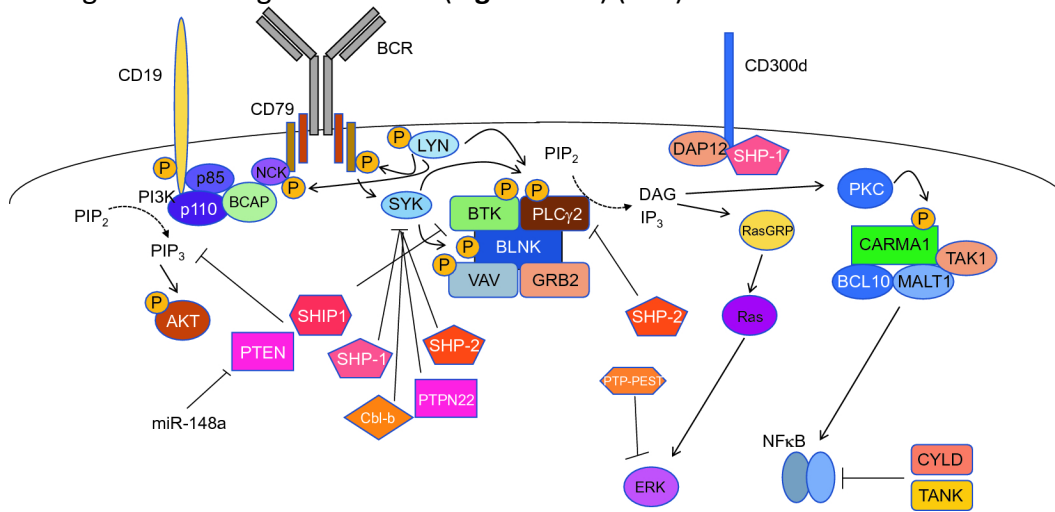


Figure 1.22. Antagonisation du signal du BCR. Tiré de (112).

Le récepteur BAFFR est un membre de la famille des récepteurs TNF et possède plusieurs ligands, eux même appartenant à la famille TNF (ex. : BAFF, TALL-1 et zTNF4). La protéine BAFF est sécrétée par les monocytes, les macrophages, les cellules dendritiques et les cellules de la moelle osseuse et favorise la survie des cellules B en se liant à son récepteur, BAFFR. Ce dernier peut enclencher la voie non-canonique dépendante de NF- κ B2, mais elle reste beaucoup plus lente et complexe que la voie canonique induite par la signalisation du BCR et passant par NF- κ B (**Figure 1.23**). La voie de signalisation induite par BAFFR va rallonger la durée de vie de la cellule B en stabilisant la mitochondrie et en initiant la synthèse de protéines (113).

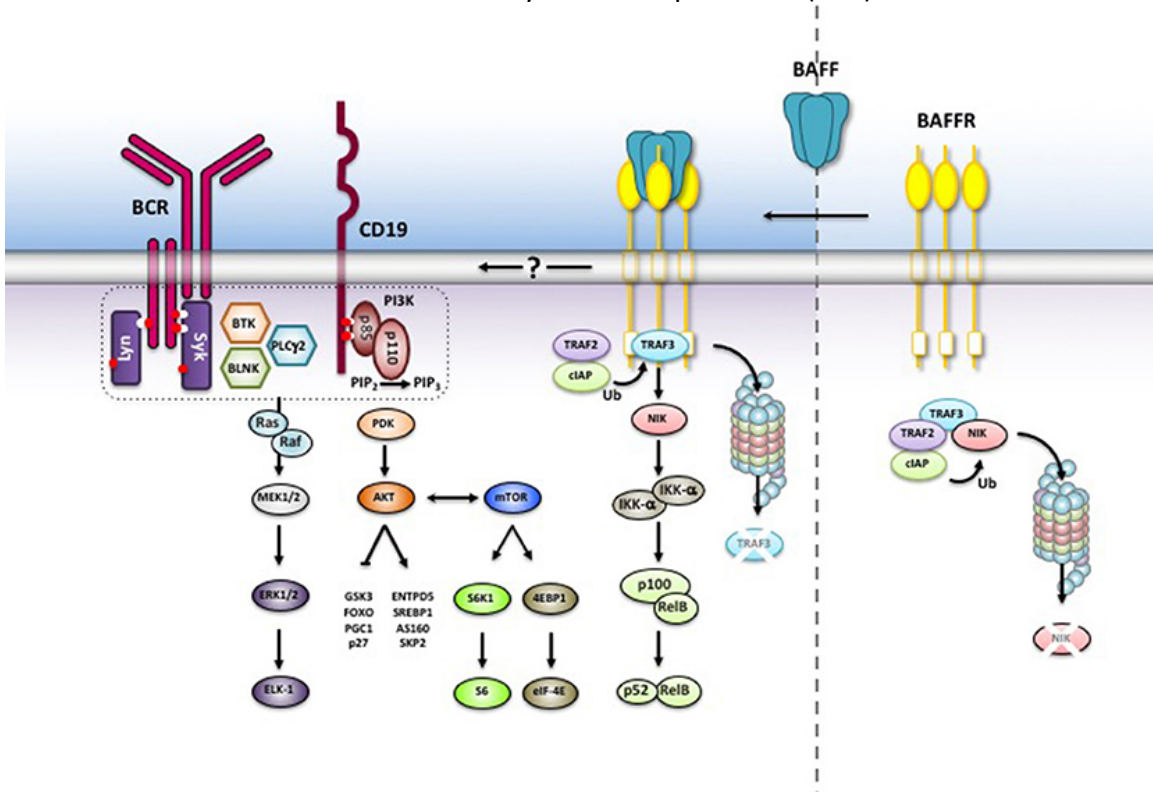


Figure 1.23. Signal intracellulaire induit par le récepteur BAFF. Tiré de (113)

1.2.5.2 Protéines impliquées dans les fonctions spécifiques aux neutrophiles

Les neutrophiles sont des cellules immunitaires dites polynucléées car elles possèdent de nombreuses vésicules dans leur cytoplasme qui pendant longtemps, ont été confondu avec des noyaux cellulaires. Ces vésicules sont le produit de la granulopoïèse qui se développe en parallèle avec la différenciation des neutrophiles. Les vésicules de la granulopoïèse se divisent en 4 stades distincts : granule azurophile (AG), granule spécifique (SG), granule gélatinase (GG) et la vésicule sécrétrice (SV) (**Figure 1.24**) (114). La maturation des granules est régulée par des FT, des micro-ARN et des protéines du cycle cellulaire.

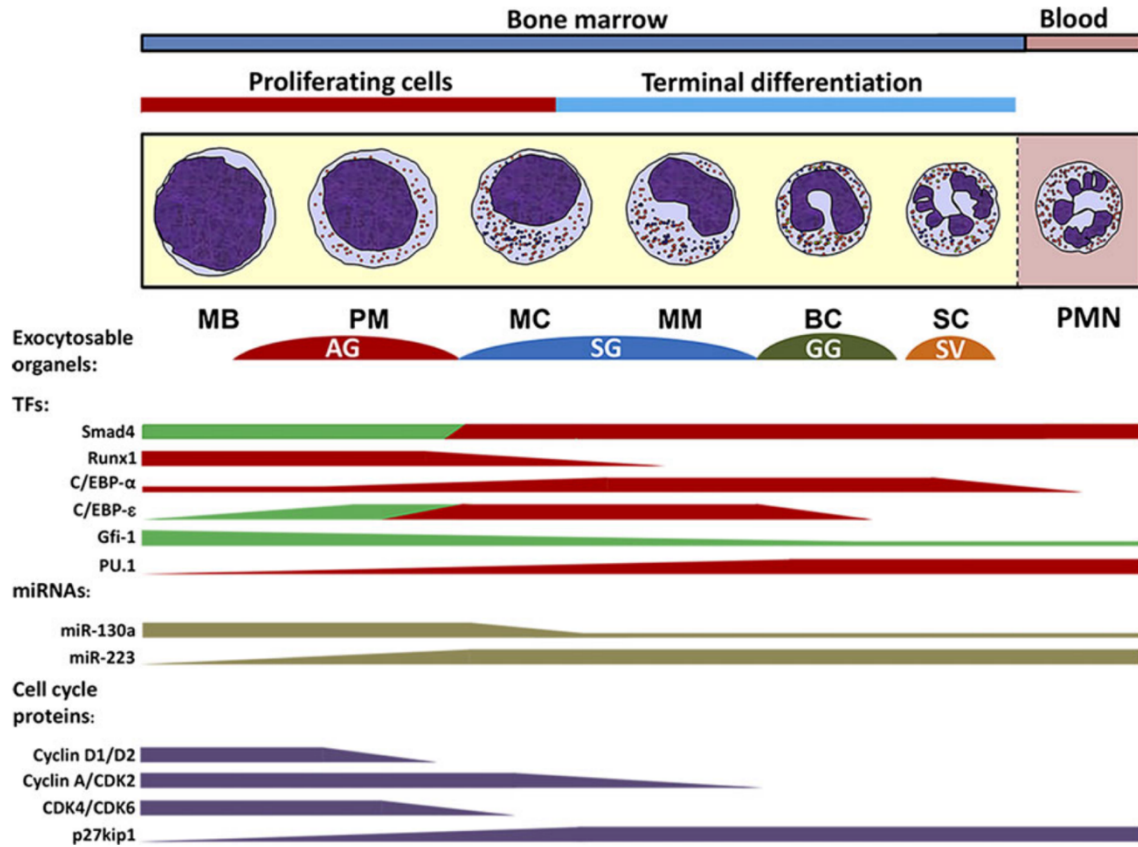


Figure 1.24. Illustration schématique de la granulopoïèse. Le profil de l'ARNm des facteurs de transcription est représenté en vert et en rouge, est représenté le profil des protéines. Le profil d'expression de Gfi-1 n'est pas connu. MB: Myeloblasts; PM: Promyelocytes; MC: Myelocytes; MM: Metamyelocytes; BC: Band cells; SC: Segmented neutrophilic cells; PMN: Polymorphonuclear leukocytes. Tiré de (114).

Les neutrophiles circulants sont recrutés sur un site d'inflammation grâce à la diapédèse. Ce processus se divise en quatre grandes étapes présentées dans la **Figure 1.25** : le roulement, l'adhésion, le glissement et la transmigration (115, 116). Ces différentes étapes sollicitent de nombreuses protéines d'adhésion telles que les sélectines et les intégrines, ainsi que des protéines de la matrice extracellulaire (collagène, laminine, fibronectine et fibrinogène). Lors de la diapédèse, l'ensemble des SV fusionnent avec la membrane plasmique des neutrophiles alors que respectivement 40%, 25% et 7% des GG, SG et AZ fusionnent (114). Ce mécanisme, impliquant la polymérisation d'actine et l'interaction des vésicules/granules avec le complexe SNARE en présence de Ca^{2+} , est la première étape de la dégranulation. Par la suite, les microtubules vont être importants dans la dégranulation spécifique des AG alors que la dissolution du réseau cortical de l'actine F permet la dégranulation de l'ensemble des granules et vésicules présentes à la membrane (114). La dégranulation va permettre la libération de molécules microbicides qui vont tuer les micro-organismes (116).

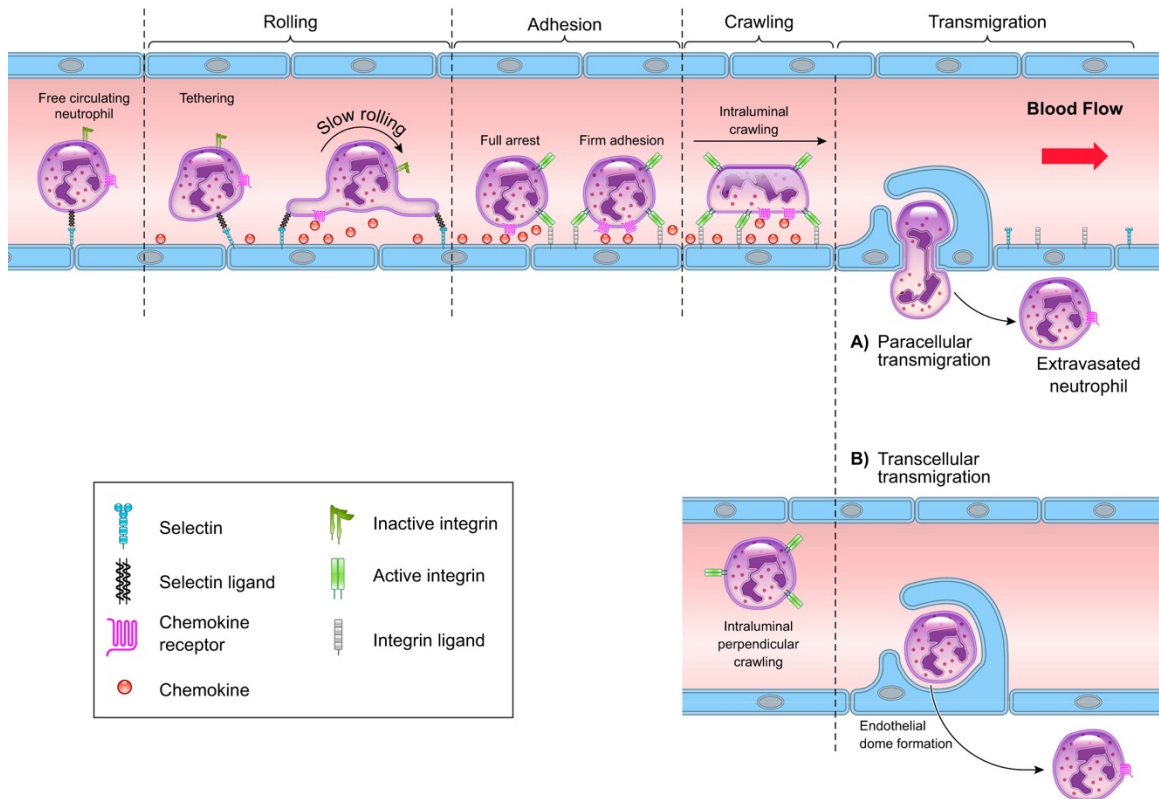


Figure 1.25. Cascade classique de recrutement des neutrophiles. Tiré de (115).

Arrivés au site d'inflammation, les neutrophiles vont pouvoir être activés, ce qui va enclencher de nombreux mécanismes au sein de la cellule, en plus de la dégranulation. Deux principaux mécanismes sont la formation et libération de NETs (*neutrophil extracellular traps*) et la production de ROS (*Reactive Oxygen Species*) (116-118). Contrairement à la phagocytose, la production de ROS et la dégranulation, les NETs ont une activité antimicrobienne capable de tuer ou suspendre la prolifération de champignons et de bactéries, appelée NETosis. Les NETs sont des filaments de chromatines nucléaires relâchés dans la matrice extracellulaire par les neutrophiles et constitués d'enzymes et de peptides étant induit par la production de ROS (ex. : défensines et cathélicidines (NE et MPO)) (**Figure 1.26**) (116, 117). Les enzymes élastase et myéloperoxydase sont responsables de la présence de MPO et NE sur la chromatine, lors de la formation des NETs dans le noyau (117).

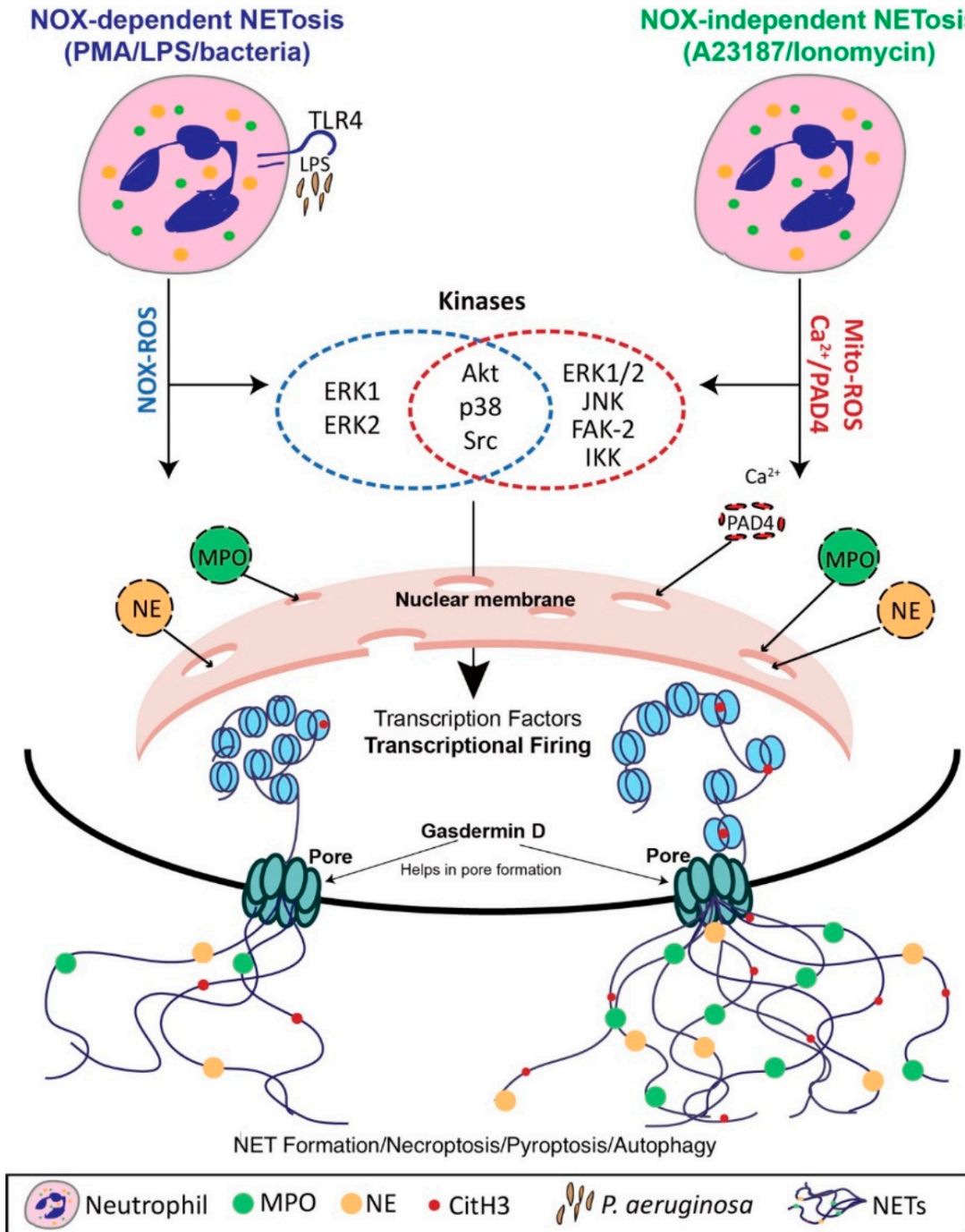
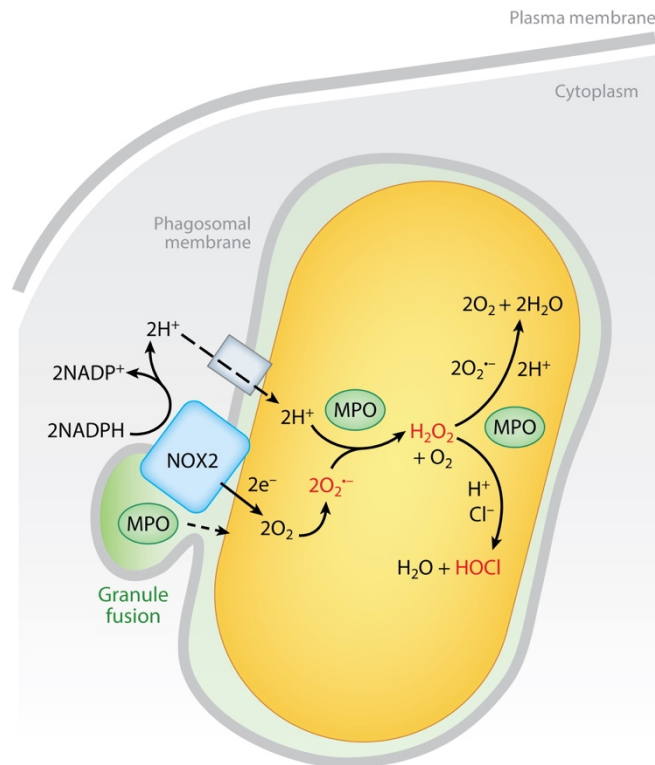


Figure 1.26. Modèle de formation des NETs. Tiré de (117).

La production de ROS permet aux neutrophiles de synthétiser l'ensemble du spectre des oxydants radicaux et non-radicaux des réactions redox biologiques (réaction d'oxydoréductase) (118). Les oxydants les plus intéressants produits par les neutrophiles sont le superoxyde et le peroxyde d'hydrogène, tous deux synthétisés par MPO et dépendant de la disponibilité de NOX2 activé (Figure 1.27) (118).




 Winterbourn CC, et al. 2016.
Annu. Rev. Biochem. 85:765–92

Figure 1.27. Schématisation de la protéine NOX2 activée à la surface de la membrane d'un phagosome ayant ingéré une bactérie (jaune), fusionné à une granule (contenu de la granule en vert). Tiré de (118).

1.2.5.3 Protéines impliquées dans les fonctions spécifiques aux macrophages

Le zinc (Zn) est un cation hydrophile qui ne peut pas traverser passivement la membrane cellulaire. Les transporteurs membranaires de la famille des SLC39 et SLC30 jouent un rôle important dans l'homéostasie du Zn où les SLC39 importent le Zn dans la cellule alors que SLC30 l'exporte (119). Par exemple chez l'humain, SLC39A10 participe à l'homéostasie du Zn et SLC39A8 est impliquée dans les réponses à l'inflammation. Le Zn est impliqué dans de nombreuses fonctions immunes des macrophages : phagocytose, présentation d'Ag et immunomodulation (**Figure 1.28**) (119). La quantité de Zn doit être basse dans les monocytes pour permettre leur différenciation en macrophages alors que les particules d'oxyde de Zn sont essentielles pour induire la migration, l'adhésion et l'absorption du cholestérol par les macrophages. En revanche, l'excès de Zn peut induire l'apoptose des macrophages. Le Zn joue un rôle important dans la signalisation de la réponse inflammatoire chez les macrophages. Il va agir comme un activateur des récepteurs TLR, des MAPK, de la PKC ainsi que des phosphodiesterases et du FT NF- κ B. Par exemple, le Zn est impliqué dans le signal de TLR4 en réponse à la phosphorylation de IRAK1 en dégradant la pro-caspase 1 et NLRP3 et ainsi permet d'inhiber la production de IL1 β après une stimulation par des lipopolysaccharides (LPS) ou un pathogène microbien (**Figure 1.28**) (119). Dans la voie de signalisation de TLR4, le Zn est requis pour la déphosphorylation de p38, MEK1/2 et ERK1/2, ce qui permet l'activation de NF- κ B et la libération des cytokines pro-inflammatoires

TNF- α , IL-1 β et IL-6 (119, 120). Dans certaines maladies chroniques, la quantité de Zn dans la cellule peut être dérégulée et entraîner des comportements anormaux chez les macrophages (**Figure 1.28**). Par exemple, dans la polyarthrite rhumatoïde, la déficience en Zn a pour conséquence d'augmenter la quantité de TNF- α , IL-1 β et IL-8 dans la lignée monocytes/macrophages.

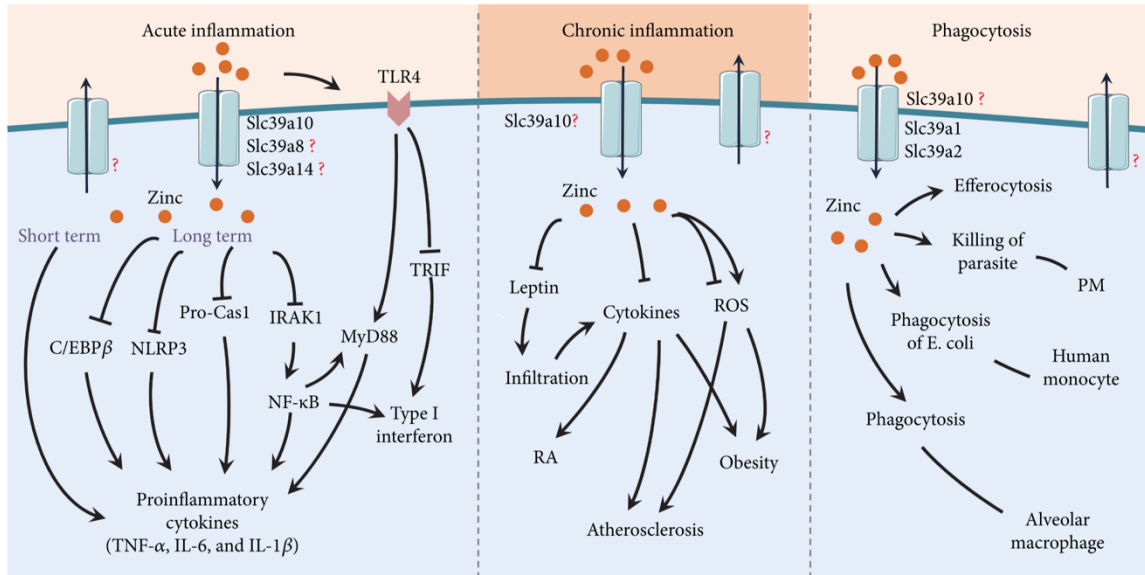


Figure 1.28. Rôle et homéostasie du zinc dans les fonctions des macrophages. Tiré de (119).

Une des fonctions principales des macrophages est la phagocytose. Ce mécanisme permet d'engloutir une cellule ou un pathogène dans une vésicule phagocytaire, d'enclencher une digestion interne, de transformer la vésicule phagocytaire en un débris cellulaire pro-inflammatoire puis de le relâcher dans la matrice extracellulaire. Ce mécanisme est initié par la liaison d'une chimiokine ou d'un ligand aux TLR sur les récepteurs Fc γ , TLR ou GPCR (121-123). Cette liaison produit un changement de morphologie des macrophages dû à la polymérisation des fibres d'actine (**Figure 1.29**). Les voies de signalisation induites dans les macrophages, vont être dépendantes du types de liaison, présentés dans la **Figure 1.29**.

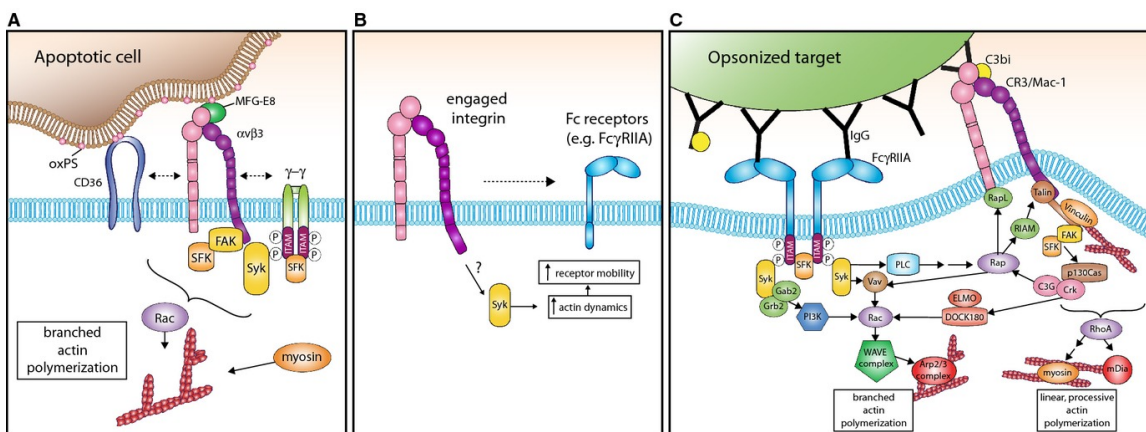


Figure 1.29. Synergie des récepteurs durant la phagocytose. **A.** L'engagement coopératif des cellules apoptotiques est acheminé par CD36. **B.** L'engagement des intégrines augmente la mobilité de Fc γ RIIA et possiblement d'autres récepteurs par des mécanismes indéterminés à travers

l'activation des kinases de la famille Src et Syk. C. Signalisation par de multiples récepteurs engagés par la même particule, peut être additif. Tiré de (121).

Aujourd'hui, de nombreuses connaissances ont été accumulées sur les cellules du système immunitaire. Ces cellules sont beaucoup étudiées car leur prélèvement est peu invasif et rapide. Comme énuméré précédemment, les protéines présentes à la surface des différents types cellulaires immunitaires sont très bien connues et sont maintenant utilisées comme marqueurs de surface pour distinguer les populations cellulaires. L'étude de la différenciation, le développement et l'activation des cellules immunitaires a permis d'identifier les protéines de types cytokines et chimiokines impliquées dans la communication cellulaire et indispensables au bon fonctionnement du système immunitaire. Mais surtout, de nombreuses études se sont intéressées au rôle des cellules immunitaires et de leurs fonctions cellulaires. Ces dernières ont pu identifier les voies de signalisation présentes au sein de la cellule et certaines des protéines qui y sont impliquées. Malgré toutes ses connaissances acquises, il reste encore de nombreux mécanismes cellulaires dont l'analyse reste à approfondir qui pourraient être expliqués, entre autres, par le transcriptome et sa régulation.

1.3 Étude du transcriptome humain

1.3.1 Analyse de l'expression génique: Outils et méthodes

Depuis de nombreuses années, des méthodes de manipulation de l'ARN ont été élaborées pour estimer l'abondance du transcriptome humain. Les données de ces méthodes, telles que *EST*, micro-puces, séquençage de l'ARN et séquençage de l'ARN par cellule unique (scRNA-Seq), sont utilisées pour mesurer le niveau d'expression des gènes, transcrits, exons ou introns d'un transcriptome. Ces données étant importantes en quantité, plusieurs outils ont été développés afin de traiter ces grands ensembles de données. La complexité de mesurer le niveau d'expression de l'ensemble des corps transcriptomiques a amené certains développeurs à optimiser leurs outils pour l'expression par gènes, l'expression par transcrits ou l'expression par exons. Parmi les outils optimisés pour l'expression par gènes se trouvent TopHat (124), RSEM (125) et DEXSeq (126), par exemples. D'autres outils comme StringTie (127, 128), Salmon (129), Sailfish (130) et IsoformEx (131) mesurent principalement l'expression par transcrits. Pour certains des outils de ces deux catégories, il est possible d'avoir aussi l'expression par exons.

Ces outils donnent la mesure de la quantité de lectures (appelée en anglais *reads*) associées à chaque gène, transcrit ou exon du transcriptome. Les données de comptes bruts des lectures ne permettent pas de quantifier l'expression des transcrits ou gènes car elles ne prennent pas en compte la longueur de ces derniers. La correction par la longueur est primordiale car, pour un même temps de PCR, les ARN de petite taille vont être beaucoup plus amplifiés que les grands ARN et donc, nous obtenons une surestimation de l'abondance des lectures provenant de petits ARN. Pour pallier cela, les données de comptes bruts des lectures peuvent être converties en différentes unités dépendamment de l'approche d'analyse utilisée par la suite. Si nous souhaitons faire une comparaison inter-échantillon, les données devront être converties en CPM (*Counts Per Million*) ou TPM (*Transcripts Per Million*). Pour une comparaison intra-échantillon, l'unité de mesure sera FPKM (*Fragments Per Kilobase Million*) ou RPKM (*Reads Per Kilobase Million*).

Après l'obtention de la mesure de l'expression, ces données vont pouvoir être analysées sous plusieurs angles. Par exemples, les différences d'expression entre plusieurs conditions (gènes et transcrits) ou la différence d'expression des isoformes au sein d'un type cellulaire (transcrits et exons). Pour l'analyse de l'expression différentielle par gènes, des outils tel que WGCNA (132) permettent d'identifier les gènes qui ont un même profil d'expression à travers un ensemble d'échantillons pouvant provenir de différents types cellulaires et/ou tissus. L'analyse de l'expression différentielle par transcrits est plus complexe car cette analyse doit tenir compte d'un plus grand nombre de paramètres. Certains outils présentés dans le **Tableau 1.9** ont relevé le défi, mais se limitent à la comparaison de l'expression entre deux conditions. rMATS est le seul outil de cette liste qui permet de comparer deux conditions avec plusieurs réplicas pour chacune d'entre elles (133). Des études ont utilisé ces outils pour analyser l'expression différentielle par transcrits. Certaines d'entre elle sont présentées dans le **Tableau 1.10**. Ces études offrent des pistes d'analyse, mais se limitent à la comparaison de deux conditions distinctes et donc, ne permettent pas de comparer plusieurs types cellulaires d'un même système biologique tel que le système immunitaire.

Tableau 1.9. Exemples d'outils d'analyse de l'expression différentielle par exons/transcrits.

Auteurs	Nom de la Méthode	Type d'analyse	Type de données	Description de la méthode
Schafer S. et al. ⁽¹³⁴⁾	<i>Percent spliced in index (PSI)</i>	Épissage alternatif	Données de séquençage utilisant l'ARN	Score: Rendement de l'épissage d'un exon donné dans la population de transcrits d'un gène (sommaire des événements d'épissage alternatif) Quantification du nombre de transcrits incluant ou excluant un exon donné : Index (0-100%) estime la fraction d'isoformes ARN contenant l'exon donné Possibilité de prédire l'épissage: couplage des résultats PSI avec une analyse de différenciation d'épissage.
Kartz Y. et al. ⁽¹³⁵⁾	<i>Mixture-of-isoforms (MISO) model</i>	Épissage alternatif	Données de séquençage utilisant l'ARN et données d'épissage du facteur hnRNP H1	Estime l'abondance de l'expression des exons et isoformes d'ARN après épissage alternatif (méthode PSI optimisée) : Intervalle de confiance Identifie les exons et isoformes d'ARN différenciellement régulés hnRNP H1 : Identifie la régulation du clivage alternatif et polyadénylation
Shen S. et al. ⁽¹³³⁾	<i>Replicate multivariate analysis of transcript splicing (rMATS)</i>	Épissage alternatif	Données de séquençage utilisant l'ARN	Détection de la différenciation de l'épissage alternatif à partir de données RNA-seq répétées Modèle hiérarchique pour modeler les niveaux d'inclusions des exons (Méthode PSI optimisée) Modèle spécifique pour l'analyse des <i>paired</i> répétés (Distribution normale bivariée et valeur de corrélation) Modèle statistique flexible de test d'hypothèses définies par l'utilisateur : hypothèses nulle et alternative pour la différenciation de l'épissage alternatif
Wu J. et al. ⁽¹³⁶⁾	<i>SpliceTrap</i>	Épissage alternatif	Données de séquençage utilisant l'ARN Base de données TXdb (exons humains)	Méthode pour quantifier les niveaux d'inclusion des exons et construire des profils d'épissage alternatif (ex.: saut d'exons, sites d'épissage 5' ou 3', rétention d'introns) Base de données TXdb sur les exons humains utilisée pour estimer le niveau d'expression de tous les exons (inférences Bayésienne indépendantes) Possibilité de déterminer le niveau d'inclusion de chaque exon dans des conditions <i>single-cell</i> sans génome de référence Possibilité d'étudier les changements de niveaux d'exon unique et identifier les éléments régulateurs cis impliqués
Middleton R. et al. ⁽¹³⁷⁾	<i>IRFinder: Intron retention (IR)</i>	Épissage alternatif Régulation des événements IR	Données de séquençage utilisant l'ARN	Détection des événements de rétention d'introns et des codons terminaux prématurés (nonsense-mediated decay (NMD)) Plus grande précision que MISO et DEXseq, identifie les événements IR pour des données avec une faible couverture ou un faible alignement sur le génome de référence Mesure du niveau d'épissage et de l'abondance intronique Ratio: Portion des transcrits possédant un intron non éliminé par les mécanismes épissages

Tableau 1.10. Exemples d'analyse de l'expression différentielle d'exons/transcrits dans plusieurs types cellulaires.

Auteurs	Espèces	Type de données	Types cellulaires	Méthode d'analyse	Résultats
Ergun A. et al. ImmGen Consortium ⁽¹³⁸⁾	Souris	Micro-puces et Séquençage d'ARN	Cellules CD4+ T versus B ----- Entre les cellules SC, DC, Gn, MF, Mo, NK, NKT, T ab, T gd et B	<i>Exon skipping</i> (Ratio inclusion/exclusion; Ratio de l'expression exon/gène): Expression par transcrite (Séquençage d'ARN) ----- Méthode k-means optimisée; <i>Clustering</i> de 4,321 exons marqués à travers les types cellulaires du projet ImmGen : Expression par exon (Micro-puces et Séquençage d'ARN)	Liste d'exons spécifiques à chaque type cellulaire pour des gènes partagés ----- Partitionnement de la distribution des exons en 28 clusters
Schaub A., Glasmacher E. ⁽¹³⁹⁾	Humain et souris	Micro-puces et Séquençage d'ARN	Comparaison au sein de la différenciation de la lignée macrophage, cellules dendritiques, cellules T ou cellules B	Revue de littérature: Rôle de l'épissage dans le contrôle de la différenciation cellulaire Rôle de l'épissage dans le contrôle de l'immunité innée Mécanismes d'épissage important dans la fonction des cellules T et B	Exemple de résultats (Tableau 1.9)
Byun S., Han S., Zheng Y., Planelles V., Lee Y. ⁽¹⁴⁰⁾	Humain	Séquençage d'ARN	Cellules T infectées par HIV-1 versus non infectées	Méthode PSI : Expression par transcrits	Dans les cellules T infectées: 427 gènes candidats avec une expression différentielle des exons obtenus par épissage alternatif (20 gènes encodant pour une protéine de surface, 35 pour une kinase et 121 pour une protéine impliquée dans le système immunitaire) Analyse des interactions protéine-protéine: Identification de 6 sous-réseaux essentiels pour le cycle de multiplication virale (ex.: Régulation de la transcription par TP53, CMH de classe I, transition G2/M et phase tardive du cycle de vie du HIV)
Ullrich S., Guigó R. ⁽¹⁴¹⁾	Humain et souris	Séquençage d'ARN	Comparaison au sein de la différenciation de la lignée des cellules B	Méthode IR pour mesurer les délais entre la transcription et la translation pour la différenciation cellulaire: Expression par transcrits	Augmentation de IR dans la translation de différenciation des cellules B dans l'hématopoïèse aux organes secondaires.
West K.O. et al. ⁽¹⁴²⁾	Souris	Séquençage d'ARN	Macrophages (RAW 264.7)	Étude du facteur d'épissage hnRNP M répresseur de l'expression des gènes de l'immunité innée: Knockdown (KD) Étude de hnRNP M mutée spécifiquement sur ses sérines (S85, S431, S480, S574, and S636) Étude de données Séquençage d'ARN provenant de cellules infectées et comparaison à une stimulation par LPS (ex: <i>Salmonella enterica serovar Typhimurium</i> (SL1344) ou <i>Recombinant Vesicular stomatitis virus</i> (VSV; Indiana serotype))	KD: Hyper-induction d'un unique régulateur de gènes inflammatoires et antimicrobiens Mutation: impacte la fonction d'épissage et des niveaux d'expression des gènes domestiques et amoindrit l'induction des gènes spécifique au système immunitaire innée Infection: pas de différence avec une stimulation par LPS
Rotival M., Quach H., Quintana-Murci L. ⁽¹⁴³⁾	Humain	Séquençage d'ARN	Différenciation versus activation des monocytes	PSI : Expression par transcrits Activation des monocytes : infection par <i>Mycobacterium tuberculosis</i> sQTL/eQTL: Étude de la régulation des variations génétiques avec les changements d'épissage alternatif et l'expression génique	Contrôle génétique par l'épissage alternatif pour 993 gènes dans les monocytes au repos ou stimulés. Les changements dans l'expression génique induisent les changements d'épissage. sQTL/eQTL: 1,464 événements d'épissage alternatif significativement associés à un SNP (21% des 4 739 événements observés); 11% des sQTL chevauchent des eQTL (155 sQTL): majoritairement, les sQTL causaux ont un effet modéré sur l'expression génique. Indépendance entre les déterminants génétiques pour l'usage d'isoformes ARN et l'expression génique.
Liu H. et al. ⁽¹⁴⁴⁾	Humain	Séquençage d'ARN	Différenciation versus activation des monocytes	Méthodes rMATS, MISO et SpliceTrap : Expression différentielle par gènes et transcrits	Différenciation des monocytes en macrophage et activation: millier de changements pour l'expression génique et l'épissage alternatif; Augmentation des inclusions d'exons.
Shalek A.K. et al. ⁽¹⁴⁵⁾	Souris	Séquençage d'ARN de cellules uniques	Cellules dendritiques dérivées de la moelle osseuse	Méthode MISO pour estimer le PSI : Expression par exons	Liste des gènes avec des exons spécifiques à chaque condition
Kalam H., Fontana M.F., Kumar D. ⁽¹⁴⁶⁾	Humain	Séquençage d'ARN	Macrophages infectés versus non infectés	Méthodes MATS par analyse bayésienne : Expression par gènes et transcrits Infection: <i>Mycobacterium tuberculosis</i> (Mtb)	Dans les macrophages, Mtb impacte l'épissage alternatif.

1.3.2 Construction d'un réseau de co-expression génique : Outil WGCNA

L'expression génique est en partie responsable de la fonction cellulaire. Ces fonctions sont le reflet de processus biologiques dont des mécanismes de régulation ou des voies de signalisation, par exemples. Pour que ces processus aboutissent, un ensemble de protéines doit être présent dans la cellule. Ces protéines travaillent en synergie, donc elles sont fortement corrélées. Cela suggère que l'expression des gènes qui encodent ces protéines doit, elle aussi être corrélée. De ce fait, l'étude de réseaux de co-expression génique permettrait de mieux comprendre ces processus, et donc, la fonction cellulaire.

Weighted gene co-expression network analysis (WGCNA) est un outil de construction de réseaux de corrélation pour trouver des modules (clusters) de gènes co-exprimés (132). Cet outil prend en entrée des données d'expression génique de type micro-puces ou séquençage d'ARN. Dans de nombreux cas, WGCNA a été utilisé pour identifier des gènes clés dans des voies de signalisation associées à une maladie (147-150).

Pour construire la matrice de corrélation, WGCNA va tout d'abord calculer, pour chaque paire de gènes de l'ensemble de données, le coefficient de corrélation Pearson. Pour éliminer le bruit des données et optimiser les résultats pour les ensembles de données composés d'un petit nombre d'échantillon, une valeur β doit être prédéfinie. Pour déterminer la valeur β , l'outil WGCNA estime que les réseaux d'expression génique suivent approximativement une topologie de type *scale-free* (151, 152). Mais pour choisir au mieux la valeur β , il doit maximiser le modèle basé sur la topologie *scale-free* et maintenir un haut niveau moyen de connectivité entre les gènes. Pour cela, deux graphes sont dessinés représentant chacun de ces deux paramètres et permettant ainsi de choisir au mieux la valeur β (**Figure 1.30**). Dans le manuel d'utilisation de WGCNA, il est conseillé de trouver une valeur β proche de $R^2 > 0,80$, pour la topologie de type *scale-free*. Il est conseillé d'utiliser une valeur $\beta = 12$ avec un réseau de type *signed* (153). *Ce dernier* permet de conserver l'information du signe du coefficient de corrélation de chaque paire de gènes.

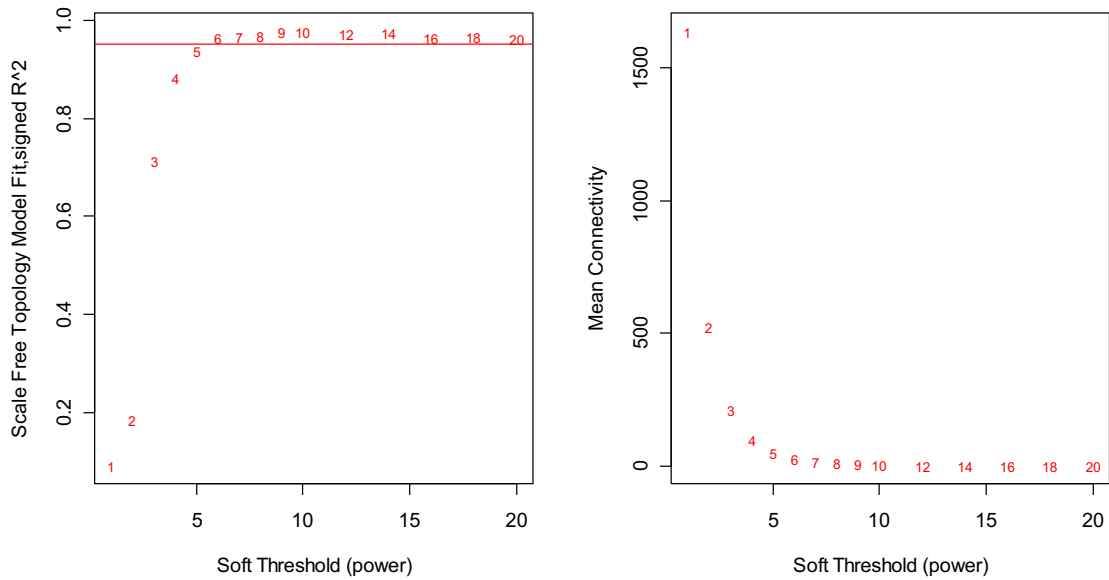
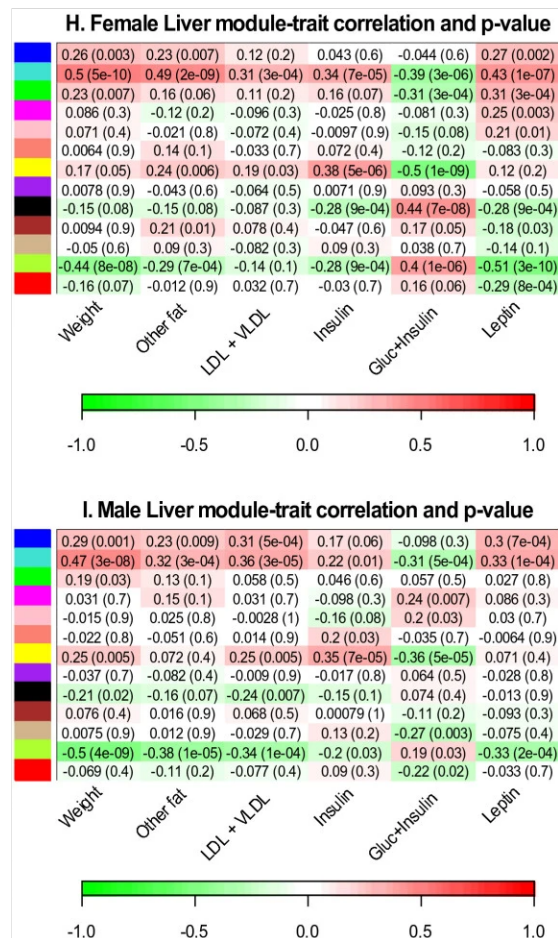


Figure 1.30. Définition de la valeur β en tenant compte de la topologie de type scale-free et la connectivité moyenne. Le premier graphe représente le modèle de la topologie de type scale-free (axe des y : R^2) et l'identification des différentes valeurs de β (*power*). Le second graphe représente la connectivité moyenne (axe des y) versus la valeur β (*power*). La ligne rouge sur le premier graphe correspond à $R^2 = 0,80$. Tiré de [WGCNA User Manuel for version 1.0.x](#).

Par la suite, la matrice de corrélation de Pearson va subir de nombreux changements afin d'identifier les modules de gènes dans les données étudiées. Premièrement, la matrice de corrélation de Pearson va être transformée en matrice adjacente puis, la fonction TOM (*topology overlap matrix*) va permettre de construire une matrice de dissimilarité utilisée pour la construction d'un cluster hiérarchique (fonction *hclust*) et ainsi obtenir un dendrogramme, où chaque module est défini par une couleur et un nombre. Ce dendrogramme permet de visualiser les modules de gènes ainsi que la distance entre chacun. Dans le cas où le nombre de modules est important (>50), car il y a plusieurs modules avec un petit nombre de gènes et qui sont très proches dans le dendrogramme, les fonctions *cutreeDynamic* et *MergeCloseModules* vont permettre de fusionner ces modules et ainsi redéfinir les modules, leurs couleurs et leurs nombres. Il est aussi possible de présenter les modules à travers un *HeatMap* où chaque module est représenté par la première composante principale obtenue avec la méthode eigengene. Une valeur de corrélation (fonction *cor*) et une valeur p (fonction *corPvalueStudent*) sont calculées permettant ainsi d'identifier le niveau d'association entre le module et chaque condition d'étude (exemples : type cellulaire, état cellulaire, trait clinique, etc.) (**Figure 1.31**).

Figure 1.31. HeatMap de modules de gènes significativement associés à des traits cliniques chez les souris femelles et males. Module représenté par un carré de couleur à droite du graphe.



L'échelle de couleur représente la corrélation entre le module de gènes et le trait clinique

associé. Corrélation et valeur p, entre parenthèses, dans le carré à l'intersection du module de gènes et le trait clinique associé. Tiré de (154).

1.3.3 Analyse de l'expression génique et sa régulation dans les cellules immunitaires : Projet ImmGen

L'humain et la souris partagent près de 90% de leurs gènes ainsi qu'une forte conservation entre leurs transcriptomes. La comparaison des transcriptomes dans ces deux espèces pour 7 types cellulaires immunitaires, a montré que la conservation de l'expression génique se trouve entre 51 et 70% (155). Une partie des gènes conservés entre ces deux espèces ont un profil d'expression distinct. Par exemple la cytokine IL15, qui cible les cellules NK et T, est exprimée par les monocytes, les cellules B et les granulocytes chez l'humain mais pas chez la souris (155). Au total, 169 gènes orthologues ont un profil d'expression distinct entre les deux espèces. Ces différences étant minimales, l'étude des cellules immunitaires chez la souris permet de donner des pistes d'études chez l'humain.

La collaboration de laboratoires d'immunologie et de biologie analytique ont créé le projet ImmGen (*The immunological Genome Project*) (156, 157). Ce consortium a pour objectif d'étudier l'expression génique et sa régulation dans les cellules immunitaires en passant de la cellule souche multipotente jusqu'aux cellules immunitaires actives. Il s'intéresse principalement à l'expression d'ARN, la structure de la chromatine, l'organisation nucléaire ainsi que les relations ARN-protéine. L'étude de l'expression des gènes a été réalisée sur l'expression globale des gènes mais aussi sur les transcrits de ces mêmes gènes. Deux types d'étude distinctes ont permis d'observer des différences d'expression de transcrits entre les types cellulaires mais aussi entre les différents stades cellulaires au sein d'un type cellulaire (138, 139). Une première étude parue en 2013 a comparé les niveaux d'expression des exons d'un même gène entre les cellules T CD4+ et les cellules B. Ils ont pu observer un certain nombre de gènes exprimés dans les deux types cellulaires mais pour un ensemble unique d'exons (**Figure 1.32**) (138). Le calcul de ratio d'exon absent (*skipped*) exprimé dans les cellules B versus les cellules T CD4+ identifie 11 gènes avec au moins un transcrit spécifiquement associé aux cellules B et inversement, 21 gènes ont au moins un transcrit spécifique aux cellules T CD4+. Dans une seconde étude, le profil d'expression des gènes a été étudié dans les différents stades de différenciation au sein d'un type cellulaire, récapitulé dans le **Tableau 1.11** (139).

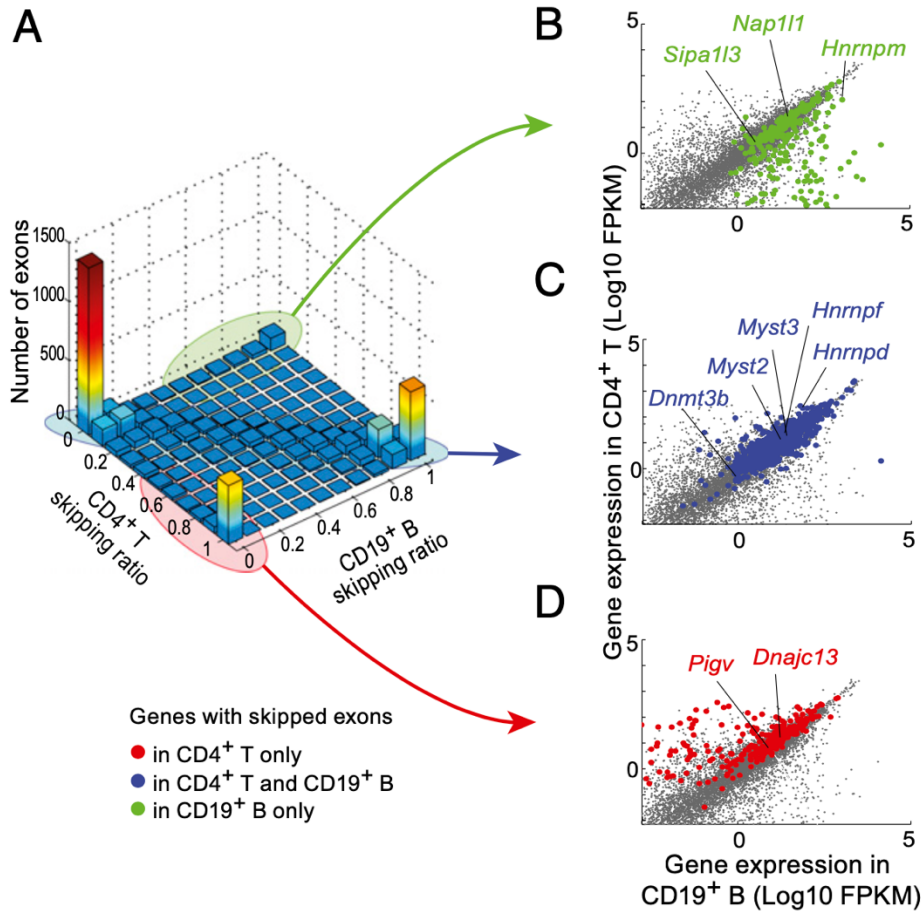
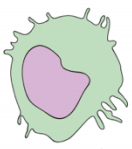

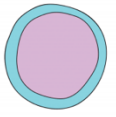
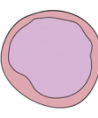


Figure 1.32. Représentation de l'abondance relative des exons dans les cellules T CD4⁺ et B. **A.** Ratio défini par le nombre de lecture de Séquençage d'ARN avec l'exon absent (*skipped*) relatif au nombre total de *read* du gène représenté graphiquement pour les cellules B et T CD4⁺. **B-C-D.** Expression des gènes intégrant les gènes avec des exons exclusivement absents pour les cellules B ou T CD4⁺ (respectivement B-vert et D-rouge) ou avec des exons équitablement absents dans les deux types cellulaires (C-bleu). Tiré de (138).

Tableau 1.11. Récapitulatif d'analyse d'épissages alternatifs, d'isoformes de transcrits différentiellement exprimés dans différents stades cellulaires et maladies associées aux erreurs d'épissage ou mutations. Tiré de (139).

	Genome-wide analysis	Described examples	Disease association
 Macrophages	<ul style="list-style-type: none"> 791 AS events SF3A1 sensitive ⁽⁵¹⁾ 81 AS events LPS sensitive ⁽⁵¹⁾ Unspliced transcripts remain at chromatin ⁽¹²⁾ 	TLR4 ⁽⁵⁵⁾ MyD88 ⁽⁵²⁻⁵⁴⁾ IL-6 ^(40, 41) IL-10 ^(40, 41)	<ul style="list-style-type: none"> AML and SFPQ ^(38, 39) AML and U2AF1 ⁽⁹⁷⁾
 Dendritic cells	<ul style="list-style-type: none"> link between maturity states and splicing patterns ⁽⁴⁷⁾ widespread variability in AS events on a single-cell level upon LPS stimulation ⁽⁴⁷⁾ 	MyD88 ⁽⁵²⁻⁵⁴⁾ IL-6 ^(40, 41) IL-10 ^(40, 41)	<ul style="list-style-type: none"> AML and SFPQ ^(38, 39) AML and U2AF1 ⁽⁹⁷⁾
 B cells	<ul style="list-style-type: none"> More than 90% of multi-exon genes alternatively spliced ⁽⁴⁶⁾ 	IgM/IgD ⁽⁵⁹⁻⁶³⁾ IgM ⁽⁶⁴⁻⁶⁷⁾ IgE ⁽⁶⁸⁻⁷²⁾	<ul style="list-style-type: none"> CLL and SF3B1 ⁽⁹³⁾
 T cells	<ul style="list-style-type: none"> 1,319 AS events upon ligation of the TCR ⁽⁴⁵⁾ 1,575 AS events upon TCR and CD28 stimulation ⁽⁴⁵⁾ Enrichment of AS in genes related to immune response ⁽⁴⁴⁾ Unspliced transcripts remain at chromatin ⁽¹³⁾ 	CD44 ⁽⁸¹⁻⁸³⁾ CD45 ^(80, 84-89) MALT1 ⁽⁷⁷⁾ FOXP3 ^(78, 79)	<ul style="list-style-type: none"> Multiple sclerosis ⁽⁸⁹⁾ Type-1 Diabetes ⁽¹⁰¹⁾

1.3.4 Analyse de la régulation de l'expression génique : Projet ENCODE

En 2003 est né le consortium ENCODE (*The Encyclopedia of DNA Elements*), une collaboration internationale de groupes de recherches fondée par le NHGRI (*National Human Genome Research Institute*). Ce consortium a été créé afin d'annoter l'ensemble du génome et du transcriptome de 4 espèces, dont l'humain et la souris (**Figure 1.33**). Pour définir les méthodes expérimentales et analytiques adaptées pour identifier et cataloguer les éléments fonctionnels, une phase pilote a été réalisée sur 1% du génome humain (158). Le 1% du génome étudié par le projet ENCODE correspond à 30 méga bases (MB) composé de 44 régions génomiques dont 14 de ces régions renferment des connaissances biologiques importantes et les 30 autres régions ont été choisies aléatoirement. Ces régions génomiques ont été analysées dans plus d'une centaine de lignées cellulaires dont des lignées provenant de cellules B, de cellules T CD4+ et de lymphoblastes. Cette phase pilote s'est intéressée à la transcription du génome (cartographie et identification des transcrits, des pseudogènes, des ARN ne codants pas pour une protéine, des transcrits primaires n'ayant pas subis d'épissage alternatif), la régulation de la transcription (cartographie des régions génomiques régulatrices, des SIT, des liaisons à l'ADN de protéines régulatrices et l'annotation des histones sur leur composition, modification et déplacement) et la réplication de l'ADN

(cartographie de la réplication de l'ADN dans le temps). L'analyse croisée de ces trois domaines d'étude du génome humain ont permis d'étudier la relation entre l'architecture de la chromatine et des domaines génomiques ainsi que l'évolution acquise et la variabilité des populations. Ces méthodes expérimentales et analytiques n'ont pas été réalisées pour l'ensemble des lignées cellulaires de la phase pilote. Cette phase pilote a pris fin en 2012 et depuis 2013 jusqu'à nos jours, le consortium ENCODE applique l'ensemble des méthodes expérimentales et analytiques sélectionnées lors de la phase pilote à l'ensemble du génome (159). La plateforme www.encodeproject.org loge l'ensemble des données brutes obtenues par les méthodes expérimentales et le site web UCSC permet la visualisation de ces résultats lors de la navigation à travers le génome humain et de souris. Des projets fondés par le NIH (*National Institutes of Health*), tels que les projets Roadmap (*The Roadmap Epigenomics Mapping Consortium*), GGR (*The Genomics of Gene Regulation*) et community, ont été ajoutés à la plateforme du projet ENCODE pour compléter les données obtenues (**Figure 1.34**) (160, 161).

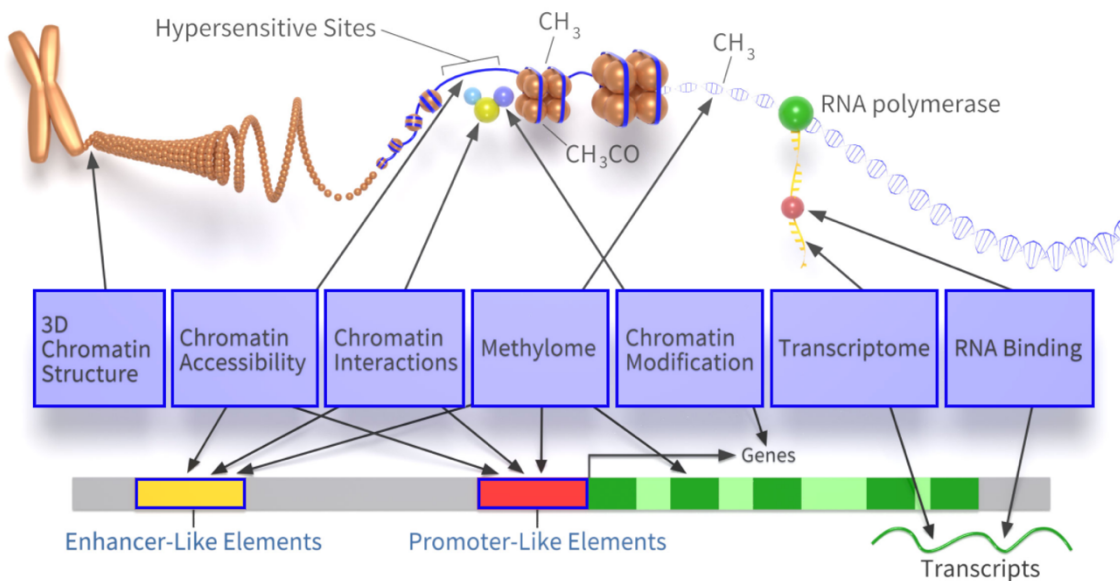


Figure 1.33. Schématisation des caractéristiques étudiées par le projet d'annotation génomique et transcriptionnelle. Tiré de Projet ENCODE

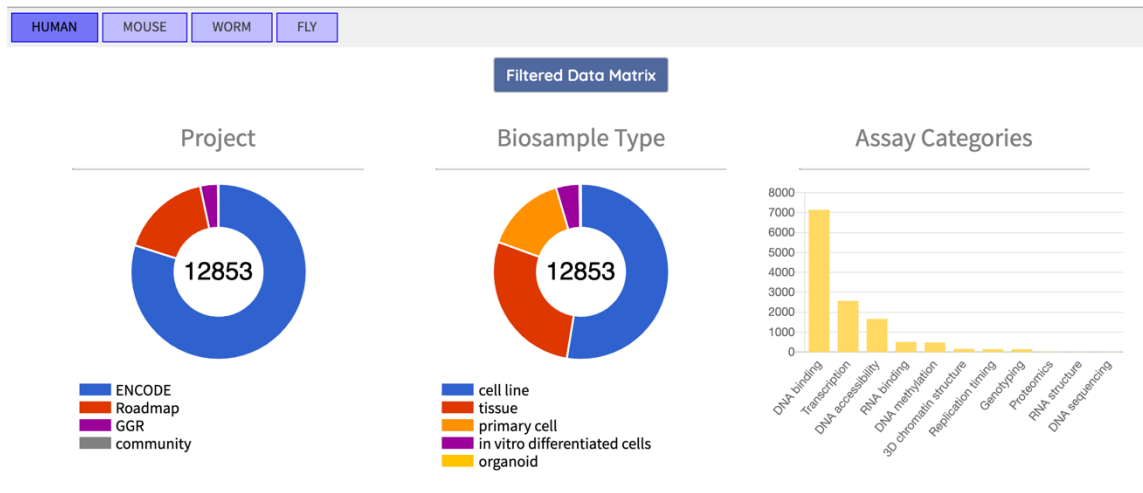


Figure 1.34. Représentation graphique des études du projet encyclopédie d'ENCODE chez l'humain.
Tiré de [Projet ENCODE](#)

Le consortium du projet ENCODE s'est intéressé aux éléments de régulation de l'ADN chez l'humain (162). Plus de la moitié des [méthodes expérimentales](#) ont été réalisées dans les [lignées cellulaires](#) et moins d'un quart dans les [cellules primaires](#) (**Figure 1.34 et Tableau 1.12**). Le tableau 1.7 montre que la majorité des expériences réalisées dans les cellules immunitaires identifient les régions ouvertes de l'ADN soit, accessibles par les protéines liants l'ADN tels que les FT. Mais en contre partie seulement 4 expériences étudiant la liaison d'un FT donné à l'ADN ont été réalisées dans les cellules primaires exclusivement immunitaires (monocytes CD14+, cellules B et cellules NK). Dans les lignées cellulaires de [type immunitaire](#) (cellules B et T), 247 expériences ChIP-seq ont été réalisées pour identifier les sites de liaison des FT. Plus de 240 de [ces expériences](#) ont été faites dans des lignées de cellules B pour 156 FT (163-165). Cela permet d'avoir une vue d'ensemble des potentielles régions de liaison de ces FT et ces données peuvent être extrapolées aux cellules B primaires. La méthode ChIP-seq permet d'identifier les régions d'ADN où se lie un FT étudié mais cette méthode n'est pas précise et les morceaux d'ADN liés au FT peuvent atteindre jusqu'à 300 pb alors qu'en moyenne un site de liaison d'un FT fait une dizaine de pb (166, 167). Le croisement des données provenant de plusieurs méthodes expérimentales réalisées au sein du consortium ENCODE a permis d'identifier certains sites de liaison d'un FT donné dans le génome d'un type cellulaire spécifique. L'ensemble de ces résultats sont présentés sur la plateforme [Factorbook](#) créée par le consortium du projet ENCODE (168-171).

Tableau 1.12. Projet encyclopédie d'ENCODE - Expériences réalisées dans les cellules immunitaires primaires. Tiré de [Projet ENCODE](#).

BIOSAMPLE ←	ASSAY →	Assays														
		DNase-seq	Mint-ChIP-seq	Histone ChIP-seq	Control Mint-ChIP-seq	Control ChIP-seq	polyA plus RNA-seq	TF ChIP-seq	WGBS	microRNA-seq	total RNA-seq	CAGE	polyA minus RNA-seq	small RNA-seq	RNA microarray	Repli-chip
primary cell		358	236	171	57	38	10	4	4	3	3	2	2	2	1	1
T-cell		60	6	11	1	3	1		1							1
activated CD4-positive, alpha-beta T cell		83														
CD4-positive, alpha-beta T cell		46	3	18	1	4	2									
CD14-positive monocyte		7	24	21	7	3	2	1	1			1	1	1		
CD4-positive, alpha-beta memory T cell		1	35	15	7	4				1						
naive thymus-derived CD4-positive, alpha-beta T cell		3	31	16	7	3				1						
CD8-positive, alpha-beta T cell		22		19		4	2			1						
B cell		6		25		8	2	2	1			1	1	1		
activated CD8-positive, alpha-beta memory T cell		42														
CD8-positive, alpha-beta memory T cell		3	25	6	6	1										
immature natural killer cell			30		7											
naive thymus-derived CD8-positive, alpha-beta T cell			24		6											
CD4-positive, CD25-positive, alpha-beta regulatory T cell			17	6	5	1										
naive B cell		1	21		5											
activated T-cell neutrophil		13	6		1						3					
T-helper 1 cell		12		11		2		1								
T-helper 17 cell		4		7		2										1
effector memory CD4-positive, alpha-beta T cell		6		6		1										
natural killer cell		3		6		1	1		1							
CD4-positive naive resting alpha-beta T cell		6	3		1											
IgD-negative memory B cell			8		2											
regulatory T cell		5		4		1										
T-helper 2 cell		7														
CD8-positive naive resting alpha-beta T cell		5														
activated B cell suppressor macrophage			3		1											
inflammatory macrophage		4														
T-helper 9 cell		3														
activated CD4 positive, naive alpha-beta T cell		2														
activated CD4-positive, alpha-beta memory T cell		2														
activated CD8-positive, alpha-beta T cell		2														
central memory CD4-positive, alpha-beta T cell		2														
effector CD4-positive, alpha-beta T cell		2														
gamma-delta T cell		2														
T follicular helper cell		1														
T-helper 22 cell		1														

1.4 Mise en contexte et introduction des articles

Le chapitre d'introduction a permis de présenter le génome humain et les mécanismes impliqués dans sa régulation et dans la production de protéines, et plus spécifiquement, de ces différentes isoformes. La seconde partie de l'introduction a servi à présenter les cellules du système immunitaire, et principalement, leurs caractéristiques fonctionnelles qui permettent de les distinguer malgré le fait qu'ils partagent certaines propriétés et fonctions immunitaires. Pour finir, différents outils et projets étudiant le transcriptome ont été présentés.

Les chapitres 2 et 3 sont deux articles qui présentent chacun une approche différente, mais complémentaire, de l'étude du transcriptome. La première approche analyse la co-expression génique pour mieux expliquer les fonctions spécifiques et partagées entre les cellules immunitaires primaires ainsi que la régulation de l'expression de gènes co-exprimés dans les cellules B. Une seconde partie de cet article présente une analyse de l'expression des gènes codant pour des facteurs de transcription et leur implication dans l'établissement et le maintien de la différenciation cellulaire.

Le chapitre 3 présente une approche analysant les différences dans les profils d'expression des transcrits d'un gène afin d'identifier les transcrits ayant profil d'expression spécifique et liés à un ou plusieurs types cellulaires. Dans le cas où ce transcrit spécifique code pour une isoforme de la protéine qui lui est propre, une analyse de la structure et de la fonction de l'isoforme est entreprise, afin de déterminer si cette isoforme est spécifiquement liée aux types cellulaires où elle est exprimée. Cela permet d'identifier les protéines qui ont des isoformes avec des fonctions différentes et dont leur expression dépend du type cellulaire immunitaire où la protéine est exprimée.

2 Chapitre 2 – A transcriptome-based approach to identify functional modules within and across primary human immune cells

2.1 Contribution des auteurs

Auteurs : Saraï Mola, Sylvain Foisy, Gabrielle Boucher, François Major, Claudine Beauchamp, Mohamad Karaky, Philippe Goyette, Sylvie Lesage, John David Rioux.

Statut de l'article:

Publié dans *PLoSOne* le 29 mai 2020.

15 (5): e0233543. DOI: 0.1371/journal.pone.0233543

Conception et mise au point des expériences :

Saraï Mola, Gabrielle Boucher, Sylvain Foisy, Claudine Beauchamp, Philippe Goyette, Sylvie Lesage et John David Rioux.

Expériences effectuées, analyses des résultats et présentation des figures:

- 2 Claudine Beauchamp a recruté les individus et a réalisé l'isolation cellulaire ainsi que l'extraction d'ARN, soit 5% du projet
- 3 Le nettoyage des données Séquençage d'ARN et mesure de l'expression génique ont été fait par Sylvain Foisy, soit 5% du projet
- 4 Gabrielle Boucher a réalisé les analyses préliminaires et l'écriture du script utilisé avec l'outil WGCNA, soit 5% du projet
- 5 J'ai appliqué l'approche d'analyse avec l'outil WGCNA ainsi que l'étude des résultats, soit 80% du projet.
- 6 Les 5% restant ont été partagés entre les autres auteurs de l'article soit Mohamad Karaky et Philippe Goyette

Rédaction de l'article : J'ai rédigé 50% de l'article en collaboration avec mon directeur, John David Rioux, qui a rédigé les 50% restant.

2.2 Article

A transcriptome-based approach to identify functional modules within and across primary human immune cells.

Saraï Mola¹, Sylvain Foisy¹, Gabrielle Boucher¹, François Major^{2,3,4}, Claudine Beauchamp¹, Mohamad Karaky¹, Philippe Goyette¹, Sylvie Lesage^{5,6}, John David Rioux^{1,4,*}.

Published: May 29, 2020. <https://doi.org/10.1371/journal.pone.0233543>

¹ Centre de recherche, Institut de cardiologie de Montréal, Montréal, Québec, Canada

² Unité de recherche en ingénierie des ARN, Institut de recherche en immunologie et en oncologie, Montréal, Québec, Canada

³ Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec, Canada

⁴ Département de biochimie et médecine moléculaire, Université de Montréal, Montréal, Québec, Canada

⁵ Centre de recherche, Hôpital Maisonneuve-Rosemont, Montréal, Québec, Canada

⁶ Département de microbiologie, infectiologie et immunologie, Université de Montréal, Montréal, Québec, Canada

Keyword: human primary immune cells; RNA-Seq; functional modules, transcriptional control

2.3 Abstract

Genome-wide transcriptomic analyses have provided valuable insight into fundamental biology and disease pathophysiology. Many studies have taken advantage of the correlation in the expression patterns of the transcriptome to infer a potential biologic function of uncharacterized genes, and multiple groups have examined the relationship between co-expression, co-regulation, and gene function on a broader scale. Given the unique characteristics of immune cells circulating in the blood, we were interested in determining whether it was possible to identify functional co-expression modules in human immune cells. Specifically, we sequenced the transcriptome of nine immune cell types from peripheral blood cells of healthy donors and, using a combination of global and targeted analyses of genes within co-expression modules, we were able to determine functions for these modules that were cell

lineage-specific or shared among multiple cell lineages. In addition, our analyses identified transcription factors likely important for immune cell lineage commitment and/or maintenance.

2.4 Introduction

The human immune system protects us from microbes (bacteria, viruses, fungi, and parasites) that penetrate the physical and chemical barriers of the body. It fulfills surveillance functions in order to detect and eliminate aberrant cells that result from infection, cancer, and senescence. In contrast, the immune system is also at the heart of pathogenic syndromes and chronic diseases that involve virtually all organ systems (e.g. cardiovascular, gastrointestinal, respiratory, musculoskeletal), and are associated with significant morbidity and mortality (172-176).

Given the central importance of the immune system in human health and disease, extensive work by numerous research groups over the past 100+ years has provided a detailed functional understanding of the cells and organs within the immune system. This includes how the multiple different cell populations, which make up the complex immune system, differentiate from a common pluripotent progenitor cell, how different cells communicate through an extensive system of cytokines, chemokines, and associated receptors, and how signaling pathways can regulate cell response to various stimuli (177-179). This functional understanding has been made possible by a combination of human and animal model systems, sensitive and specific reagents, such as antibodies directed against cell surface and intracellular proteins, and the development of technological platforms to isolate and study-specific cell populations based on multi-parameter flow cytometry, fluorescence-activated cell sorting, and mass cytometry (80, 176).

Genome-wide transcriptomic-based approaches have been instrumental in the identification of molecular classifiers of leukemias, interferon and granulopoiesis signatures in the blood of autoimmune patients, cytokine response pathways of T cells in psoriasis and the activation of macrophages, to name a few (180-182). Transcriptomic analyses have also been employed to perform surveys of the different cell populations within the myeloid and lymphoid lineages of the mouse immune system in different contexts (e.g. age, differentiation and activation states, tissue location) (156, 183). More recently, the use of single-cell RNA sequencing to evaluate the transcriptome has enabled the identification of previously unrecognized rare cell subsets of the human immune system (184-187). Together, these studies suggest that genome-wide transcriptomics is an important tool for studying how the human immune system responds to different conditions (e.g. stimuli, genotypes, health status, response to therapy).

The identification of gene sets that robustly represent different conditions has the potential to become powerful biomarkers, even without fully understanding the impact of the observed transcriptomic changes. The functional interpretation of changes in transcriptome between different conditions, however, remains an important challenge. A common approach to addressing this problem is to first perform gene co-expression analyses, for identifying gene sets that potentially are involved in common biological functions, followed by annotation-based analyses (e.g. enrichment of annotation terms, guilt-by-association) for attributing specific

functions to the observed gene sets (153, 188, 189). The success of this approach depends on numerous factors such as the complexity of the samples being tested, the true extent of the differences in the transcriptomes between the different conditions, the degree of experimental variability in the collection, processing, and sequencing of the samples (153, 188, 189).

In the current study, we were interested in testing this approach in the context of circulating human immune cells because of the well-characterized and distinctive nature of the different immune cells in terms of cell-specific markers (and thus genes expressed) and functions, as well as the relative ease by which highly purified immune cell populations can be isolated from peripheral blood. Specifically, we analyzed the transcriptomes of nine distinct immune cell populations from the peripheral blood of twelve healthy volunteers using a combination of global and targeted approaches to functionally annotate modules of co-expressed genes. We were able to attribute functions to most of the modules that were either cell population-specific or shared across more than one cell population. In addition, we identified TFs associated with lymphoid vs. myeloid lineages, T and B cell fates, many of which were previously known to be involved in lineage differentiation, as well as others that we propose as new candidates for having a role in lineage commitment and/or lineage maintenance.

2.5 Materials and methods

2.5.1 Subject selection criteria, ethics committee approval, and blood sample procedures

The project was submitted to and approved by the MHI Institutional Ethics Committee «Comité d'éthique de la recherche et du développement des nouvelles technologies ». Volunteers were then solicited via postings across the Montreal Heart Institute (MHI) and directed to the MHI Biobank Center for screening, informed consent form completion, and venipuncture. To be selected, the following criteria were used: (1) adult males with no history of (or ongoing) chronic inflammatory disease; (2) no current use of anticoagulants; (3) no use of oral steroids or cyclosporine in previous four weeks; (4) no use of NSAIDs in previous three days. After signing the informed consent form, volunteers were assigned a randomly generated ID number by the MHI Biobank Center; they were then called for blood withdrawal according to the availability of the cell sorter platform. The assigned ID number was provided with the blood samples to ensure anonymity. We recruited 12 individuals over a period of nine months. Researchers can request access to the individual-level data from the current study by contacting the Montreal Heart Institute ethics committee at the following institutional email address: cer.icm@icm-mhi.org.

2.5.2 Blood collection and cellular isolation

For lymphocytes and monocytes isolation, blood was collected in Vacutainer CPT Mononuclear Cell Preparation Tube - Sodium Citrate (BD Biosciences, San Jose, CA). Centrifugation was performed according to the manufacturer's instructions. Mononuclear cells were collected and washed three times in saline buffer, then filtered through a 70um mesh cell strainer. The cells were further purified with anti-CD14 coated microbeads using LS columns and the QuadroMACS separator stand (Miltenyi, Cambridge, MA), according to the manufacturer's instructions. Non-specific binding of antibodies was prevented on both CD14+ and CD14- enriched fractions by

using FcR blocking reagent (Miltenyi). A portion of each fraction was used for immunophenotyping.

The CD14⁺ enriched mononuclear fraction was stained with a goat anti-mouse IgG2a antibody (Serotec, Kidlington, United Kingdom). CD14⁻ enriched mononuclear fraction was stained with antibodies against CD3 (clone UCHT1), CD4 (clone RPA-T4), CD8 (clone RPA-T8), CD19 (clone HIB19), TCR $\gamma\delta$ (clone B1) (Biolegend, San Diego, CA) and CD56 (clone B159, BD Biosciences). Monocytes, CD4⁺ T cells (CD3⁺CD4⁺CD19⁻TCR $\gamma\delta$ ⁻), CD8⁺ T cells (CD3⁺ CD8⁺CD19⁻TCR $\gamma\delta$ ⁻), TCR $\gamma\delta$ T cells (CD3⁺CD19⁻TCR $\gamma\delta$ ⁺), NK cells (CD3⁻CD19⁻CD56⁺) and B cells (CD3⁻CD19⁺) were sorted on a FACSAria™ III cell sorter (BD Biosciences) based on cell size, granularity, doublet exclusion, live/dead marker and surface expression of targeted molecules. Purity was > 94% (median 98.1% with 97.3% and 98.9% for the 1st and 3rd quartiles, respectively).

For neutrophil isolation, blood was collected in Vacutainer® EDTA collection tubes (BD Biosciences). CD15⁺ fraction was enriched with StraightFrom™ Whole Blood CD15 MicroBeads using Whole Blood columns with the QuadroMACS separator (Miltenyi), according to the manufacturer's instructions. Non-specific binding of antibodies was prevented on the enriched CD15⁺ fraction by using FcR blocking reagent (Miltenyi). Cells were stained for CD14 (clone M5E2) and CD16 (clone 3G8) (Biolegend). Neutrophil population (CD14⁻CD16⁺) was sorted on a FACSAria™ III cell sorter (BD Biosciences) based on cell size, granularity, doublet exclusion, live/dead marker and surface expression of targeted molecules. Purity was > 99%.

2.5.3 Macrophage differentiation and activation

CD14⁺ enriched mononuclear cells were frozen in the vapor phase of liquid nitrogen until the day of culture. After thawing, the cells were cultured at 37°C 5% CO₂ in RPMI 1640 medium with GlutaMAX™, 10% heat-inactivated fetal bovine serum, 100 units/mL penicillin, 100 µg/mL streptomycin (ThermoFisher Scientific, Mississauga, Canada) and 50ng/ml M-CSF (Millipore, Etobicoke, Canada) for a total of 8 days, with media renewal every 2 days, to obtain macrophages. Activated macrophages were also generated by stimulating the macrophage cultures with 1ug/ml LPS (Sigma, Oakville, Canada) during the last 24h of culture.

2.5.4 Immunophenotyping

Following FcR blocking reagent (Miltenyi) step, CD14⁺ and CD14⁻ enriched mononuclear cells were stained with antibodies described in **S2.1 Table**. *In vitro* generated macrophages were blocked with FcR blocking reagent (Miltenyi) and stained with antibodies listed in **S2.2 Table**. Cells were analyzed on a BD™ LSR II flow cytometer (BD Biosciences). Final immunophenotyping results were generated using FlowJo software (FlowJo LLC, Ashland, OR) according to size, granularity, live/dead marker and surface expression of targeted molecules. Results are summarized in **S2.1 Fig**.

2.5.5 RNA extraction and sequencing

The total number of samples obtained for each population from each individual is summarized in **S2.3 Table**. Total RNA from all cell samples was isolated using a Qiagen RNeasy Plus Mini kit (Qiagen, Toronto, Canada) with an additional step of RNase-free DNase set (Qiagen) DNaseI

digestion; quantification, as well as quality control, were performed on Agilent BioAnalyzer 2100 (Agilent, Santa Clara, CA) using an Agilent RNA 6000 Nano Kit, and aliquots for sequencing were prepared and kept at -80°C until further processing.

RNA sequencing was performed at McGill/Genome-Québec Innovation Center. Total RNA was extracted and all, but one sample had RNA integrity number (RIN) greater than 8 (RIN ranged between 8.4 and 10; the one sample with RIN<8 was excluded). RNA samples were transformed into barcoded DNA libraries using Truseq Stranded mRNA library preparation kits (Illumina, San Diego, CA). Paired-end sequencing, generating 2x100bp reads, was performed on an Illumina HiSeq2000 with raw FASTQ sequences downloaded from the platform's server for local pre-processing and analysis.

2.5.6 Bioinformatic processing of sequence files

Raw FASTQ formatted sequence files were collected from the McGill/Genome-Québec Innovation Center sequence service. Primary QC analysis was performed using FastQC (v0.10.1) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Following a manual inspection of all FastQC result files to check for low-quality samples, raw sequence files were trimmed of low-quality bases from reads and removing low-quality reads altogether as well as removing potential adapter contamination with Trimmomatic (v0.33) (190) using a modified adapter FASTA file with the following parameters: ILLUMINACLIP with seed mismatches set to 2, palindrome clip threshold set to 30 and simple clip threshold set to 15, TRAILING set to 20 and MINLEN set to 50. Quality- and adapter-trimmed FASTQ files were inspected again with FastQC to confirm global quality.

Alignments were performed with TopHat 2 (v2.1.0) (124) using iGenome index and annotation files built from build 38 of the human genome (Illumina). Subsequent filtering of the aligned reads was done using SAMTools (v1.6) (191) to keep only properly-paired and single-mapped reads in the final alignments. Summary statistics from all steps for each cell type are shown in **S2.3 Table**.

For gene expression analysis, R/Bioconductor was used with a variety of additional analysis modules. Alignment files were processed with easyRNASeq (v1.8.8) (192) to measure gene abundance and create the gene expression matrix of raw read counts that was further processed using edgeR (v3.12.1) (193, 194) and Limma (v3.26.9) (195). edgeR package was used to assemble the expression count matrix and annotations with DGEList function and to filter out non- and low-expressed genes based on a cpm threshold. edgeR was also used to obtain per sample normalization factors. The function voom from the Limma package was used to normalize data from gene expression for library sizes scale with TMM factors and then to stabilize variance in the data. The various cell types collected in the study have expression profiles (e.g. neutrophils) with too much dissimilarity to allow for a normalization approach based on the hypothesis of relatively similar expression, for example, the normalization by quantile (**S2.2 Fig**) [29].

2.5.7 Principal component and hierarchical clustering analyses

Analyses of transcriptomic data using Principal Component Analysis (PCA) and hierarchical clustering approaches were performed to provide a global view of the transcriptomic data. These approaches enabled us to identify similarities between samples and structure in the dataset. This allowed a first assessment of the underlying principal sources of variation. After filtration and normalization, principal components were computed on all detectable transcripts with *prcomp* in R thus representing multiplicative differences between samples. The first few PCs (196-198) were investigated because they represented large axes of variation, using graphical display (**Fig 2.1**). In complementary, unsupervised hierarchical clustering was used to investigate the grouping of all samples. For this analysis on the entire normalized transcriptome, Pearson correlation was chosen to evaluate the distance between samples with R base functions, *cor*, and *dist*. Hierarchical clustering was used to build our data dendrogram with *hclust* R function (**Fig 2.2A**). Finally, ANOVA was used to determine the percent of gene expression variance explained by cell types. The expression for each gene was centered and scaled to a mean and variance of respectively 0 and 1. Analysis of variance was performed on each gene to obtain the sum of squares (SS) explained by cell types. Multivariate variance explained was estimated by adding explained SS from all genes and then dividing by the total SS.

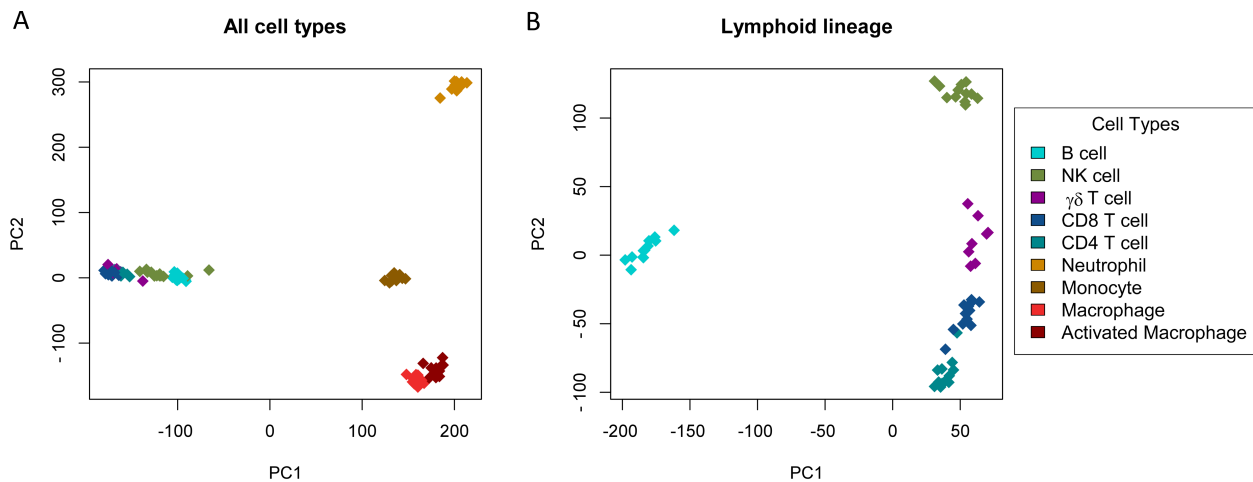


Figure 2.1. Transcriptomic data distinguish immune cell types. Representation of the first two principal components on RNA-seq data from all individuals for all samples (Panel A) and for lymphoid samples alone (Panel B). The R function *prcomp* was used to perform principal component analysis, and each symbol represents an individual RNA-seq sample.

<https://doi.org/10.1371/journal.pone.0233543.g001>

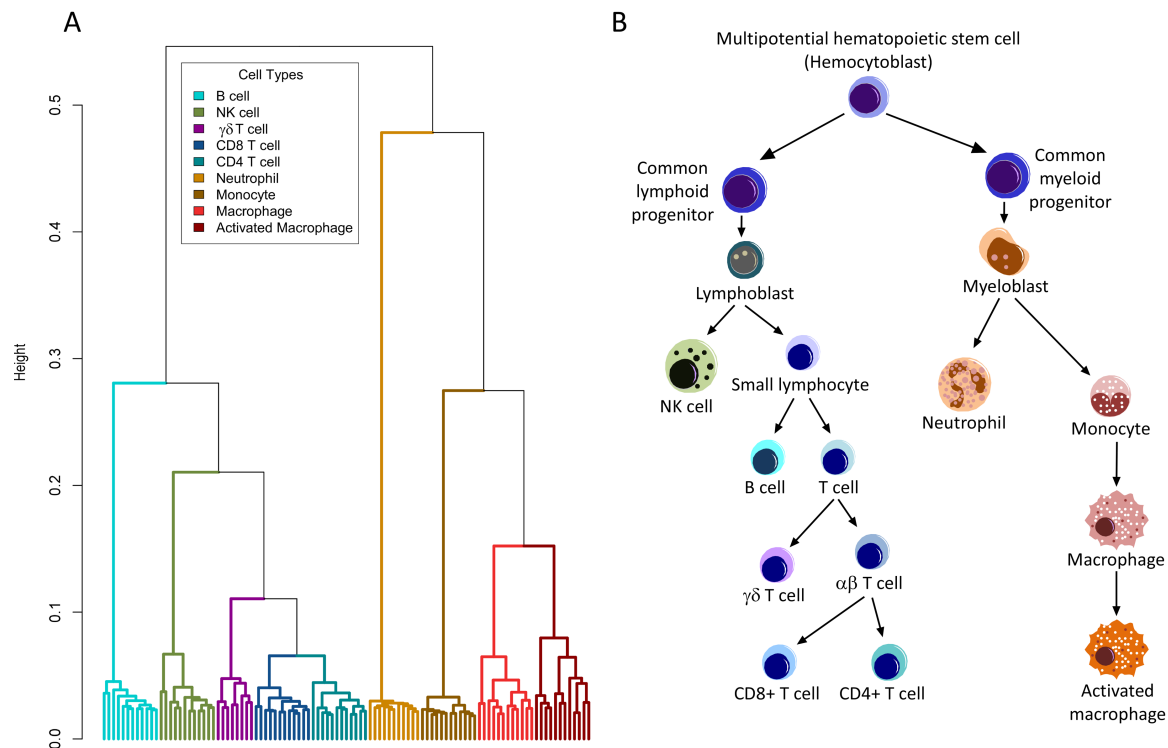


Figure 2.2. Unsupervised hierarchical clustering of RNA-seq data reflects hematopoietic differentiation. A dendrogram representation of the unsupervised hierarchical clustering of RNA-seq data from across all individuals and cell types (Panel A) has a structure that closely resembles the known differentiation scheme of immune cells (Panel B). Each dendrogram leaf represents a single RNA-seq sample and Y-axis represents the distance based on Pearson correlation. Black arrows represent a decision in the known differentiation scheme of immune cells. <https://doi.org/10.1371/journal.pone.0233543.g002>

2.5.8 Definition of co-expression modules

Starting with the hypothesis that genes with a similar function will often be co-expressed, we built co-expression modules with the R package WGCNA (v1.64-1) using the RNA-Seq data from our nine primary immune cell types. The objective of this tool is to cluster genes based on similar expression profiles between samples (132, 199). For building co-expression networks with WGCNA, we choose a signed topology to compute gene co-expression similarity. Signed topology makes it easier for interpretation purposes or relevant biology, such as making distinction gene repression and activation (153). A connectivity matrix between genes was then computed that describes how strongly genes are connected to all others. For that purpose, a “soft thresholding” parameter β must be chosen (199). This parameter is an exponent to the correlation matrix that determines the emphasis put on higher vs lower correlations. The *pickSoftThreshold* function was used to help in choosing β . After analysis of network topology for various β values, a value of 12 was selected as a good trade-off between scale-free topology and connectivity (S2.3 Fig). Therefore, to compute dissimilarities between genes, we used the WGCNA functions *adjacency* and *TomsimilarityFromExpr* (Topology Overlap Matrix Similarity From Expression) with the following parameters: Pearson correlation, signed topology and a β

of 12. Based on the TOM dissimilarity measure, *hclust* function was used to build a hierarchical clustering of the genes. Finally, *cutreeDynamic* function with *minClusterSize* and *deepSplit* parameters to 20 and 2, respectively (200), was used to cut the hierarchical tree into gene modules (**S2.4 Fig**). To merge closely related modules among our 63 modules, *MergeCloseModules* function was applied with *cutHeight* parameter of 0.05. Each module was represented by its first principal component using the function *moduleEigengenes*. Correlation and corresponding p-values of association between a given module and a cell type were then evaluated from these principal components using *cor* and *corPvalueStudent* functions from the WGCNA package.

2.5.9 Global and targeted functional annotation

We used a combination of global and targeted analyses of gene annotations to identify potential biological functions represented in each module. For a global analysis of genes within each module, we used DAVID (Database for Annotation, Visualization, and Integrated Discovery) (201) to test for enrichment of functional annotations, specifically using the annotation from Gene Ontology (GO) terms for Biological Processes, Molecular Functions, and Cellular Components, Up-Keywords, Interpro and KEGG, (202-205). In **S2.4 Table**, we report all annotations for which the enrichment was \geq two-fold and a $P < 0.05$.

For the targeted analyses, we undertook a more detailed approach taking advantage of the information within GeneCards (206), PubMed (207) and Google Scholar (208) databases, using gene symbols as query parameters. This approach was applied to two gene lists per module; the first list consisted of the set of genes most highly expressed in each module, with the hypothesis that the highly expressed genes may provide insight to important functions within each module. Specifically, we examined the known functions of the top 2% expressing genes (corresponding to the 98th percentile) which we called these the *Top Expressing Genes* (TEGs). The second list consisted of the genes that had the most important impact on the first principal component, regardless of their level of expression. Specifically, we ranked genes based on their contribution to the first principal component of the module in question with the objective of examining the known functions for the genes that accounted for 20% of PC1, or the top 20 genes if this exceeded 20 genes, we called these the

). For modules 22, 26, 38, 41, we also performed this detailed functional annotation to the entire set of genes within the modules in order to compare with the results obtained from these targeted analyses.

2.5.10 Analyses of transcription factors and transcription factor binding sites in B cells

Genes within modules were annotated as TFs based on the catalog of human TFs published by Lambert and colleagues (26). In this resource, based on a collection of databases including TRANSFAC, HT-SELEX, UniProbe, and CisBP to identify human TFs and their binding site motifs (209-218), a TF was defined as a protein capable of a) binding DNA in a sequence-specific manner and b) regulating transcription. In total, they identified 1,639 human TFs, and binding motifs for two-thirds of these (see <http://humantfs.cbr.utoronto.ca/index.php>) Transcriptional

control of genes within each module was analyzed in the following fashion: (1) the transcription start site (TSS) was located for each gene using the GRCh38 Ensembl database using the biomaRt tool (219); (2) promoter regions were then defined as -1000 to +500 base pairs around these TSS; while somewhat arbitrary, this is a commonly used definition of the proximal promoter (220, 221); and (3) these promoter regions were examined for the presence of TFBS as defined by ChIP-Seq data from the ENCODE (Encyclopedia of DNA Elements) Project (162). As this latter data was referenced to the human genome build GRCh37, TFBSs were converted to GRCh38 version using liftOver (222). This ENCODE dataset consisted of ChIP-Seq data for a total of 161 TFs in 80 immortalized cell lines. Of the nine immune cell populations within the current study, only the B lymphocyte population was represented in this dataset. For the modules associated with B lymphocytes, we restricted our analyses to the 76 TFs that were studied in the GM12878 immortalized B cell line. Enrichment of TFBS in the promoter regions within each of these modules was calculated with the hypergeometric distribution. Enrichment p-value threshold was 0.05.

2.5.11 Transcription factors potentially associated with cell lineages and states.

We examined all 1600 human TFs (26) for expression patterns across the cell types that were consistent with having a role in cell fate decisions, independent of co-expression modules described above. Following the differentiation scheme depicted in **S2.5 Fig**, we considered each branch point like a cell fate decision and looked for TFs that appeared to be specific for a given fate (e.g. myeloid vs. lymphoid; neutrophil vs. monocyte; B cell vs. T cell; etc.). Several TFs were very well known to be implicated in cell fate decisions (e.g. PAX5 for B cell differentiation; EOMES for differentiation of cytotoxic cells; FOXP3 for the differentiation of CD4+ T cells). Based on the expression pattern of these well-known TFs involved in differentiation of immune cells, we defined the following algorithm for prioritizing TFs potentially having a role in cell fate decisions (**S2.5 Fig and S2.5 Table**): (1) cell types resulting from a cell fate decision (e.g. lymphoid) should have an average number of read counts for the given TF at least 15-fold greater than the average number of read counts for cells from the same branch point that do not have the same fate (e.g. myeloid); (2) each cell type resulting from a cell fate decision should have an average number of read counts >100 for the given TF in order to reduce artifacts resulting from the greater variance at low read counts; (3) the cell type resulting from a cell fate decision with the lowest read count for the given TF should be at least four-fold greater than the cell type with the highest read count from the same branch point that does not have the same fate (e.g. IKZF3 in CD4+ T cells vs LPS-activated macrophage). Finally, we identify which of these candidate TFs are known to be implicated in associated cell fate decisions based on a search of the PubMed, GeneCards, Entrez, UniProt, SwissProt.

2.6 Results

2.6.1 Data generation: isolation of immune cells and RNA-Seq analysis

For the current study, we isolated neutrophils, B cells, CD4+ and CD8+ T cells, NK cells, $\gamma\delta$ T cells and monocytes from the peripheral blood of 12 healthy volunteers. In addition, we differentiated *in vitro* the monocytes into mature macrophages, as well as stimulated these

macrophages with LPS, thus generating a total of nine different cell populations. We obtained highly pure immune cell populations, and immunophenotyping analyses confirmed the representation of subpopulations that can be expected in the circulation of healthy adults (**S2.1 Fig**). A transcriptomic analysis of the total RNA extracted from these cells was performed using a paired-end sequencing approach, that resulted in a very good depth of coverage (2.48×10^7 read counts per sample on average; **S2.3 Table**), and thus can be expected to provide quantitative measures of transcript abundances even for less common transcripts (223)

2.6.2 Structure of transcriptomic data is a reflection of the known differentiation scheme for immune cell subsets

Prior to embarking upon in-depth analyses of this RNA-Seq data, we wanted to assess whether the patterns of gene expression could potentially be correlated to differences in functions between cell populations. Principal component analysis (PCA) revealed a clear separation of the cell types on the first two components, representing the nine populations into four very distinct clusters (neutrophils, lymphocytes, monocytes, and macrophages), with all individuals clustering tightly within each cell population (**Fig 2.1A**). This is concordant with the observation that 86% of the total multivariate variance is between the cell types and the first two components capture a large proportion of the total variance (43% and 25 % respectively). We also applied a similar PCA only on lymphoid cell types, which resulted in five distinct clusters representing the five different lymphoid populations: B cells, CD4+ and CD8+ T cells, NK cells and $\gamma\delta$ T cells (**Fig 2.1B**). Hence gene expression in our data set clearly represents more differences between immune cell populations than between individuals.

Using an unsupervised hierarchical clustering we found that the dendrogram reflected the differentiation scheme of pluripotent stem cells into myeloid and lymphoid cell types, except for B and NK cells, which were interchanged in our hierarchical clustering (**Fig 2.2**). Given these observations, we propose that there is a great potential to identify differences in gene expression between immune cell populations that will reflect the functional biology of these populations.

2.6.3 Defining potential functional groups using a combination of gene co-expression and functional annotation

Knowing that there is a significant correlation structure in the transcriptome captured by the RNA expression data sets, and those co-expressed genes are often related functionally (224), we hypothesized that defining modules of co-expressed genes would be a relevant starting point for exploring potential functional units relevant to these immune cells. We, therefore, used a weighted gene co-expression network analysis (WGCNA) approach to define modules of co-expressed genes (132). Using this approach, we obtained 45 modules with an average of 318 genes per module (ranging from 26 to 1,945 genes; **Table 2.1**). As can be seen in **Fig 2.3**, each module could be associated with one or more immune cell populations. For example, **modules 22 and 38** were highly associated with B cells ($r = 1.00$, $P = 1.97 \times 10^{-106}$; $r = 0.74$, $P = 8.05 \times 10^{-19}$; respectively) and **module 41** with B cells and monocytes ($r = 0.63$, $P = 1.41 \times 10^{-12}$; $r = 0.46$, $P = 8.32 \times 10^{-07}$; respectively), whereas **module 26** was strongly associated with NK, $\gamma\delta$ T and CD8+ T

cells ($r = 0.68$, $P = 1.71 \times 10^{-15}$; $r = 0.41$, $P = 1.78 \times 10^{-05}$; $r = 0.32$, $P = 9.80 \times 10^{-4}$; respectively), all cell populations that perform cytotoxic functions.

Table 2.1. Summary of key functions identified for modules of co-expressed genes.

Module ID	# Genes in module	Associated cell types ***									Main functions identified
		B	NK	γδ T	CD4+	CD8+	Neut	Mon	MΦ	LPS	
1	1212								0.3	0.23	Mitochondrial functions, ribosome biogenesis
2	1945										Broad range of basic functions
3	619	0.26		0.2	0.27	0.29					Transcriptional regulation via zinc finger proteins
4	593						0.48	0.32	0.35	0.39	Lysosomal function; neutrophil polarity; chemoattraction
5	489								0.62	0.69	Chemokine production, microtubule network; extracellular exosomes
6	748								0.49	0.43	Mitochondrial metabolism
7	452						0.68	0.44			Azurophil granules; cytoskeleton dynamics; ruffle
8	411	0.37	0.18	0.25	0.38	0.37					Transcriptional regulation (e.g. of cytokine genes) via zinc finger proteins
9	427							0.44	0.55	0.48	Endocytosis & ROS production
10	309						0.71			0.44	Actin remodeling & membrane trafficking, nicotinamide metabolism
11	296						0.96				Production of Azurophil granules
12	548						0.85	0.39			Cytoskeleton remodeling; cell polarity; chemokine signaling
13	401	0.33	0.22	0.23	0.32	0.32					Cell-cell adhesion; actin remodeling; interferon response
14	246	0.24					0.64				Ubiquitin-like conjugation; control of intracellular membrane traffic; transcriptional regulation
15	595						0.9				Defensin-specific granules
16	426						0.91				NET-associated histone production
17	414	0.34	0.21	0.24	0.32	0.32					Transcriptional regulations; control of immune cell proliferation
18	226	0.2	0.23	0.22	0.21	0.23	0.27				N.D.
(19)	321	0.36					0.35				Cytoskeleton remodeling
20	429							0.39	0.5	0.47	Lysosomal/Endosomal functions
21	388						0.61				Neutrophil polarization, trafficking, and exocytosis; ROS & NET production
22	195	1									Maintenance of B cell lineage; BCR signaling
23	637	0.35		0.22	0.34	0.34					Ribosomal functions
24	177		0.45	0.36	0.31	0.37					Control of Wnt signaling via beta-catenin TCF complex
25	147						0.42	0.33	0.28	0.32	Superoxide-generating NADPH oxidase activator activity
26	129		0.68	0.41		0.32					Cytotoxic granule composition; non-self-recognition
27	120		0.49	0.37	0.31	0.4					Modulation of receptor signal transduction
28	279							0.6	0.65		Bioenergetics (e.g. fatty acid oxidation, protein turnover)
29	93								0.98		Anti-microbial functions (chemokine production & zinc metabolism)
30	91			0.33	0.52	0.51					Regulation of T cell signaling
(31)	81				0.66	0.57					Modulation of immune response including via CD28, netrin 4 and semaphorin 5A
(32)	229						0.24		0.5	0.65	Vesicular trafficking (lysosomes, endosomes, exosomes)
33	67	0.34	0.18	0.23	0.26	0.28					N.D. (associated with lymphocytes)
(34)	61	0.32						0.27	0.47	0.45	Antigen presentation; vesicle function and trafficking; RAS signaling
(35)	58							0.54		0.68	Response to hypoxia; autophagy; regulation of inflammation
36	52							0.76		0.47	Endocytosis and membrane trafficking
37	46	0.39						0.28	0.36	0.36	Intracellular vesicle formation and function, including lysosomes
38	45	0.74							0.3	0.35	Regulation of BCR signaling
(39)	43		0.91	0.22							Golgi apparatus
40	121						0.88				Transcriptional regulation via zinc finger proteins
41	39	0.63						0.46	0.22		MHC class II antigen processing and presentation
42	38		0.26	0.19	0.22	0.22			0.28	0.32	Heterogeneous functions; IL7R signaling
43	35							0.92			N.D.
(44)	34						0.36			0.87	Interferon response
(45)	26	0.56					0.44	0.27			Regulation of intracellular membrane trafficking

Module ID: Module number; Module ID in parentheses: the associated functions are more speculative as enrichment is more limited; ***: only correlations between modules and cell types where $P < 0.05$ are indicated; B: B cells; NK: NK cells;

$\gamma\delta$ T: $\gamma\delta$ T cells, CD4+: CD4+ T cells; CD8+: CD8+ T cells; Neut: neutrophils; Mon: monocytes; M Φ : macrophages; LPS: LPS-activated macrophages. N.D.: Non-determine. <https://doi.org/10.1371/journal.pone.0233543.t001>

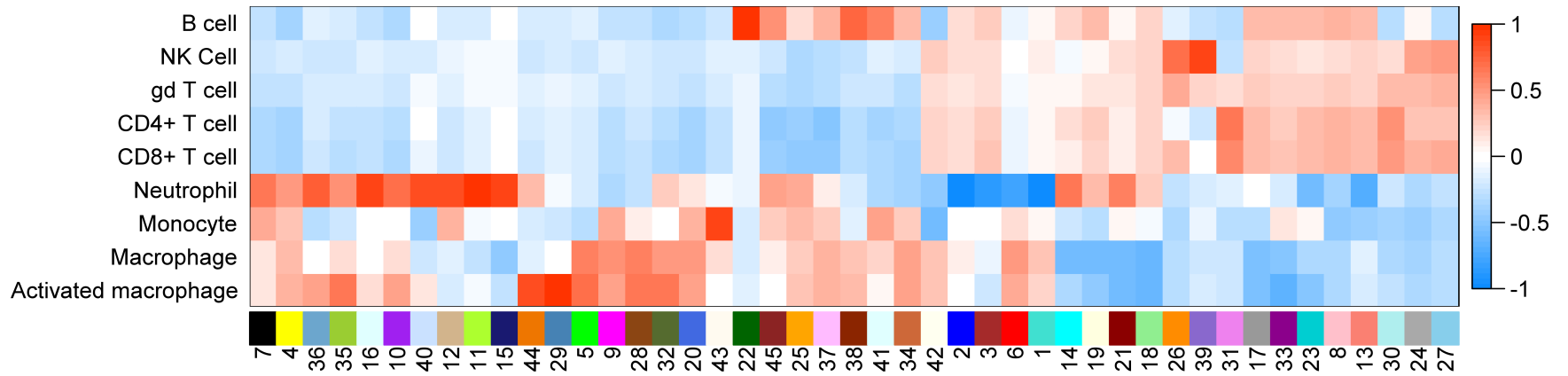


Figure 2.3. Heatmap of the correlation values of WGCNA modules with primary immune cell types. Columns represent modules computed with WGCNA and rows, primary immune cell types. Correlation values and *P* values are displayed in **S2.6 Fig.** <https://doi.org/10.1371/journal.pone.0233543.g003>

To gain a better understanding of the biology underlying these co-expression modules that were highly associated with one or more specific cell types, we used a combination of global and targeted analyses of gene annotations and identified potential biological functions represented in each module. For the global analysis, we tested for the enrichment of annotation terms within the Gene Ontology (GO), UniProt, InterPro and KEGG databases (202, 205). For the targeted analysis, we undertook two independent approaches. The first was to identify the set of genes most highly expressed in each module, with the hypothesis that the highly expressed genes may provide insight into important functions within each module. Specifically, we examined the known functions of the genes that are within the top 2% of genes expressed within the cell types associated with a given module. We call these the *Top Expressing Genes* (TEGs). The second targeted approach entailed the identification of the genes with the greatest impact on the first principal component for each module, with the hypothesis that these genes better represent the modules, regardless of the expression level. Specifically, we examined the known functions of the genes that can account for 20% of the variance captured by the first principal component; We call these the *Module Representative Genes* (MRGs). To assess whether these approaches could identify functions that are relevant to the broader set of genes within each module, we also examined the known function for the entire set of genes within **modules 22, 38 and 41**.

2.6.4 B cell-specific modules; cell activation and BCR engagements

As mentioned, **modules 22 and 38**, are highly associated with B cells (**Fig 2.3**) and contain 195 and 45 genes, respectively (**Table 2.1**). In our global analysis of **module 22**, there were eight annotation terms that were below 5% False Discovery Rate (FDR), and included “B-cell activation”, “Primary Immunodeficiency”, and “Immunoglobulin domain”, with a total of 17 shared genes being associated with one or more of the annotation terms mentioned previously (**S4 Table**). For **module 38**, 16 annotation terms with $P \leq 0.05$ were detected, but one was below 5% FDR, and none appeared to indicate functions specific to B cells (**S2.4 Table**).

Regarding TEGs in B cells, there were 12 TEGs within **module 22** and three TEGs within **module 38** (**S2.6 Table**). Querying of public databases (GeneCards, PubMed and Google Scholar) for the known functions of the TEGs within **module 22** revealed that many are involved in B cell receptor (BCR) structure and/or signaling (e.g. *IGLL5*, *CD79A*, *BANK1*, *MS4A1*, *FCRL1*) (225). There were also two genes encoding transcription factors (TF), *POU2AF1* and *PAX5*, associated with B cell biology (12, 226). Interestingly, *PAX5* also regulates the expression of *EBF1*, a gene within **module 38**, a TF crucial for B cell differentiation (227). In the case of **module 38**, two of the three TEGs have known functions in B cells: *CD22* inhibits BCR activation and *FCRLA* is involved in immunoglobulin assembly (228, 229) (**Fig 2.4**).

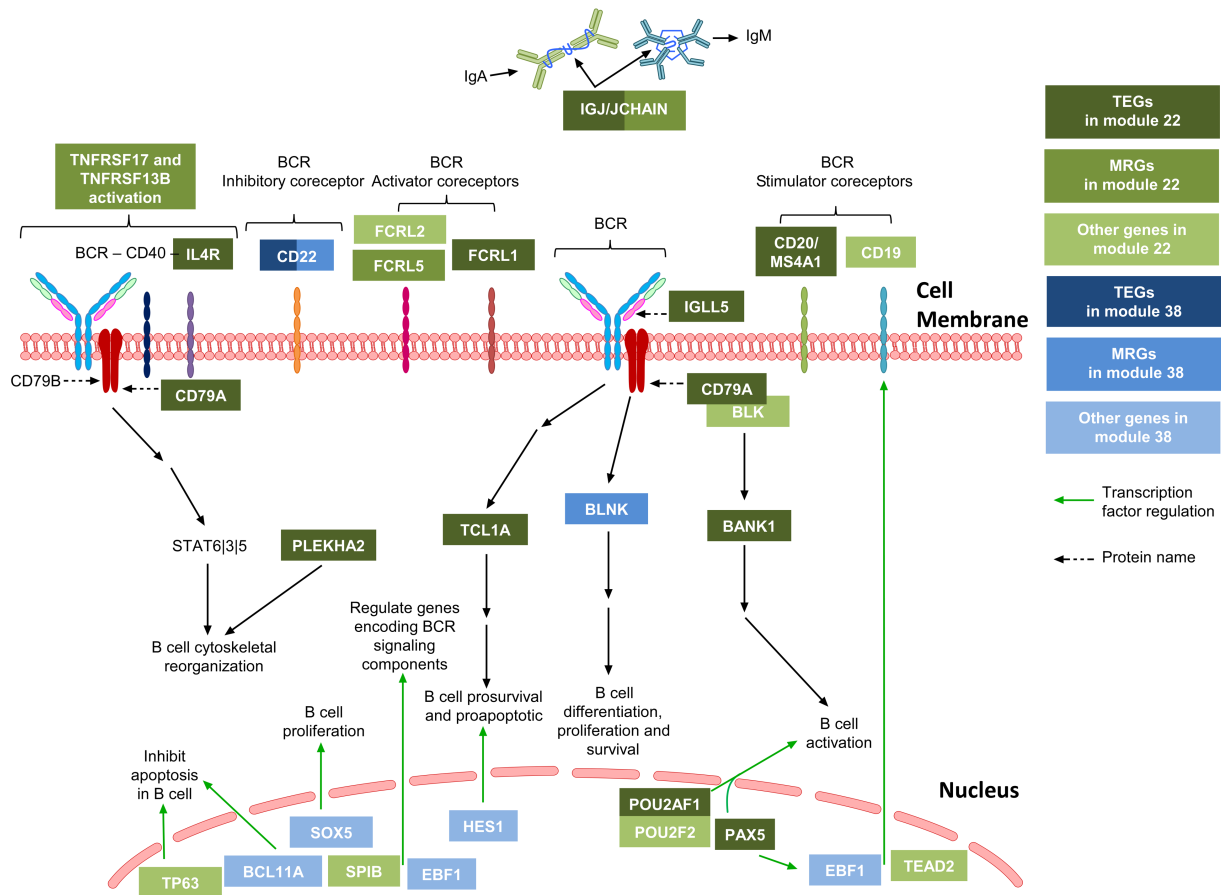


Figure 2.4. Biological functions ascertained from the analysis of genes within two B cell-associated modules. The genes within *modules 22 and 38* encode multiple genes that are central to B cell functions. These include components of B cell receptor (BCR) complex, such as inhibitory and stimulatory co-receptors CD22 and CD19, as well as TFs regulating the transcription of many of these. Genes in *modules 22 and 38* are represented in green and blue, respectively: TEGs in dark colors, MRGs in intermediate colors and other genes in light colors. To see the profile of gene expression mean of all genes of module 22 presented in this Figure refers to the heatmap in **S2.7 Fig.** <https://doi.org/10.1371/journal.pone.0233543.g004>

Regarding the MRGs, there were 28 within *module 22* and seven within *module 38* (S2.6 Table), with some overlap between the MRG and TEG lists: the *IGJ* gene (alias *JCHAIN*) encoding the immunoglobulin J chain in *module 22*, as well as the *CD22* and *FCRLA* genes in *module 38* (S2.6 Table). Among the MRGs from *module 22* are the *TNFRSF13B* (alias *B cell-activating factor*, *BAFF*) and *TNFRSF17* (alias *B Cell Maturation Antigen*, *BCMA*) genes, which play an important role in B cell activation. The remainder of these MRGs is poorly-characterized genes or non-protein coding RNAs (9 of 28). Of the poorly characterized genes, some are expressed in B cells (e.g. *HTR3A* (230)) or involved in B cell development/maturation (e.g. *KLHL14* and *VPREB3* (231, 232)). Others, such as *CHL1* and *TCL1A* have no known links to B cell function and merit further functional studies. Specifically, *CHL1* may be involved in cell adhesion and migration, as these

functions are ascribed to its homolog *L1CAM* (233) and *TCL1A* is involved in AKT activation, which is important for BCR signaling (234).

There were five MRGs from **module 38** that were also not TEGs: *BLNK*, *DENND5B*, *DIRAS1*, *RASGRF1*, and *LOC102724714*. *BLNK* encodes a critical adaptor protein that bridges BCR-associated kinase activity with downstream signaling events (235, 236). *DENND5B* is a poorly characterized gene containing a MAP kinase activating death (MADD/DENN) domain. While the function of *DIRAS1* and *RASGRF1* are not well understood, it is likely that these RAS family proteins are involved in B cell signaling.

2.6.5 A module shared by B cells and monocytes; a gene expression program enabling antigen processing and presentation

While **modules 22 and 38** are specific to B cells, **module 41** was strongly associated with both B cells and monocytes, two antigen-presenting cells (**Table 2.1**). Global analysis of the 39 genes within this module revealed multiple annotation terms (N=66) that were significantly enriched, with the three most significant (%FDR<10⁻⁸) being “endosome”, “MHC classes I/II-like antigen recognition protein”, and “MHC class II protein complex” (**S2.4 Table**). In fact, many of the annotation terms found to be enriched in this module can be explained by a core set of genes including *CIITA*, *CD74*, *HLA-DMA*, *HLA-DRA*, *HLA-DRB1*, *HLA-DMB*, *CD1A*, *CD1C*, *LY86*, and *CD180*. The five TEGs in **module 41** are part of this core set of genes (*CIITA*, *CD74*, *HLA-DMA*, *HLA-DRA*, and *HLA-DRB1*), all involved in and essential to antigen presentation in B cells and monocytes (237, 238) (**S2.8 Fig**).

Regarding the genes that are most representative of **module 41**, there were six MRGs (*CD74*, *SIDT2*, *FGD2*, *CD1A*, *MMP17*, and *HLA-DRA*), three of which are included in the core set of genes described above (*CD74*, *CD1A*, and *HLA-DRA*). Additionally, *SIDT2* and *FGD2* contribute to intracellular trafficking relevant to antigen processing (239, 240). *MMP17* is an endopeptidase that may be involved in the activation of membrane-bound precursors of growth factors or inflammatory mediators, such as TNF- α (241), which enhances antigen presentation. The combination of the global and targeted analyses clearly supports **module 41** having a role in antigen presentation in both B cells and monocytes.

2.6.6 A shared module associated with NK, $\gamma\delta$ T, and CD8+ T cells; a role in cytotoxic granule and non-self-recognition

Fig 3 shows that **module 26** is associated with NK, $\gamma\delta$ T and CD8+ T cells, strongly suggesting that it contains genes encoding for functions shared by these cell types (**Table 2.1**). The global analyses of the 129 genes within this module identified multiple significantly enriched annotation terms, including “natural killer cell-mediated cytotoxicity”, “cellular defense response” and “regulation of immune response”, all with FDR<10⁻⁷ (**S4 Table**). Genes enriched for these functions are located in the KIR locus (*KIR2DL1*, *KIR2DL3*, *KIR3DL1*, *KIR3DL2*, and *KIR2DS4*) within the human leukocyte receptor gene complex (LRC), in the NK gene complex (NKC) (*KLRC1*, *KLRC2*, *KLRC3*, *KLRC4*, *KLRD1*, *KLRF1*, and *KLRK1*), encoding proteins found in cytolytic granules (*PRF1*, *GZMB*, and *GZML*) and cell surface proteins (*CD160*, *NCR1*, and *FASLG*) that mediate and/or potentiate cell-cell interaction. The primary functions that NK, $\gamma\delta$ T, and

CD8+ T cells have in common are the recognition and killing of abnormal cells (e.g. infected with viruses or other intracellular pathogens, cancerous), and the genes noted above are essential to these functions (**Fig 2.5**). It should be noted that all of the genes have equivalent expression levels across these three cell types, except for the KIR genes which had a differential expression pattern (NK > $\gamma\delta$ T > CD8+ T).

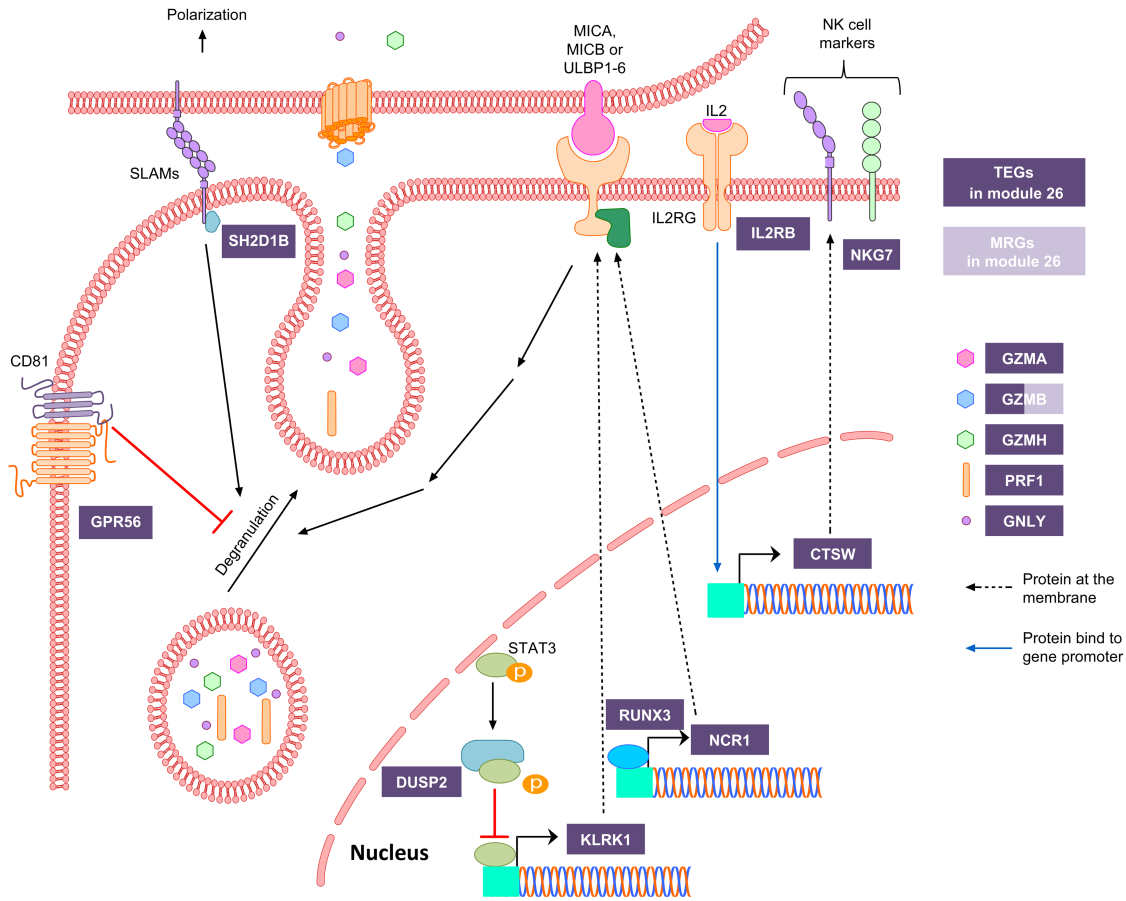


Figure 2.5. Targeted analyses of genes within module 26, associated with NK, $\gamma\delta$ T, and CD8+ T cells. The majority of genes in *module 26* encode components of cytotoxic granules and proteins involved in the recognition of non-self. Genes in *module 26* are represented in purple: TEGs in dark purple and MRG in light purple. To see the profile of gene expression mean of all genes of module 26 presented in this Figure refers to the heatmap in **S7 Fig**. <https://doi.org/10.1371/journal.pone.0233543.g005>

When examining the 18 TEGs in these cell types, many of the same genes are highlighted (*GNLY*, *GZMB*, *KLRD1*, *KLRF1*, *KLRK1*, and *PRF1*), although other TEGs, such as *GZMA* and *GZMH*, can also be attributed to cytolytic granules. *NKG7* (alias *GMP-17*), another TEG, may also be contained within these granules in all three cell types, since it is found in the membrane of NK cell granules (242). Many of the other TEGs within this module are involved in regulating the cell activation state, such as *GPR56*, an inhibitory receptor on NK cells (243), *NCR1*, that displays

either a stimulatory or inhibitory effect depending on the ligand (244), and SH2D1B that regulates signal transduction from cell-surface SLAM receptors and leads to granule polarization to cell-cell synapses (245). In addition, CD160 is potentially an activating receptor encoded outside the LRC and NKC complexes (246, 247).

Also included in this list of TEGs are a number of genes whose functions are poorly characterized in NK cells, $\gamma\delta$ and CD8+ T cells, including *CST7*, *CTSW*, and *FGFBP2*. *CST7* encodes cystatin F, which may be important for the processing and activation of granule-associated serine proteases, in particular granzymes A and B (248). *CTSW* encodes Cathepsin W, which is a putative cysteine protease that is believed to be released during target cell killing, although apparently not essential to the process of cytotoxicity (249). *FGFBP2* is highly expressed in NK cells, $\gamma\delta$ T and CD8+ T cells and essentially absent in the other immune cell types studied herein, suggesting that it plays a specific function in these cells.

Regarding the genes that are representative of **module 26**, there were 21 MRGs, with three (*GZMB*, *KLRD1*, and *S1PR5*) also being in the list of TEGs (**S2.6 Table**). Five of the MRGs are encoded within the LRC or NKC complexes (*KLRD1*, *KIR3DL1*, *KIR2DL3*, *KLRC3*, and *KLRC1*). This MRG list also contained *FASLG*, important in T lymphocyte induced cell death, as well as *B3GAT1* which encodes a key enzyme in the biosynthesis of the carbohydrate epitope HNK-1 (human natural killer-1, alias CD57). Moreover, this list contains the genes that encode the XCL1 and XCL2 cytokines – two ligands for the chemokine receptor XCR1. Although still poorly characterized, this chemokine likely mediates the chemotactic activity of cDC towards cytotoxic cells (250).

In terms of transcriptional control of genes within this module, it is interesting to note that *DUSP2*, a TEG of this module, encodes a dual-specificity phosphatase that negatively regulates the activity of the signal transducer and transcriptional activator *STAT3*, which in turn regulates the transcription of the *KLRK1* gene within the NK gene complex (251, 252). The TF *RUNX3* is also a TEG and regulates genes implicated in lymphocyte activation, proliferation, cytotoxicity, migration and cytokine production in CD8+ T and NK cells, like *NCR1* and *IL2RB* (253, 254). Although not within the list of TEGs, *ZNF683* (alias *Hobit*) drives the expression of *GPR56* in NK cells (243).

Taken together, these analyses suggest that **module 26** has two predominant functions, the first being cytotoxic granule composition, and the other being non-self-recognition, very consistent with the role of NK, $\gamma\delta$ T, and CD8+ T cells.

2.6.7 Modules associated with LPS-activated macrophages

In addition to identifying co-regulated genes associated with specific cell types, we also explored our ability to use this experimental approach to examine the transcriptional impact of modifying the activation state of a cell. Specifically, we cultured CD14+ enriched mononuclear cells from human peripheral blood with M-CSF for a total of 8 days to obtain macrophages (255, 256) and activated macrophages were generated by stimulating the macrophages with LPS during the last 24h of culture (257). As can be seen in **Fig 2.3**, **module 29** stands out as being highly associated with response to LPS cells (**Table 2.1**). A global analysis of the 93 genes within

this module identified 29 significantly enriched annotation terms (FDR<5%) that fell within two groups. The first group, including “cellular response to interferon-gamma”, “cellular response to interleukin-1” and “cellular response to tumor necrosis factor”, is driven by a common set of genes (*CCL1*, *ASS1*, *CCL20*, *EDN1*, *CCL8*, *CCL19*, *IRG1* (alias *ACOD1*), *CCL15* and *CCL17*), many of which are chemokines (*CCL* or *CXCL* genes) acting as chemoattractants for immune cells. *EDN1* is expressed by endothelial cells, although it can also be produced by macrophages co-cultured with activated T cells (258). The second group includes annotations related to “metallothionein” and is also driven by a common set of genes (*MT1G*, *MT1H*, *MT1L*, *MT1M*, and *MT2A*) known as metallothioneins; they shape the macrophage zinc pool in response to inflammatory and infectious stimuli (259).

Module 29, strongly associated with LPS-activated macrophages, contains five TEGs; *CCL8* and *MT2A*, both highlighted by the global analysis, complement factor B, known to be LPS-induced (260), *IDO1*, which catalyzes the first and rate-limiting step of the catabolism of the essential amino acid tryptophan along the kynurenine pathway (261), and, finally, *ST3GAL1*, which is highly expressed in all cell types examined, and thus is likely, not informative to this module. In addition to the TEGs, 14 MRGs were found in **module 29** including three metallothioneins (*MT1G*, *MT1H*, and *MT1M*) and one chemokine (*CCL15*) also highlighted by the global analysis. Among the other MRGs, we found two cytokines, *CSF3* and *IL36G*, as well as *PDGFRL*, whose molecular function is not established but may be involved in cell proliferation (262). Taken together, these analyses support an antimicrobial function for this module, with an important signature of chemokine production and zinc metabolism.

2.6.8 The neutrophil transcriptome is distinctive from other immune cell types

Examining the distribution of reads, before and after normalization, we observed that neutrophils have a very different pattern when compared to the other cell types (**S2.2 Fig**). Neutrophils have fewer expressed genes (11,626 genes for neutrophil and between 13,161 and 13,614 for other cell populations) and the spread in expression is larger than the other cell types. Not surprisingly, this was also observed in the PCA and hierarchical clustering analyses, representing global differences in the transcriptome of neutrophils as compared to all other cell types examined herein (**Fig 2.1A**). In addition, there were 15 transcripts with a greater than the 100-fold difference between the average number of read counts in neutrophils as compared to the average number of read counts across all other cell types combined (**see Supplementary Materials**).

Global analysis of the 13 neutrophil modules identified general functions such as “transcription” and “phosphoprotein”; however, a number of more specific functions relevant to neutrophils/myeloid cells were also identified, such as “lysosome”, “extracellular exosome”, “movement of cell or subcellular component”, and “superoxide-generating NADPH oxidase activator activity” (**S2.4 Table**). In terms of targeted analyses for these 13 neutrophil-associated modules, there were multiple TEGs (n=349) and MRGs (n=628). These analyses highlighted neutrophil polarity, lipid metabolism, and chemoattraction as functions related to **module 4**; calprotectin, azurophil granules, cytoskeleton dynamics, chemoattraction for **module 7**; actin remodeling & membrane trafficking, nicotinamide metabolism for **module 10**; neutrophil

polarization, trafficking and exocytosis for **module 21**; and endocytosis and membrane trafficking for **module 36**.

Moreover, given that the molecular aspects of many neutrophil functions have previously been characterized, we looked for the presence of the transcripts relevant to these functions within our modules. Specifically, we looked for 126 transcripts relevant to neutrophil granules (e.g. azurophil, gelatinase, secretory), antimicrobial proteins, reactive oxygen species (ROS), Neutrophil Extracellular Traps (NETs), cell cross-talk and resolution of inflammation (apoptosis and clearance, lipid mediator class switch) (114, 263-265). One hundred 11 transcripts were expressed in our dataset, with 80 of them found within our neutrophil-associated modules. Using this approach, we found that **module 11** was associated with the presence of transcripts for azurophil granules, **module 15** with defensin-specific granules, **module 21** with ROS and NET proteins, and **modules 16** with NET-associated Histone Cluster 1 family members. Extending this approach to the neutrophil modules also associated with other myeloid cells, we observe multiple transcripts relevant to neutrophil functions, either because the functions are shared between these cell types (e.g. ROS production, vesicle trafficking) or because the gene products are involved in different functions (e.g. actins are involved in NETs and other cytoskeleton-related functions).

2.6.9 Transcriptional control of genes within B cell modules

In order to gain an understanding of the transcriptional control of genes within each module, we were interested in examining whether there was an enrichment of TF binding sites (TFBS) within the promoters of the genes within its module. In order to do so, we used the empirical ChIP-Seq data from the ENCODE (Encyclopedia of DNA Elements) Project (162). Given that B lymphocytes are well represented in the ENCODE dataset - 76 TFs were studied in the GM12878 immortalized B cell line – we focused our initial analyses on the three modules most associated with B lymphocytes and described above (**modules 22, 38 and 41**). Specifically, these analyses identified that the TFBS for EZH2 was significantly enriched in the proximal promoters of the 195 genes within **module 22** and the TFBS for IKZF1 was enriched in the promoters of the 45 genes within **module 38** (**S2.7 Table**). Analysis of the promoters for genes within **module 41** associated with B cells and monocytes revealed enrichment for nine different TFBSs (BCL11A, CEBPB, EP300, EZH2, IKZF1, MTA3, NFATC1, RFX5, and SPI1) (**S2.7 Table**). Interestingly, the enrichment was the greatest for IKZF1, with over four- and five-fold enrichment in **modules 38, and 41**, respectively. The transcript for IKZF1 is found in **module 17**, which while associated with lymphoid cells (**S2.6 Fig**), has also strong expression across all immune cells (**S2.7 Table**). BCL11A is essential for lymphopoiesis and B cell development; it is also a critical component of a transcriptional network that regulates B cell fate by controlling V(D)J recombination and functions upstream of EBF1 and PAX5 in the B cell lineage differentiation pathway (266-268). RFX5, EZH2, EP300, and SPI1 all control MHC class II expression either directly or via interaction or regulation of CIITA (269-276). MTA3 has been implicated in cell fate during B lymphocyte differentiation (277). NFATC1 regulates normal homeostasis and differentiation in B cells (278). Finally, CEBPB is involved in the survival of Ly6C⁻ monocytes and have unknown function in MHC class II (279).

Interestingly, all of the TFs enriched in this analysis have average-to-high levels of expression in B cells. Among the 76 TFs that were evaluated, only two (MYC and NFATC1) have their TFBS enriched in their own modules (**modules 3 and 23**; respectively). Specifically, the empirical data provided evidence of a TFBS for MYC in the promoter region of 180 genes of the 619 genes from **module 3**. It has been shown that MYC is necessary to stimulate both proliferation and inhibited differentiation in mature B cells induced by BCR signal (280, 281). BCR signaling is also known to increase the transcription of the *NFACT1* gene, which in turn plays an important role in controlling plasmoblast/plasma cell formation (282, 283).

2.6.10 Transcription factors potentially associated with cell lineages and states.

As described above for the analysis of the B cell-specific modules, the expression of TFs playing a key role in cell fate can be maintained in mature cells. In fact, with the exception of *BACH2*, *TCF3* and *IKZF1*, all of the key TFs known to be involved in the multiple steps of B-cell-lineage differentiation from common lymphoid progenitor to mature B cell are within modules 22 and 38, despite their being identified from the transcriptome of mature B cells (**S2.8 Table**) (284-287). *BACH2*, *TCF3*, and *IKZF1* were not identified in these analyses as, while they are necessary to B cell differentiation, their expression is not unique to B cells and in fact, they are ubiquitous in the primary immune cell types studied herein. We were therefore interested in identifying additional TFs that were highly expressed in a given cell type or cell lineage. In order to do so, we examined the 1600 known human TFs (26) for expression patterns that were consistent with the differentiation scheme depicted in **Fig 2.2B**. Specifically, we considered each branch point as representing a cell fate decision and looked for TFs that appeared to be specific for a given fate (e.g. myeloid vs. lymphoid; neutrophil vs. monocyte; B cell vs. T cell). For example, we looked for TFs that were highly expressed in lymphoid cells and absent or only very weakly in myeloid cells and identified 13 TFs that are candidates for being involved in lymphoid differentiation and/or function (**S2.5 Fig and S2.5 Table**). This list includes *IKZF3*, which is a known regulator of lymphoid development and *BACH2*, a known inhibitor of myeloid differentiation (288, 289). In addition, when looking for TFs potentially involved in the differentiation step leading to monocytes, we identified *SMAD1*, which is known to be implicated in monocyte differentiation, polarization and inflammation (290, 291) and *MITF* which has not previously been associated with monocyte differentiation but is known to be a phagocyte-restricted TF in macrophage (292). Altogether, we identified 74 TFs differentially expressed in one cell type or lineage (**S2.5 Fig and S2.5 Table**), with over a third (22/74) having confirmatory evidence in the literature as being involved in the relevant differentiation step. While not definitive proof, this certainly indicates that these TFs are strong candidates for playing a role in cell fate decisions and/or maintenance.

2.7 Discussion

Cells within the immune system are generated through successive differentiation steps from a common pluripotent hematopoietic stem cell progenitor. As in all differentiation processes, certain functions are gained and others are lost along the way, such that each immune cell subset has different functions, some unique to a given cell type while others are shared. For decades, such differences have been exploited in flow cytometry-based experimentation, with

specific cell surface markers (e.g. CD3, CD4, CD8, and CD14) used to identify, qualify and quantify immune cell populations within different experimental contexts. Sequencing the transcriptome (RNA-seq) of these immune cells can provide a complementary approach to understanding the differences between immune cell types or between different experimental conditions, disease states. RNA-seq based approaches certainly come with the challenge of analyzing and interpreting the data on thousands of transcripts per sample. We, therefore, set out to generate, analyze and interpret RNA-seq data of primary human immune cells from healthy individuals in order to evaluate the feasibility and usefulness to understand some of the biology underlying these cells and could play a major role in many immune-mediated diseases.

Starting with the premise that genes with correlated expression patterns are more likely to have shared biological functions than if they have disparate expression patterns, we aimed to identify functional modules within and across primary human immune cells by defining sets of co-expressed genes. Indeed, this unsupervised approach identified co-expression modules that were either highly correlated to a single cell type or to a group of cell types (**Fig 2.3**). We then assigned potential functions to these modules using a combination of global and targeted annotation approaches. From these analyses, it was clear that modules associated with a single cell type were linked to known functions of those cell types; for example, BCR signaling pathway genes in the B cell-specific module or genes encoding proteins linked to functions of specific granules, ROS and NET production in different neutrophil-associated modules. As importantly, modules shared across different cell types also had biologic meaning; for example, MHC Class II antigen processing and presentation in B cells and monocytes, or cytotoxic granule composition and non-self-cell identification in NK, $\gamma\delta$ T, and CD8+ T cells. These results confirm that this approach can reveal co-expressed genes that do share functions.

While not the primary objective of the current study, these co-expression modules also provide an opportunity to use a “guilt by association” approach in order to identify additional genes involved in the specific functions revealed by our analyses. For example, a number of genes in **module 41** encode proteins implicated in vesicular traffic (*BLOC1S6*, *LAMP5*, *TBC1D5*, and *TRAK1*), potentially revealing novel players in the trafficking of endosomal vesicles involved in antigen processing and presentation. Additional examples are: (1) two RAS family proteins, *DIRAS1* and *RASGRF1*, that are potentially involved in B cell signaling (234); (2) *FGFBP2* potentially having a role in cytotoxic functions of NK, $\gamma\delta$ T and CD8+ T cells (250); (3) *NECAB2* potentially being an important regulator of neutrophil apoptosis, autophagy and NETs; and (4) the enzyme-couple channel *TRPM6* involved in neutrophil chemotaxis. While experimental validation is required to confirm these proposed functions, nonetheless, this illustrates how additional functional hypotheses can be generated from the information contained within the co-expression modules.

Also found within these modules are genes encoding TFs and transcriptional regulators; multiple TFs with known roles within the immune system were found within the relevant modules. Prime examples are the B cell-specific TFs *PAX5* and *EBF1* that are respectively within the B cell-specific **modules 22 and 38**, *CIITA* is known to regulate MHC class II genes found within **module 41**, and *DUSP2*, in **module 26**, is a dual-specificity phosphatase that influences indirectly the

expression of killer receptors within the NK gene complex. In order to assess whether TFs controlling the expression of genes within their own co-expression modules or in distinct modules, we focused our analyses on B cell modules, given the extensive ENCODE Chip-seq data for the 76 TFs in cell lines from the B cell lineage. While we detected an enrichment signal for 45 TFBS in 12 B cell-associated modules. Only two TFs (MYC and NFATC1) have their TFBS enriched in their own modules (280-283), supporting a network model of TFs regulating the transcription of genes in distinct co-expressed gene sets, which is consistent with current literature (153, 188, 189, 293).

While the focus of our analyses was to identify functional subgroups of genes, we did perform an analysis of TFs that were independent of the co-expression modules described above. Specifically, we searched for TFs that differentially expressed in a given cell type or lineage. Indeed, this approach identified many known lineage-specific TFs that are likely involved in maintaining lineage commitment and key functions associated with those commitments. Prime examples were IKZF3 known to be involved in lymphocyte development and homeostasis (288), PAX5 in differentiation and function of mature B cells (11), and FOXP3 in CD4+ T cell differentiation (294-296). This analysis also highlighted a few TFs previously unknown to be associated with specific lineages, such as ZNF860 in B cell differentiation, ZNF385A in monocyte differentiation, CREB5 in the myeloid lineage, TFCEP2L1 in $\gamma\delta$ T lineage, and MSC in the activation of macrophages in response to LPS (**S2.5 Fig and S2.5 Table**). While the expression patterns were certainly consistent with a role in cell fate decision and/or maintenance, with many having published confirmatory evidence (18 of 49), future studies will be required to confirm the role of the remaining candidates.

In the future, the approach described herein could be used for the analysis of circulating immune cells in a variety of contexts, such as in cross-sectional comparisons between different immune-mediated diseases, in longitudinal studies of disease progression or response to therapy, or in a comparison of circulating immune cells versus immune cells taken from peripheral inflammatory sites. It is clear that comparisons in such study designs can be complex, with different expression patterns being a result of changes in cell populations, in addition to gene expression within cells, and therefore it will be important to rely on careful immunophenotyping and cell sorting strategies. Alternatively or in complementarity, with an inevitable decrease in the costs associated with single-cell RNA sequencing (scRNA-seq), this may become a better option, with the analysis strategy presented herein also being relevant to this experimental approach.

2.8 Funding

This work was supported by grants from the Canadian Institutes of Health Research: (1) The “CIHR Emerging Team in Integrative Biology of Inflammatory Diseases” [CIHR #102170] with co-funding from Crohn’s and Colitis Canada made possible via a generous donation from the Ross McMaster Memorial Fund; (2) the IMAGINE-SPOR Chronic Disease Network [CIHR #SCA-145105]. Salary support to Saraï Mola was provided by the Canada Research Chair in Genetics and Genomic Medicine awarded to John David Rioux [CRC#230625]. Sylvie Lesage holds a

Research Scholar Emeritus award from the Fonds de recherche du Québec – Santé (FRQ-S). This work was supported in part by the infrastructure at Compute Quebec/Canada.

2.9 Acknowledgments

Authors would like to thank Martine Dupuis (Centre de recherche Hôpital Maisonneuve-Rosemont) for providing cell sorting service, and the production team at the McGill/Genome-Quebec Innovation Center for the high throughput sequencing service.

2.10 Competing interests

The authors declare that no competing interests exist.

2.11 Protocol

The protocol for the experiment described herein has been deposited in protocols.io (dx.doi.org/10.17504/protocols.io.bcagisbw).

2.12 Supporting information

S1 File: Additional module analyses. Figures S2.1 to S2.8 and tables S2.1 to S2.8.

3 Chapitre 3 - Identifying transcript-level differential expression in primary human immune cells

3.1 Contribution des auteurs

Auteurs : Saraï Mola, Claudine Beauchamp, Gabrielle Boucher, Philippe Goyette, Mohamad Karaky, Sylvain Foisy, John D. Rioux.

Statut de l'article:

Soumission à Molecular Immunology

Conception et mise au point des expériences :

Saraï Mola, Gabrielle Boucher et John David Rioux.

Expériences effectuées, analyses des résultats et présentation des figures:

J'ai mesuré l'abondance de l'expression des transcrits à partir des données nettoyées ainsi que l'application de l'approche d'étude développée par mes soins et l'analyse des résultats, soit 95% du projet

Les 5% restant sont partagés entre les autres auteurs de l'article

Rédaction de l'article : J'ai rédigé 50% de l'article en collaboration avec mon directeur, John David Rioux, qui a rédigé les 50% restant.

3.2 Article

Identifying transcript-level differential expression in primary human immune cells.

Saraï Mola^{a,b}, Claudine Beauchamp^a, Gabrielle Boucher^a, Philippe Goyette^a, Mohamad Karaky^a, Sylvain Foisy^a, John D. Rioux^{a,b,c}.

^a Centre de recherche, Institut de cardiologie de Montréal, 5000 Rue Bélanger, Montréal, QC H1T 1C8, Canada

^b Département de biochimie et médecine moléculaire, Université de Montréal, Pavillon Roger-Gaudry, C.P. 6128, Succ. Centre-ville, Montréal, Québec, H3C 3J7, Canada

^c Département de médecine, Université de Montréal, Pavillon Roger-Gaudry, C.P. 6128, Succ. Centre-ville, Montréal, Québec, H3C 3J7, Canada

Data availability statement: Researchers can request access to the individual-level data from the current study by contacting the Montreal Heart Institute ethics committee at the following institutional email address: cer.icm@icm-mhi.org.

3.3 Abstract

Background: Multipotential hematopoietic stem cells differentiate into a wide variety of immune cells with a diversity of functions, including the ability to respond to a variety of stimuli. Importantly, numerous studies have demonstrated the importance of gene transcription in defining cell identity and functions. While these studies have primarily been performed at the level of the gene, it is known that key immune genes such as *CD44* and *CD45* generate multiple different transcripts that are differentially expressed across different immune cells, and that encode proteins with different sequences and functions. Prior genomic surveys have shown that the mechanisms for generating diversity in expressed transcripts (alternate splicing, alternate transcription start sites, etc.) are very active in immune cells, but have been lacking in terms of identifying genes with multiple transcripts, that are differentially expressed, and likely to affect cell functions.

Methods: We first identified the set of genes that had at least two transcripts expressed in our RNA sequencing dataset generated from purified populations of neutrophils, monocytes and five lymphocyte populations (B, NK, $\gamma\delta$ T, CD4+ T and CD8+ T) from twelve healthy donors. Next, we developed a heuristic approach to identify genes where two or more transcripts have distinct expression patterns across lymphoid and/or myeloid populations. We then focused our annotation and interpretation on differentially expressed transcripts that affect the coding sequence. This process was repeated to identify transcripts that were differentially expressed between monocytes and populations of macrophages and LPS-stimulated macrophages derived from these monocytes *in vitro*.

Results: We found that over 55% of genes had two or more expressed transcripts, with an average ~3 transcripts per gene, and that 70% of these had at least two of the transcripts resulted in proteins with different sequences. As expected, we identified a complex pattern of differential expression for multiple transcripts encoding the *CD45* transmembrane protein, but we also found similar evidence for ten other genes (*CD300A*, *FYB1*, *GPI*, *LITAF*, *PSMA1*, *PTMA*, *RPL32*, *SEPTIN9*, *SH3BP2*, *SH3KBP1*) when comparing the expression patterns of transcripts within myeloid and lymphoid cells. We also identified five genes with differentially expressed transcripts associated with the transition from monocytes to macrophages (*FNBP1*, *KLF6*, and *SEPTIN9*) or between macrophages and LPS-stimulated macrophages (*CD44*, *OAZ2*, and *SEPTIN9*). For the most part, we found that the different transcripts of these genes are expected to impact specific biological functions, for example the different transcripts of *SEPTIN9* likely regulate the cytoskeleton in immune cells via their interactions with actins filaments and microtubules.

Conclusions: This analytic approach successfully identified multi-transcript genes that are differentially expressed across immune cells and could be applied to other transcriptomic data.

Keyword: primary human immune cells; RNA-Seq; differential transcript expression control; protein isoforms

Abbreviations: TF: transcription factor; CPM: Counts per million; TSS: Transcription start site; AS: Alternative Splicing; NMD: nonsense-mediated RNA decay

3.4 Introduction

While it is estimated that the human genome contains 20,000 to 25,000 protein-coding genes, transcriptional, post-transcriptional, and translational mechanisms are believed to generate a total of 250,000 to one million different proteins. Diversity at the transcript level can be generated via alternative transcription start sites (TSS), alternative splicing (AS), and alternative polyadenylation sites (297-299). RNA sequencing (RNA-Seq) technologies enable a precise measurement of the expression quantity not only at the gene level but also at the level of the different transcripts for each gene, generated by these various mechanisms. Specifically, transcriptome measurement with RNA-Seq can provide useful information into each transcript such as the TSS and termination site (3' polyadenylation site) as well as skipped exons and retained introns (299). Diversity in transcripts can lead to multiple protein isoforms that can differ in subcellular localization, protein-protein interaction, stimulatory or inhibitory activities, stability, DNA-binding properties, lipid-binding properties, or enzymatic activity (43, 45). Alternative transcripts can, however, also lead to proteins isoforms with disrupted structures and functions. Importantly, aberrant transcripts are targeted for nonsense-mediated RNA decay (NMD), although it has been observed that the selectivity of targeting of the NMD system for different transcripts is a means of regulating protein expression (300).

At the gene level, analysis of RNA-Seq data generated from human tissues by the Genotype-Tissue Expression (GTEx) project found that variation in gene expression is far greater among tissues (47% of total variance in gene expression) than among individuals (4% of total variance) (299). It was also found that thousands of genes are differentially expressed between tissues or show tissue-preferential expression (3), together supporting the notion that gene expression patterns can provide insight into tissue-relevant biological functions. While most of this work was carried out in human tissues or cell lines, it has further been shown via gene level transcriptomic analyses across several human immune cells that networks of co-expressed genes are associated with cell identity and biological functions (1, 2, 54).

At the level of transcripts, analyses of 29 solid organ tissues, 11 brain subregions, whole blood, Epstein-Barr virus-transformed lymphocytes and cultured fibroblasts from skin from the GTEx Project found that alternative TSS and termination sites accounted for the majority of alternate transcripts, and are most likely to lead to functional proteins as compared to alternate splicing events that involves untranslated exons and thus not as likely to lead to novel protein sequences (299). This study also suggested that the main driver of cellular identity was transcription at the gene level and that differential expression at the level of individual transcripts played a complementary role (3).

Given that these studies were performed with complex tissues, to better understand the role of differential expression of different transcripts in the immune system it is important that this question be studied within different isolated immune cell populations. Importantly, the Immunological Genome Project Consortium (ImmGen) looked at transcript diversity by examining RNA-Seq data for alternate splicing events in murine B and CD4+ T cells and differentially expressed exons using micro-puce data across 172 ImmGen cell populations (138). They reported AS events in ~60% of the genes, with patterns of AS linked to lineage differentiation and cell cycle, while noting a few examples of differential splicing between murine B and CD4+ T cells (e.g. in the *Ptprc* gene) (138). Alternate splicing events have also been examined in primary human monocytes and macrophages differentiated and polarized *in vitro* from four healthy donors (144). This study found a prominent role for AS associated with the differentiation and activation events, with RBFOX2 and MBNL1 RNA binding proteins likely being regulators of AS. Given these findings, we were prompted to study nine different immune cell populations from twelve healthy donors to better understand the role of alternate TSS, AS and alternate polyadenylation sites in generating transcripts that are differentially expressed between cell populations and activation states.

3.5 Materials and methods

3.5.1 Description of dataset

All the raw data used for the current study was generated as part of our previous study (54). Briefly, in this previous study, 12 healthy volunteers were recruited and assigned a randomly generated ID by the Montreal Heart Institute (MHI) Biobank Center to ensure anonymity as previously reported. Peripheral blood was collected and then processed to isolate purified populations of neutrophils, monocytes and five lymphocyte populations (B, NK, $\gamma\delta$ T, CD4+ T and

CD8+ T). The CD14+ enriched mononuclear cells were then differentiated and stimulated to generate populations of macrophages as well as activated macrophages, as previously described (54). RNA was extracted, sequencing libraries prepared, and paired-end 2x100bp reads were generated, as previously described to (54).

This project was approved by the MHI Institutional Ethics Committee « Comité d'éthique de la recherche et du développement des nouvelles technologies ». Researchers can request access to the individual-level data from the current study by contacting the Montreal Heart Institute ethics committee at the following institutional email address: cer.icm@icm-mhi.org.

3.5.2 Bioinformatic processing of sequence files

In the previous study, primary QC analysis of raw RNA-Seq data was performed with FastQC (v0.10.1) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and sequences were trimmed of low-quality bases with Trimmomatic (v0.33) to remove potential adapter contamination (190). For more information, referred to (54). For the current analyses, HISAT (hierarchical indexing for spliced alignment of transcripts) (v 2-2.1.0) (128, 301) was applied to align sequencing reads to the reference genome. The human gene annotation file from GENCODE for build GRCh38 was used. Sorting of the aligned reads was done with SAMTools (v1.9) (191) and was also used to convert SAM files to BAM.

Transcript abundance was quantified with StringTie (v2.0.4) (127, 128). By default, StringTie returns transcript expression matrix in RPKM or FPKM. To create transcript expression matrix of read counts, the python script prepDE.py was used on files generated by StringTie (128). Prior to the analysis of transcript expression level, the data was filtered and normalized with R/Bioconductor packages (302). The expression count matrix and transcript annotations were assembled in a DGEList object with *DGEList* function from edgeR package (v3.12.1) (193, 194). To filter out non- and low-expressed transcripts, a CPM threshold of 3 CPM into at least five samples was used and defined by analysis of gene expression data through our nine immune cell types (**S2.1A Fig.**). Sample normalization factors were calculated with *calcNormFactors* function from edgeR. Normalization and variance stabilization of data were achieved with *voom* function (303) from limma package (v3.42.2) and TMM factors.

3.5.3 Principal component and hierarchical clustering analyses

Transcriptomic data of gene transcripts was analyzed with Principal Component Analysis (PCA) and hierarchical clustering approaches to provide a global view of the data. With these approaches, we identified expression profile similarities between individuals and cell types. We used the PCA approach to assess the principal axes of variation in our data. Principal components were computed from filtered and normalized transcript expression data with R function *prcomp*. A graphical display was used to represent the two first PCs the largest axes of variation (**S2.2 Fig.**). As a complementary analysis, we computed an unsupervised hierarchical clustering to identify sample grouping. As a measure of similarity between samples, Pearson correlation was computed on filtered and normalized transcript expression data with R function *cor*. Distance (dissimilarity) was then computed as 1-similarity. The data dendrogram was built based on hierarchical clustering with *hclust* R function (**Fig. S3A**). Finally, the fraction of

transcript expression variance explained by cell types was computed from ANOVA. Expression for each gene was centered and scaled to have mean and variance values respectively of 0 and 1. The sum of squares (SS) explained by cell types was obtained from analysis of each gene variance and SS sum of all genes dividing by the total SS used to estimate multivariate variance explained.

3.5.4 Transcript annotations and gene information: protein isoform encoded, transcription factor and gene involved in splicing machinery

To align reads, we used HISAT with human gene annotation file GENCODE version 32 (gencode.v32.annotation.gtf). This file provided some transcript information like Ensembl transcript ID, transcript name, start, end, strand, chromosome, gene name, gene ID, gene type and transcript type. All transcripts with protein-coding keyword in the transcript type category were defined as a transcript encoding a protein. From two UniProt file databases, we added protein UniProtKB AC and protein length to all transcripts encoding a protein isoform (HUMAN_9606_idmapping.dat and GRCh38_latest_protein.gpff).

To complete annotation of transcripts, we identified those encoding for a transcription factor (TF). For this purpose, we used the catalog of human TFs published by Lambert and colleagues (26). To identify human TFs and their binding site motifs, this catalog was based on specific databases including TRANSFAC, HT-SELEX, UniProbe, and CisBP (209-218). Their definition of a TF is a protein capable of 1) binding DNA in a sequence-specific manner and 2) regulating transcription. They identified 1,639 human TFs, and two-thirds of these have binding motifs (see <http://humantfs.ccb.utoronto.ca/index.php>). We also identified genes that encode proteins involved in RNA splicing machinery. We obtained two lists: 1) a list built from spliceosome analyses with an algorithmic model published by Akerman M. et al. (304) and 2) a list computed in ProteinAtlas.org with the keyword “mRNA Splicing”. These two lists were used to finalize transcript annotation.

3.5.5 RNA isoforms associated with cell lineages and states.

We hypothesize that some gene transcripts are involved in fate decision in primary immune cell types, which is not always captured when global gene expression is studied (54). To investigate on this hypothesis, we developed an algorithm to identify candidate genes based on: (1) presence of a query transcript which has high expression specific to one or a subgroup of cell types, i.e. below detection threshold in the other cell types, and (2) presence of alternate transcripts that have an opposite or different expression profile compared to the query transcript. An alternate transcript is defined as one showing high expression levels in at least one cell type for which the query transcript has low-undetected expression. We based our method on the geometric mean of CPM expression from each cell types.

This method was applied to all the dichotomous branchpoints along the differentiation tree for immune cells (e.g. B vs T cells) as well as the steps along the differentiation and activation of monocytes to activated macrophages. For each gene with at least two transcripts, the algorithm identifies transcripts with a “candidate profile”, defined as an expression level superior or equal to a high-expression threshold (we used 60 CPM) in at least one cell type of the first subgroup

(e.g. B cells) and an expression level inferior to the low-expression threshold (5 CPM) in all cell types of the second subgroup of the comparison (e.g. T cells). If one or more of the other transcripts for this same gene had an expression superior or equal to the high-expression threshold (60 CPM) in the second subgroup of cell types, the transcript is called alternate transcript and the transcript with a candidate profile is called query transcript. Any genes with transcripts fitting this pattern are considered candidates for additional follow up. For RNA-Seq data, all transcripts not reaching 3 CPM in at least five samples were considered not expressed and excluded from further analyses. In our analyses for comparing transcript profiles, we used a slightly higher threshold of 5 CPM on the geometric mean computed per cell type to define low expressed transcript in that cell type. This threshold was chosen to be stricter than the detection threshold, also accounting for the distribution of the averaged expression value per cell type (**S2.4 Fig.**). We observe that 15566 (57%) transcripts reach an average of 5 CPM in at least one cell type, compared to 18965 (68%) at 3 CPM. To define moderate to high expression level for transcripts, we used an arbitrary threshold of 60 CPM. This threshold corresponds to a 20-fold-change from our detection threshold of 3 CPM and 12-fold-change from our low-expression threshold of 5 CPM. As shown in **Figure S4**, only 14% of the transcripts reach 60 CPM in at least one cell type.

Next, we focused our efforts on candidate genes where the differentially expressed transcripts impacted on the protein sequence. Genes were removed from the list if the query and/or alternative transcripts: 1) encode for the same protein isoform, or 2) do not code for a protein. Finally, we identify if query transcripts of filtered candidate genes are known to have an isoform specific function based on searches with PubMed, GeneCards, UniProt, NextProt, and Google Scholar. To support the candidate gene results, correlation and corresponding p-value matrices are calculated for each candidate genes in cell type comparison used to select its. For this purpose, we use all individual expression data, pooled by subgroup of cell types. To identify a subgroup of interesting candidate genes, all query transcripts are compared to all alternate transcripts for each candidate gene and the comparisons with a significant ($p < 0.05$) negative correlation ($r < 0$) were used to select the candidate gene subgroup. Negative correlation represents an opposite correlation between the two comparative transcripts.

3.6 Results

3.6.1 Alternate promoter usage and differential splicing leads to a large diversity of transcripts in human immune cells

It is known that individual genes can give rise to multiple different RNA transcripts that result from alternate promoter usage and differential splicing events. As it is possible that different RNA transcripts of a given gene can encode proteins with different functional characteristics and/or stability that impact on the functions of specific cell types, we were interested in systematically identifying genes with alternative transcripts that were differentially expressed in primary immune cells. As a first step in this process, we characterized the diversity of transcripts in nine different primary immune cell populations from 12 healthy volunteers for which we had previously characterized the transcriptome at the gene level (54). Specifically, these were neutrophils, B cells, CD4+ T cells, CD8+ T cells, NK cells, $\gamma\delta$ T cells and monocytes isolated from

the peripheral blood. In addition, the monocytes were differentiated *in vitro* into macrophages, with a subset of these macrophages being activated with LPS.

To achieve transcript-level analyses, individual RNA sequence reads were mapped to known transcripts using a hierarchical indexing for spliced alignment of transcripts (HISAT), with individual transcripts being quantified with StringTie and annotated with the GENCODE annotation file (127, 128, 301). After data filtering and normalization, we detected a total of 33,283 transcripts from 13,424 genes (**Fig. 2.1**). Of these, 5,591 genes were found to have a single expressed transcript, whereas ~58% of the genes (n=7,833) had two or more transcripts detected in our data (**Figure 2.1**). While most genes with multiple transcripts had up to ten different transcripts, approximately 2% expressed between 11 and 33 different transcripts (**Figure 2,1**). Furthermore, we observed that the ratio of genes with a single transcript, versus those with multiple transcripts, appeared to be relatively similar from one cell type to another, including in neutrophils that overall expressed fewer genes (**S2.1 Table and S2.5 Fig**). In the subset of genes with multiple transcripts that are protein-coding, we observed that this diversity arose due to different TSS usage (~11 %), alternate splicing (~6 %) or a combination of both (~81 %).

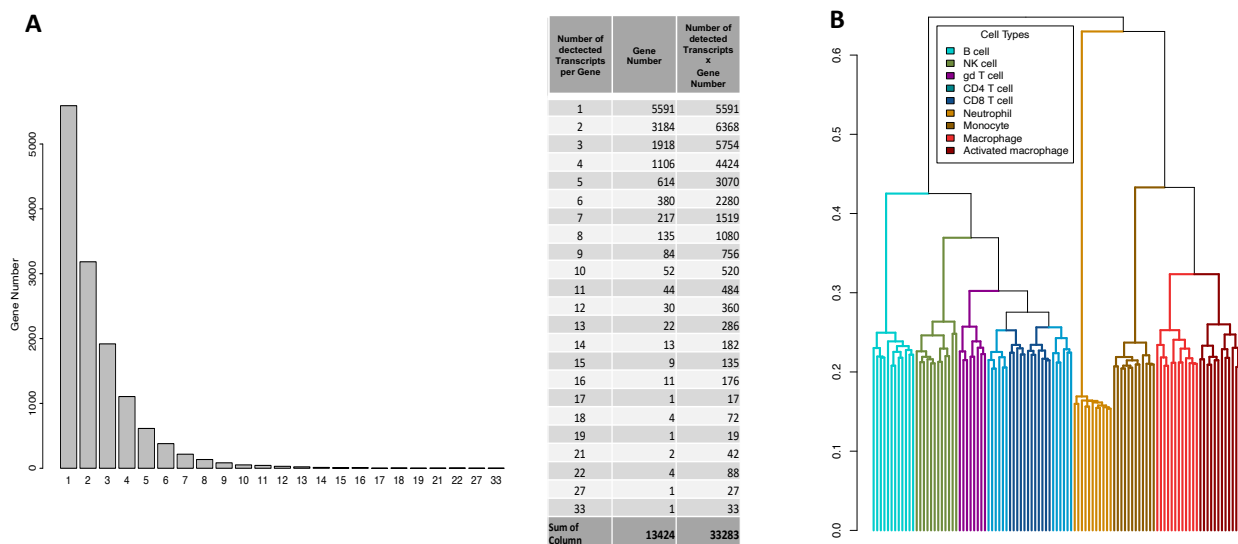


Figure 3.1 Transcript distribution in our data and unsupervised hierarchical clustering of transcript RNA-Seq data. A) Distribution of the number of detected transcripts per gene (histogram), with tabular representation. **B)** Dendrogram representation of the unsupervised hierarchical clustering of all samples (cell types and individuals) based on transcripts. The tree has a structure that closely resembles the known differentiation scheme of immune cells (see **Figure S3**). Each dendrogram leaf represents a single RNA-Seq sample, the Y-axis represents the distance based on Pearson correlation.

Principal component analysis (PCA) of this data revealed a clear separation of the cell types on the first two components, representing the nine populations into four very distinct clusters (neutrophils, lymphocytes, monocytes, and macrophages), with all individuals clustering tightly within each cell population (**S2.2 Fig**). This is concordant with the observation that 54% of the

total variance is between the cell types with the first two principal components capturing a large proportion of the total variance (25% and 16 % respectively). This supports there being greater differences in transcript expression between immune cell populations than between individuals, consistent with what we previously observed in our gene-level analyses of immune cell transcriptome (54). Using an unsupervised hierarchical clustering of this data we found that the dendrogram also reflected the differentiation scheme of pluripotent stem cells into myeloid and lymphoid cell types, although there was some overlap between CD4+ and CD8+ T cells (**S2.3 Fig.**). Given these observations, we propose that there is potential to identify differences in gene expression at the level of isoforms between immune cell populations that reflect the functional biology of these populations.

3.6.2 Systematic bioinformatic screen identifies genes where two or more transcripts have distinct expression patterns in primary human leukocytes.

It is well established that all cells of the immune system derive from a common pluripotent stem cell, where gene expression programs play a central role in cell fate decisions as well as maintenance of specific cellular functions in the resulting cell lineages. We therefore used the differentiation scheme (**S2.3A Fig.**) to guide our analyses of differential expression at the level of individual transcripts. Specifically, for each of the genes with two or more transcripts, we compared the expression pattern of the different transcripts for the cell populations on either side of a given branchpoint in the differentiation scheme (myeloid vs lymphoid, B cells vs T cells, $\alpha\beta$ T cells vs $\gamma\delta$ T cells, NK cells vs B and T cells, monocytes vs macrophages, and macrophages vs activated macrophages; **Table 1 and 2.**

Table 3.1 **Genes with at least two transcripts with distinct expression patterns between cell types or cell lineages.**

Comparison ¹	GROUP A	GROUP B	GROUP C
	Candidate Genes with Differentially Expressed Transcripts (DET)	Subset of GROUP A Affecting Coding Sequence	Subset of GROUP B where Query and Alternate Transcripts are Negatively Correlated and P<0.05
1.1 Monocytes vs Lymphocytes	SEPTIN9, PLEC, PPT1, ENTPD1, FYB1, PTPRE, RPS6KA1, HIPK3, NPC2, CD300A, NCOA4	CD300A, ENTPD1, FYB1, HIPK3, NPC2, PLEC, PTPRE, RPS6KA1, SEPTIN9	CD300A, FYB1, SEPTIN9
1.1 Lymphocytes vs. Monocytes	PTMA, JAK1, CDC42SE1, CNN2, RPL13A, GPI, ATM, FYB1, ATP2A3, ST6GAL1, CYFIP2, TSC22D3, EIF3F, CD164, LITAF, FAM107B, FGD3, RPL32, PTPRC, FCGR3A, RUNX3, IL4R, PARP8,	ANXA6, ARL4C, ATP2A3, CD164, CD300A, CNN2, ESYT2, FYB1, GPI, IL4R, LITAF, PSMA1, PTMA, PTPRC, RPL32, RUNX3, SH3KBP1, SWAP70, TLE5, TSC22D3	CD300A, FYB1, LITAF, PSMA1, PTPRC, SH3KBP1, GPI, PTMA, RPL32,

	ANXA6, FGR, SEPTIN7, ESYT2, TLE5, PSMA1, SPN, SH3KBP1, CD300A, SWAP70, CD53, SYK, ARL4C		
1.2 NK vs (B cells, T cells)	ITGB2, SUN2	ITGB2, SUN2	
1.2 (B cells, T cells) vs NK	RACK1, SH3BP2, RPS16, FYB1, ST6GAL1, RPS18, RPL4, FGD3, CD3E, CD53, SYK	FYB1, RPS16, SH3BP2	SH3BP2
1.3 B cells vs T cells	SH3BP2, ST6GAL1, CD53	SH3BP2	SH3BP2
1.3 T cells vs B cells	CNN2, ATM, ITGB2, LITAF, FAM107B, FGD3, PTPRC, ITM2B, RUNX3, ETS1, SH3KBP1	CNN2, ETS1, ITGB2, LITAF, PTPRC, RUNX3, SH3KBP1	LITAF, PTPRC
1.4 (CD4+ T cells, CD8+T cells) vs $\gamma\delta$ T cells	.	.	.
1.4 $\gamma\delta$ T cells vs (CD4+ T cells, CD8+ T cells)	.	.	.
1.5 CD4+ T cells vs CD8+ T cells	.	.	.
1.5 CD8+ T cells vs CD4+ T cells	.	.	.
B cell specific	ST6GAL1, SWAP70, CD53, SYK	SWAP70	.
NK cell specific	.	.	.
$\gamma\delta$ T cell specific	.	.	.
CD4+ T cell specific	.	.	.
CD8+ T cell specific	.	.	.

Table 3.2 Genes with at least two transcripts that are differentially expressed between monocytes, macrophages and LPS-activated macrophages

Comparison ¹	GROUP A	GROUP B	GROUP C	We first
	Candidate Genes with Differentially Expressed Transcripts (DET)	Subset of GROUP A Affecting Coding Sequence	Subset of GROUP B where Query and Alternate Transcripts are Negatively Correlated and P<0.05	
2.1 Monocytes vs Macrophages	CD302, ANXA1, SEPTIN9, RPL27A, DAZAP2, LTA4H, NFKBIA, EEF1A1, PTPRC, ALDH2, KLF6, RPL29	SEPTIN9, DAZAP2, NFKBIA, EEF1A1, PTPRC, KLF6	KLF6, SEPTIN9	
2.1 Macrophages vs Monocytes	SEPTIN9, CTSD, CAPG, ITGB2, ANXA2, NPC2, FNBP1, FURIN	SEPTIN9, CTSD, ITGB2, ANXA2, NPC2, FNBP	FNBP1, SEPTIN9	
2.2 Macrophages vs Activated macrophages	GSN, SLCO2B1, LIPA	GSN, SLCO2B1, LIPA	.	
2.2 Activated macrophages vs Macrophages	CD44, OAZ2, NR3C1, SLAMF7, SNX10	CD44, OAZ2, SLAMF7	CD44, OAZ2	

focused our analyses on monocytes and the five lymphoid cell populations (B, NK, $\gamma\delta$ T, CD4+ T and CD8+ T) and included the 6,985 genes with at least two different transcripts (**S2.2 Table**). The neutrophils were not included at this stage as their overall transcriptome was quantitatively different from the other populations (**S2.1 and S2.5 Fig.**). In order to identify which of these 6,985 genes have RNA isoforms that are differentially expressed, for each gene we: (1) defined as a “query” any transcript that had high expression (geometric mean > 60 CPM) in the population/group of interest (e.g. lymphoid cells) and low expression (geometric mean < 5 CPM) in the opposing population/group of the branchpoint (e.g. myeloid cells), and then (2) searched for a different transcript (called “alternate”) of the same gene that was high in the latter population/group (geometric mean > 60 CPM), regardless of its expression level in the former population/group. Using this approach, we identified 55 genes with at least two transcripts that were differentially expressed, with some genes being identified in multiple comparisons (**“Group A” in Table 2.1**). These include genes encoding cell surface receptors or transporters, TF, transcriptional regulators, ribosome components, enzymes, endosomal and lysosomal proteins, and proteins involved cytoskeleton organization.

As a next step, we focused our efforts on the subset of genes where the differentially expressed transcripts encode for different protein isoforms resulting from either different TSS usage,

different splicing events, or a combination of both. This identified a subset of 33 candidate genes (**“Group B” in Table 2.1; S2.2 Table**). To better quantify the differential expression of these transcripts, we also calculated the correlation and associated p-value between query and alternate transcripts and identified 11 genes (*CD300A*, *FYB1*, *GPI*, *LITAF*, *PSMA1*, *PTMA*, *PTPRC*, *RPL32*, *SEPTIN9*, *SH3BP2*, *SH3KBP1*) where there was a negative correlation between the query and alternate transcripts with nominal significance $P < 0.05$ (**“Group C” in Table 2.2; S2.2 Table**). We considered these as being strong candidates for genes that have multiple transcripts with different patterns of expression and giving rise to protein isoforms with potentially different functions.

Of these 11 genes where at least two different transcripts have significantly divergent patterns of expression, the *PTPRC* (Protein Tyrosine Phosphatase Receptor Type C) gene is of particular interest as it encodes the CD45 protein. Studies in humans and mice from over two decades ago demonstrated that CD45 is a key molecule in antigen receptor signaling and lymphocyte development, with different protein isoforms being present at the surface of cells given the cell lineage and differentiation state (305) (306). Our analyses detected a total of five *PTPRC* transcripts and identified a distinct pattern of expression for *PTPRC-201* (encoding CD45RO), *PTPRC-209* (encoding CD45RABC) and *PTPRC-203* (encoding protein isoform M3ZCP1) (**Figure 2.2**). All cell types had strong expression of *PTPRC-201* and *PTPRC-209*, although interindividual variability of expression was detected for the former in B cells and for the latter in macrophages, with strong expression of *PTPRC-203* in most T cell lineages, except for CD4+ T cells which showed higher interindividual variability. The *PTPRC-203* isoform is believed to generate a protein that has an extracellular region that is identical to CD45RBC but a cytoplasmic region with its two protein tyrosine phosphatase domains either partially truncated or deleted, and thus likely to be catalytically inactive while still capable of binding its ligand.

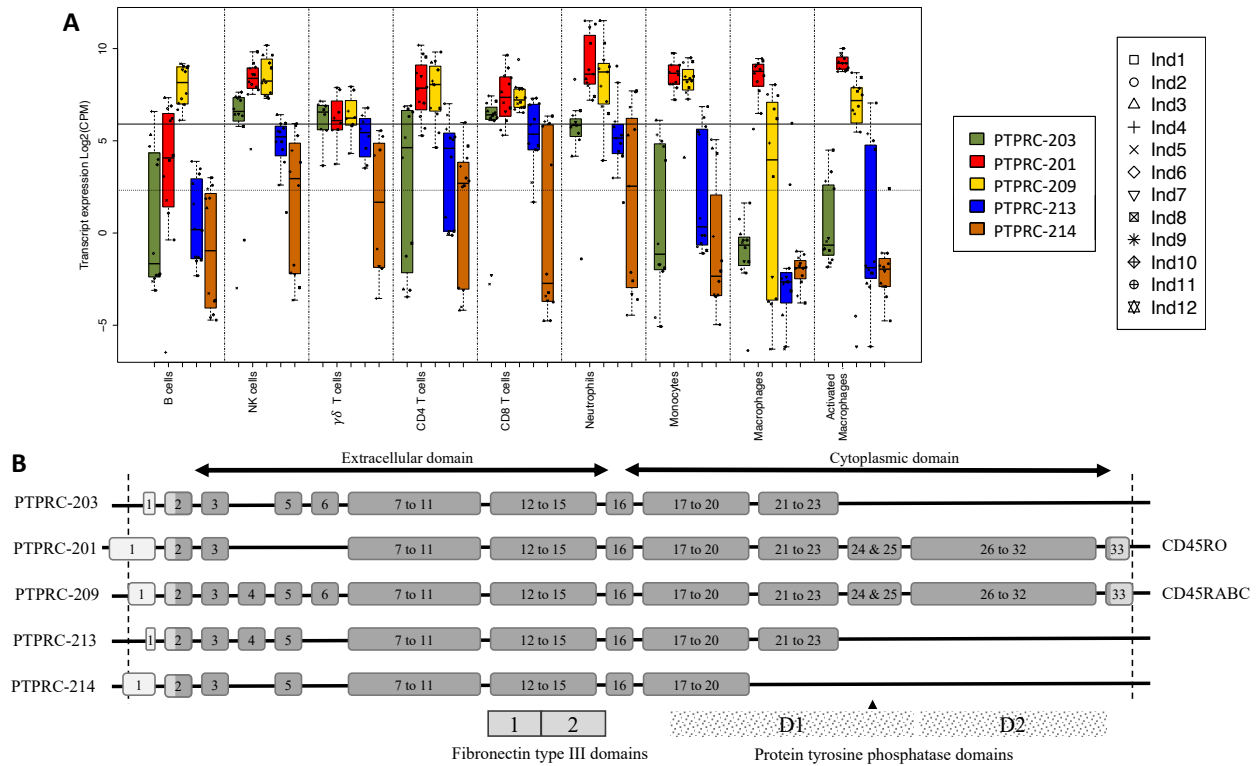


Figure 3.2 Expression pattern and encoded protein structures of *PTPRC* transcripts across immune cell populations. **A)** Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations. Dotted line represents the minimal threshold to identify low-expression transcripts ($\log_2(5 \text{ CPM})$) in our dataset; the dark line represents the minimal threshold value used to define moderately/highly expressed transcript ($\log_2(60 \text{ CPM})$). Each individual is represented by a different symbol. The Box illustrates the four quartiles of data, with the middle bold line indicating the median value; the box then includes 50% of the data points, 25% are above and 25% below. The whiskers extend up to 1.5 times the interquartile range (height of the box). Data points outside this range are generally considered outliers. **B)** Graphical representation of the different alternate transcripts, identified in our data, and their respective protein isoforms for the *PTPRC* gene. The transcript name from GENCODE is on the left while the protein name from UniProt is on the right when available. Exonic regions in dark grey are coding sequences, while regions in light grey are untranslated regions. The junctions between light and dark grey indicate the location of translation start and stop sites. The *PTPRC* gene encodes CD45, which is a type 1 transmembrane glycoprotein, having a single transmembrane domain, an extracellular region containing two fibronectin type III domains and a cytoplasmic tail containing two phosphatase domains. The arrowhead represents the ligand binding domain. *PTPRC-209* codes for the canonical protein isoform. The locations of the functional protein domains within the protein isoforms are also shown.

While not as extensively studied as *PTPRC*, there is also prior evidence in the literature for the FYN Binding Protein 1 (*FYB1*), Prothymosin alpha (*PTMA*), and SH3 Domain Containing Kinase

Binding Protein 1 (*SH3KBP1*) genes having differentially expressed transcripts that encode proteins with differing functions (see **Box 1** for details in **Annexe: Supplementary Text**). In terms of *FYB1* and *PTMA*, we have identified a greater complexity in the number and patterns of transcripts expressed in immune cells than previously reported, and for *SH3KBP1* studies had primarily been limited to T cells. Intriguingly, although not previously studied in terms of differential transcript expression, one of our other candidates, *SH3BP2* encodes a known *SH3KBP1* interacting protein (see **Box 1** for details in **Annexe: Supplementary Text**).

For the remaining “**Group C**” genes (*GPI*, *LITAF*, *CD300A*, *RPL32*, *PSMA1*, *SEPTIN9*), differential transcript expression has not been previously reported in immune cells, but all were found to have distinctive patterns of expression in our data (**Box 1**). For example, *GPI* was found to have a transcript specific to cytotoxic cells, as well in activated macrophages, and another that is more ubiquitous. *CD300A* had a transcript primarily expressed in myeloid cells and another specific to NK and $\gamma\delta$ T cells. *RPL32*, on the other hand, had multiple transcripts with moderate to high expression in lymphoid cells, with a few of these with variable low-to-no expression in myeloid cells. Finally, genes like *PSMA1* and *SEPTIN9* had multiple transcripts with complex expression patterns. *SEPTIN9* is a member of the septin family cytoskeletal GTPase, and is involved in many biological processes including cytokinesis, polarization, vesicle trafficking, cell migration, and apoptosis (307). It has previously been reported that different isoforms of *SEPTIN9* exist in mammary epithelial cells, with isoforms differing in their tumorigenesis potential (308). In the current study, *SEPTIN9-201* was strongly and consistently expressed in myeloid cells and essentially absent from lymphoid cells although some expression was observed in B cells in some individuals (**Figure 2.3**). The *SEPTIN9-202* transcript was highly expressed in all primary cell populations examined but decreased in expression when monocytes were differentiated into macrophages and further decreased when these were activated by LPS. Interestingly, *SEPTIN9-208* had an almost identical pattern of expression to *SEPTIN9-202* but was absent in neutrophils. The *SEPTIN9-203* transcript, in opposition, was only observed in macrophages and LPS-activated macrophages. All four of these *SEPTIN9* transcripts encode proteins that differ in length due to different TSS being used as well as alternate splicing, resulting in variable N-terminal sequences. These different N-terminal sequences are expected to differentially bind other septins within its complex and/or to impact on the binding and bundling of microtubules likely impacting on its downstream functions (307, 309).

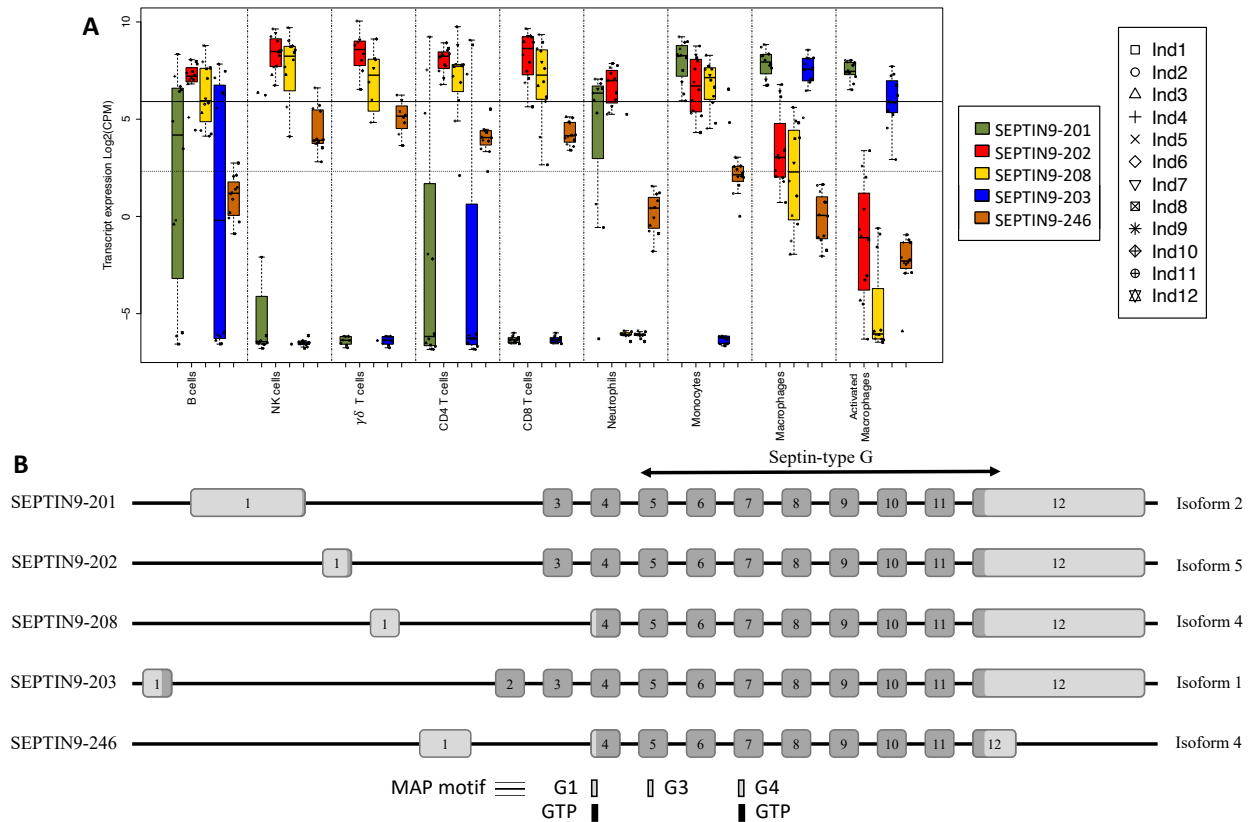


Figure 3.3 Expression pattern and encoded protein structures of *SEPTIN9* transcripts across immune cell populations. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **Fig. 2**). **B)** Graphical representation (as in **Fig. 2**) of the different *SEPTIN9* transcripts, identified in our data, and their respective protein isoforms. *SEPTIN9* encodes a cytoskeleton protein. *SEPTIN9* polymerizes with other members of the septin family of proteins to generate protein complexes that form filaments. *SEPTIN9* possesses microtubule associated protein (MAP)-like motif as well as a GTPase domain with conserved motifs G1, G3, and G4 that are characteristic of GTP-binding proteins (310). *SEPTIN9-203* codes for the canonical protein isoform.

3.6.3 Identifying differentially expressed isoforms of protein-coding genes in monocytes, macrophages and activated macrophages

We then applied this same approach to compare the transcriptomes of human monocytes, macrophages and macrophages activated by LPS from this dataset. Specifically, we detected a total of 23,756 different transcripts encoded by 11,301 distinct genes, with 5,773 of these genes expressing two or more transcripts (**S2.3 Table**). Performing the same analysis as described above we identified 27 genes that have differentially expressed transcripts across the monocyte, macrophage and activated macrophages (**Table 2.2 and S2.4 Table**). These genes encode cell surface receptors or transporters, TFs, ribosome components, enzymes, endosomal and lysosomal proteins, and proteins involved in signaling cascades or cytoskeleton organization

("Group A" in Table 2.2), with 17 affecting the protein coding sequence ("Group B" in Table 2.2; S2.4 Table). Examining the individual-level data for these 17 genes, we identified five where there was a negative correlation between the query and alternate transcripts with nominal significance $P < 0.05$ ("Group C" in Table 2.2; S2.4 Table). These are *CD44*, *FNBP1*, *KLF6*, *OAZ2*, and *SEPTIN9*. Specifically, *FNBP1*, *KLF6*, and *SEPTIN9* had transcripts that were significantly differentially expressed between monocytes and macrophages, and *CD44*, *OAZ2*, and *SEPTIN9* had transcripts that were significantly differentially expressed between macrophages and macrophages activated by LPS exposure (see Box 1 for details in Annexe: Supplementary Text). *SEPTIN9* was also identified above in the broader analyses of lymphoid and myeloid cells as described above.

Of these five top candidates, *CD44* is well known to have multiple different protein-coding transcripts. Also known as Cluster of Differentiation 44, CD44 is a cell surface adhesion molecule involved in cell-cell and cell-matrix interactions of immune cells. We detected six different protein coding transcripts of *CD44* in our dataset (Figure 2.4). All six transcripts were highly expressed in LPS-activated macrophages, of which four (*CD44-201*, *CD44-206*, *CD44-208*, *CD44-209*) encode known protein isoforms (CDw44, CD44-epidermal, CD44, and CD44E, respectfully), while *CD44-202* and *CD44-207* are highly truncated. Monocytes, on the other hand, strongly expressed *CD44-201* and *CD44-209* but did not express *CD44-208*. Important interindividual variance in transcript *CD44-206* expression was observed, with this transcript being undetectable in most. Macrophages had a similar pattern of transcript expression, although the differentiation from monocytes to macrophages led to an increase in expression of transcripts *CD44-206* and *CD44-208* in some but not all individuals. Interestingly, the expression pattern in lymphocytes and neutrophils was most like that of monocytes. As previously reported, this sequence diversity is primarily focused within the stem region where there is insertion of one or more variant exons that will affect binding of known ligands such as hyaluronic acid (HA), osteopontin (OPN) and matrix metalloproteinases (MMPs), as well as downstream signaling (311). Importantly, it has previously been shown that CD44 regulates TLR2-linked macrophage activation and subsequent inflammatory responses, is required to maintain inflammatory monocytes within the circulation during inflammation, and promotes phagocytosis in macrophages (312, 313) (314).

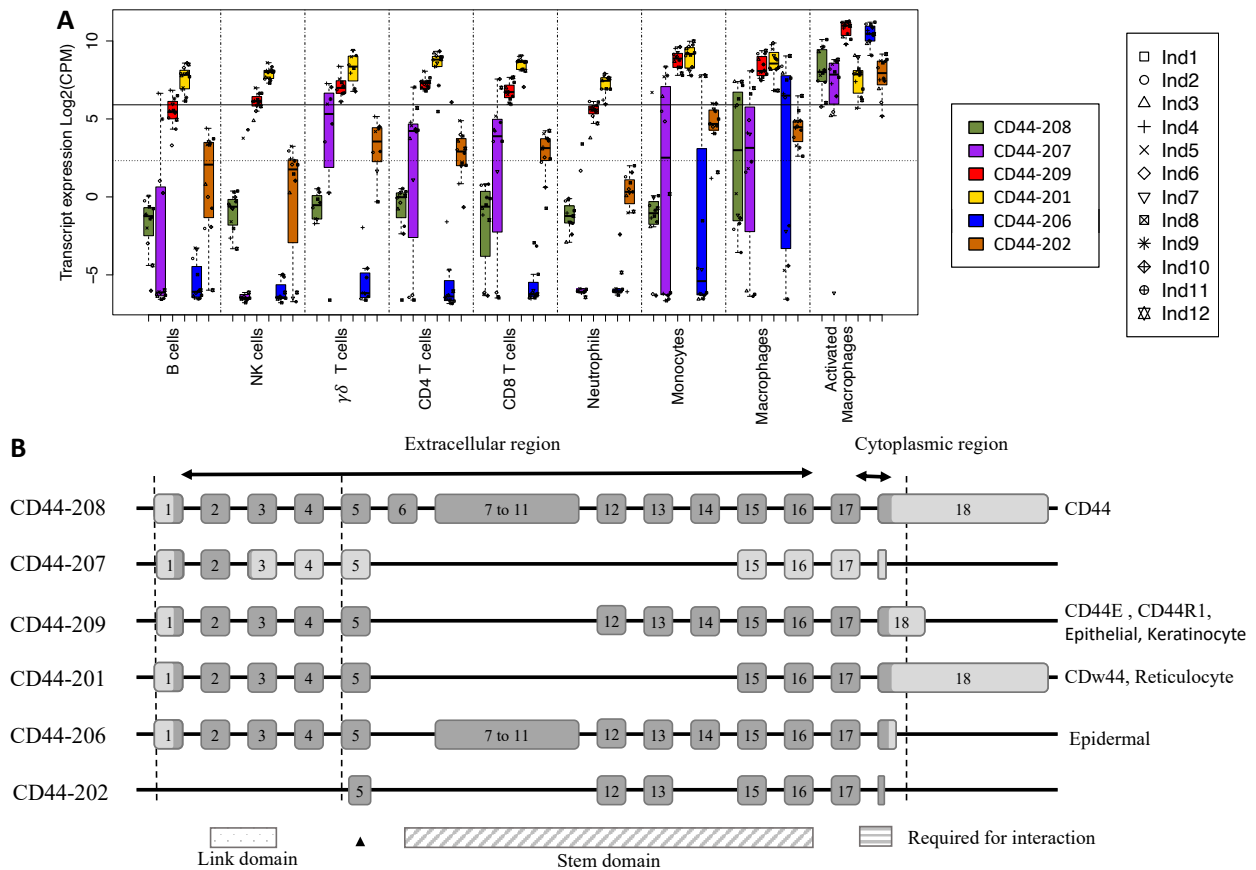


Figure 3.4 Expression pattern and encoded protein structures of *CD44* transcripts across immune cell populations. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **Fig. 2**). **B)** Graphical representation (as in **Fig. 2**) of the different alternate transcripts, identified in our data, and their respective protein isoforms for the *CD44* gene. The *CD44* protein is a cell membrane molecule that mediates cell-cell and cell-matrix adhesion and consists of a link domain important for hyaluronan binding, the stem region, and a cytoplasmic region. The dark grey portion of exon 1 encodes the protein's signal peptide, whereas the dark grey portion of exons 16 and 17 code for transmembrane region. The arrowhead represents the Arg/Lys-rich domain. The box "Required for interaction" is the site of interaction with EZR, MSN and RDX and for co-localization to microvilli. *CD44-208* encodes the canonical protein isoform.

This approach identified Krüppel-like factor 6 (*KLF6*) which encodes a transcription factor that has been implicated in cellular differentiation and cell proliferation, but its aberrant expression is involved in many cancer types (e.g.: glioblastoma, ovarian, lung, and prostate carcinoma). The interaction of *KLF6* with the c-jun oncoprotein gives it a potential role as a tumor suppressor (315, 316). This protein is also known to be involved in B cell growth and development (315). We detected three different protein-encoding transcripts in our data that were differentially expressed between monocytes and macrophages (**Figure 2.5**): *KLF6-204* that is strongly expressed in monocytes but is essentially lost once these are differentiated into macrophages; *KLF6-206* expressed weakly in monocytes and strongly induced in macrophages and LPS-

activated macrophages; and *KLF6-207* that was most strongly expressed in LPS-activated macrophages. Consistent with our results, *KLF6-206* (aka isoform 1) has been reported to regulate macrophage polarization to the M1 phenotype and its suppression by miR-181a promoted M2 phenotype (316, 317), although our data suggests that *KLF6-207* should also be examined in this regard. Also very relevant, KLF6 has been shown to be involved in macrophage motility and the regulation of pro-inflammatory gene expression including *MCP1*, *CXCL2*, and *IL8*, by cooperating with NF- κ B and suppressing PPAR γ and BCL6 (316-319). KLF6 has a role in the hypoxic-response transcriptional program in macrophages and promotes pro-inflammatory response (316, 320). KLF6 transactivates the expression of inducible nitric oxide synthase (nNOS) and endothelial NOS (eNOS) to regulate NO production (316).

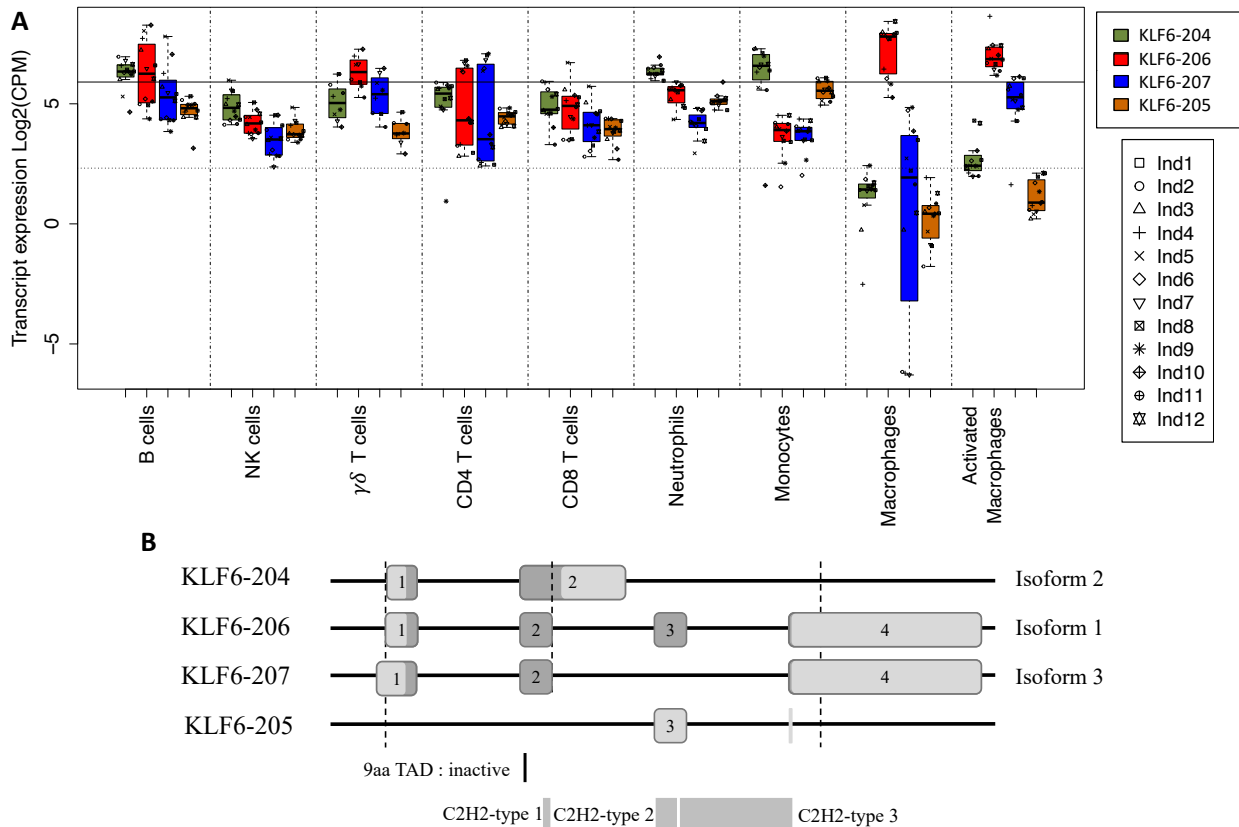


Figure 3.5 Expression pattern and encoded protein structures of *KLF6* transcripts across immune cell populations. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations. (as described in Fig. 2). B) Graphical representation (as in Fig. 2) of the different alternate transcripts, identified in our data, and their respective protein isoforms for the *KLF6* gene. This gene encodes the KLF6 protein, which is a zinc finger containing nuclear protein. The 99aaTAD domain is a transactivation domain and its inactive in KLF6 protein. The three C2H2-type zinc finger structural motifs bind to GC-rich regions in DNA. *KLF6-206* codes for the canonical protein isoform.

3.7 Discussion

It has long been recognized that there is a diversity of cell types within the immune system that play different roles in the protection of the human host from germs, parasites, and abnormal cells. This diversity is characterized by the differences in proteins expressed at the cell surface, in how cells respond to different stimuli, in the signaling cascades that enable immune cells to carry out their important functions, and much more. The diversity and flexibility of the immune system and its constituent cells depends in part on gene transcription. Indeed, targeted, and genome-wide transcriptomic studies have provided an understanding of how the regulation of gene expression plays an important role in immune cell functions. These studies, however, have relied on an understanding of the transcriptome at the gene level despite the knowledge that individual genes have the capacity to generate multiple different transcripts that differ at the level of their untranslated and/or translated sequences. This diversity of transcripts arises from differential use of transcription start sites (TSS) and/or alternate splicing events. In the current study, we examined the expression differences of the different transcripts, or RNA isoforms, across nine different immune cell types from peripheral blood cells of twelve healthy donors and identified how different transcripts from individual genes are differentially expressed between cell populations and activation state.

In this study we observed that ~60% of the genes expressed in the immune cells that were studied herein expressed two or more different transcripts. This is consistent with findings from a combination of microarray and RNA-Seq data generated for mouse immune cells (138). In addition, while transcript-level data is noisier than gene-level data in any given dataset, non-supervised clustering of the current data performed well and reflected the identity and relationship between cell types, with some overlap between CD4 and CD8 T cells, suggesting that this approach is capturing relevant biological differences (**Figure 2.1**).

With the objective of determining the extent to which this diversity in transcripts can contribute to the diversity of functions across immune cells, our approach focused on whether there are genes with multiple different transcripts where at least two transcripts have distinct patterns of expression across the primary cell types studied herein. This analysis did not examine all pairwise combinations of cell types but rather focused on comparisons that were based on the known branchpoints in hematopoietic differentiation. A heuristic approach was used based on identifying a transcript with high expression in one cell population or group on one side of a branchpoint, with a different transcript for the same gene having a distinct pattern of expression in the cells on the other side of the branchpoint. We next focused our analyses on the instances where the different transcripts encoded different protein sequences. Among the genes identified in this process was the “gold standard” for differential splicing in immune cells – the *PTPRC* gene encoding CD45. The three most abundant transcripts that we observed (*PTPRC-201*, *PTPRC-203*, *PTPRC-209*) for this gene differ not only in their expression patterns but also in their 5'UTR, alternate usage of exons 4-6, and C-term sequence (**Figure 2.2**). It has previously been demonstrated that alternate splicing of exons 4, 5 and 6, leads to different extracellular domains, and that the different CD45 protein isoforms have different thresholds of sensitivity and differential disposition to autoimmunity (305, 321). This approach identified 10 other genes having transcripts with differential expression when focusing on primary immune

cells, including other examples with prior evidence in the literature such as *FYB1*, *PTMA*, and *SH3KBP1*. We then applied this same approach to identify genes with differential transcript expression that are potentially involved in the transition from monocytes to macrophages to activated macrophages. Again, this approach identified a well-known gene (CD44) as well as novel genes with differential patterns of expression between monocytes and macrophages, and/or between macrophages and macrophages activated by LPS.

A novel finding of particular interest identified in both analyses is that of *SEPTIN9*, where we found a predominant isoform in lymphocytes (*SEPTIN9-202*; a.k.a. i5), another in myeloid cells (*SEPTIN9-201*; i2) and with differentiation to macrophages a switch to another transcript (*SEPTIN9-203*; i1). Septins are known to be overexpressed in multiple tumor types and are regulators of the cytoskeleton via their interactions with actins filaments and microtubules. It has very recently been demonstrated that *SEPT9* mediates the binding of septins to microtubules via a short microtubule-associated protein (MAP) motif (**Figure 2.3**) (310). It is important to note that this MAP motif is encoded by exon 2, which is unique to (*SEPTIN9-203*; i1). The current study provides the first evidence that *SEPT9* transcripts are differentially regulated in immune cells. Microtubules, however, are known to play multiple different roles in immune cells such as cell migration, first studied in the chemotactic response of neutrophils, immune synapse formation in T cells and antigen presenting cells, and transport of mitochondria which is required for NLRP3 inflammasome activation (322, 323). Recent studies in osteosarcoma, adenocarcinoma, retinal pigment epithelial and embryonic kidney cell lines have revealed differential expression of *SEPT9* transcripts and that the *SEPTIN9-203/i1* isoform is the one that binds to microtubules via a 25 amino acid N-terminal motif. This binding was found to result in the stabilization of microtubule octamers (310, 324). Importantly, it has previously been shown that macrophage activation results in increased levels of stabilized cytoplasmic microtubules, which then can be used for intracellular trafficking (323). For example, it been shown that MMP-9 containing vesicles are trafficked along stable subsets of microtubules, leading to the extracellular secretion of MMP-9, which is a key component enabling endothelial transmigration (325). Other isoforms of *SEPTIN9*, on the other hand, could not bind microtubules but were found to bind actin filaments (310, 324). The actin cytoskeleton plays a central role in a multitude of cellular processes that are essential to immune cell functions, but further work will be necessary to determine if and how the different *SEPTIN9* isoforms impact on these.

We are confident in the validity of our approach as it identified genes previously reported to have multiple protein-encoding transcripts that are differentially expressed across immune cells (e.g. CD45, CD44). It should be noted, however, that the thresholds employed herein could be modified to be more stringent or more permissive, affecting the number of candidates requiring follow-up analyses. It should be emphasized that although we focused on transcripts that affect protein sequence, the other genes that we identified herein are of potential interest as different transcripts can have other biological impacts such as on mRNA stability (326-328). It is also important to understand the caveats that relate to the current approach. The first caveat is that this approach is based on short read genome-wide transcriptomics without validation of the structure of the transcripts or of the resulting proteins. It should be said, however, that this

concern motivated the decision to not perform *de novo* transcript assembly, but to use reference transcripts in the mapping of sequence reads to individual transcripts. This is also the reason that we limited our analyses to transcripts that encode different protein isoforms found within the UniProt Knowledgebase (329). A second caveat is that since this is a transcriptomic approach, there is no direct measurement of the abundance of the different protein isoforms. While genome-wide studies examining the correlation between mRNA expression levels and protein abundance have found this correlation to be relatively poor, it has been found that for differentially expressed genes this correlation is significantly better and thus supports the usefulness of inferences from mRNA expression (330-332). Another caveat is that while the different protein isoforms encoded by the transcripts that we identified may have important biological differences, for example in terms of protein-protein interactions, stimulatory or inhibitory activities, stability, DNA-binding properties, etc., it is known that alternative transcripts can, however, also lead to proteins isoforms with disrupted structures that could target the transcripts for nonsense-mediated RNA decay. While this is less likely, again as we have focused our approach on protein isoforms with evidence within the UniProt Knowledgebase, we have not directly performed functional validation. In fact, what these caveats highlight is the intended purpose of the current approach, which is to identify genes where there is evidence that differential transcript expression plays a role in modulating the function of a gene or biological pathway and to help generate biologically plausible hypotheses that require functional validation.

In conclusion, the approach presented in this article successfully identified known genes with multiple transcripts that are differentially expressed across human immune cells, as well as identified multiple novel candidates where this also appears to be the case. As differential expression of genes and transcripts provide complementary information, this approach should prove useful for the analysis of other datasets, such as when comparing transcriptome profiles between patients and healthy individuals to identify pathogenic pathways or between patients with different clinical outcomes to identify blood-based biomarkers.

3.8 Funding

This work was supported by grants from the Canadian Institutes of Health Research: (1) The “CIHR Emerging Team in Integrative Biology of Inflammatory Diseases” [CIHR #102170] with co-funding from Crohn’s and Colitis Canada made possible via a generous donation from the Ross McMaster Memorial Fund; (2) the IMAGINE-SPOR Chronic Disease Network [CIHR #SCA-145105]. Salary support to Saraï Mola was provided by the Canada Research Chair in Genetics and Genomic Medicine awarded to John David Rioux [CRC#230625]. Sylvie Lesage holds a Research Scholar Emeritus award from the Fonds de recherche du Québec – Santé (FRQ-S). This research was enabled in part by support provided by Calcul Québec (<http://www.calculquebec.ca>) and Compute Canada (www.computecanada.ca).

3.9 Competing interests

The authors declare that no competing interests exist.

4 Chapitre 4 – Conclusion

4.1 Conclusion sur l'étude du transcriptome pour mieux comprendre les fonctions des cellules immunitaires

Toutes les cellules immunitaires circulantes sont le produit de la différenciation des cellules souches hématopoïétiques multipotentes. L'hématopoïèse, qui est le processus à l'origine de cette différenciation, débute dans la moelle osseuse et permet, à elle seule, de construire la plus grande partie des cellules indispensables au système immunitaire. Ce processus de différenciation permet la perte et le gain de fonctions cellulaires qui par la suite vont être spécifiques ou partagées parmi les types cellulaires immunitaires primaires. Les différences caractérisant chaque population de cellules immunitaires sont utilisées pour les identifier, les quantifier et mesurer la pureté cellulaire d'un échantillon de cellules. La cytométrie de flux permet de mesurer l'ensemble de ces paramètres par l'identification de marqueurs de surface (ex. : CD14 pour les monocytes, CD15 pour les neutrophiles et CD19 pour les cellules B). Pour que ces marqueurs soient présents à la surface cellulaire, les gènes codant pour ces protéines doivent être exprimés au préalable dans la cellule. Il existe aujourd'hui de nombreux outils permettant de quantifier le transcriptome d'une population ou d'une cellule dont le séquençage de nouvelle génération (ex. : Séquençage d'ARNm). L'analyse du transcriptome est une bonne approche pour mieux comprendre les différences et ressemblances entre plusieurs types cellulaires immunitaires, conditions expérimentales ou états pour une maladie. La quantification du grand nombre de données obtenues par séquençage d'ARNm est un défi que certains développeurs ont relevé, ce qui a permis d'offrir un choix non négligeable d'outils mesurant l'expression par gènes, transcrits ou exons (ex. : TopHat, DEXSeq, StringTie, Salmon et Sailfish). L'analyse et l'interprétation de ces mesures de l'expression transcriptomique provenant de cellules immunitaires primaires saines restent tout autant complexe mais permettraient de mieux comprendre la régulation de l'expression génique ainsi que les fonctions cellulaires spécifiques et partagées. Les connaissances obtenues pourront par la suite être utilisées dans des études de maladies auto-immunes ou ciblant le système immunitaire.

Dans cette thèse, deux axes d'analyse du transcriptome ont été exploités. Le premier axe a été développé autour de l'hypothèse que les gènes qui sont impliqués dans une fonction biologique commune ont des profils d'expression corrélés. Pour cela, l'étude de l'expression génique avec l'outil WGCNA a permis de construire des modules de gènes fortement corrélés entre eux. L'analyse des listes de gènes co-exprimés composant ces modules, a permis de comprendre un peu mieux les fonctions cellulaires spécifiques et partagées entre les cellules immunitaires primaires. Le second axe est complémentaire au premier. Certains gènes du génome humain ont la capacité d'encoder plus d'une isoforme d'une protéine. Cette compétence génique interroge sur le lien qu'il pourrait exister entre l'expression spécifique d'une isoforme d'une protéine et un type cellulaire donné, et plus particulièrement la fonction cellulaire dans laquelle cette isoforme est impliquée. Basé sur ce questionnement, le second axe de la thèse étudie les différences des

profils d'expression des transcrits d'un gène dans les cellules immunitaires primaires. Cette analyse a permis d'identifier un certain nombre de gènes dont une isoforme est spécifiquement présente dans un ou plusieurs types cellulaires immunitaires. Pour un petit nombre de ces gènes, le rôle de la protéine au sein de la cellule est connu et dépend de l'isoforme exprimée.

4.1.1 La co-expression génique et la régulation transcriptionnelle permettent de mieux comprendre la fonction cellulaire

Dans les cellules immunitaires primaires, l'analyse des profils d'expression génique avec l'outil WGCNA a permis d'obtenir des modules de gènes co-exprimés. Le premier article de cette thèse, présente les résultats d'analyses globales et ciblées d'annotations des fonctions associées aux gènes co-exprimés dans chaque module. Dans les modules fortement associés aux cellules B, ces analyses sont complétées par l'étude d'enrichissement de sites de liaison de FT dans la région promotrice des gènes co-exprimés. Pour cela, les données ChIP-Seq du projet ENCODE, dans les lignées de cellules B, ont été utilisées. Pour finir, l'étude des données d'expression des gènes qui codent pour des FT, ont permis d'identifier des FT possiblement impliqués dans l'établissement et le maintien de la différenciation cellulaire dans l'hématopoïèse.

L'approche d'analyse par modules de gènes avec l'outil WGCNA a permis d'obtenir 45 modules contenant entre 1 945 et 26 gènes co-exprimés (**Figure 2.3 et Tableau 2.1**). Le HeatMap de la **Figure 2.3** montre que chaque module a une corrélation qui lui est propre avec chaque type cellulaire et que pour certains d'entre eux, on peut observer une forte corrélation avec un type cellulaire ou un sous-groupe de populations immunitaires. Par exemple, le module 22 est fortement corrélé aux cellules B alors que le module 41 est fortement associé aux cellules B, mais aussi aux monocytes. Le module 26, quant à lui, est spécifiquement corrélé aux cellules NK, T $\gamma\delta$ et T CD8+. L'analyse globale de chaque module a été réalisée avec l'outil DAVID pour calculer l'enrichissement d'annotations géniques. Cette analyse a montré que le module 22 est enrichi avec des gènes impliqués dans la fonction des cellules B (ex. : activation des cellules B, immunodéficiência primaire et voie de signalisation Akt-PI3K), le module 41 contient des gènes impliqués dans le CMH de classe II (ex. : mécanisme de présentation d'Ag) et le module 26 est enrichi pour des fonctions cytolytiques (ex. : cytotoxicité conduite par les cellules NK et cytolysse) (**Tableau S2.4**). Cette analyse est complétée par deux approches d'analyse ciblées. La première consiste à identifier pour chaque module le ou les types cellulaires auxquels il est le plus fortement associé et ensuite d'identifier la liste des « *Top Expressing Genes (TEGs)* » dans ces populations, qui sont présents dans le module étudié. Les TEGs sont les 2% des gènes avec l'expression la plus élevée dans un type cellulaire donné, ce qui correspond aux gènes des 99^{ème} et 98^{ème} percentiles dans notre étude. La seconde analyse ciblée utilise la première composante principale pour identifier les 20% des gènes qui représentent le module. Ils sont appelés « *Module Representative Genes (MRGs)* » dans l'article. Ces analyses ciblées ont confirmé que le module 22 contient des gènes impliqués dans les fonctions des cellules B et que le module 41, associé aux cellules B et aux monocytes, est composé de gènes essentiels à la présentation d'Ag du CMH de classe II. Pour le module 26 qui lui, est fortement associé aux cellules NK, T $\gamma\delta$ et T CD8+, les listes de TEGs et de MRGs contiennent des gènes qui codent pour les protéines composant les granules cytotoxiques ainsi que les protéines de la famille des LRC et NKC, qui sont des récepteurs impliqués dans l'identification des cellules du non-soi (**Tableaux 2.1 et S2.6**).

Le projet ENCODE a pour objectif d'annoter l'ensemble du génome et du transcriptome humain, dont les éléments de régulation de l'ADN (ex. : FT et histones). À l'aide de données ChIP-Seq, ils ont identifié des sites de liaison (SL) pour plus de 160 FT, dont 76 dans des lignées de cellules B. Malheureusement, ces FT n'ont pas été étudiés dans d'autres lignées cellulaires qui pourraient être représentatives de nos cellules immunitaires primaires. De ce fait, l'analyse d'enrichissement de SL dans la région promotrice des gènes co-exprimés d'un module a été limitée aux modules fortement associés aux cellules B, soit 12 modules au total comprenant les modules 22, 38 et 41 (**Tableau S2.7**). Sur les 76 FT, 45 ont été enrichis dans au moins un des 12 modules associés aux cellules B. Parmi eux, se trouve les FT MYC et NFATC1, seuls FT qui ont un enrichissement de SL dans le module qui contient le gène qui les encode. Une grande majorité des FT enrichis dans ces 12 modules sont connus pour être impliqués dans la régulation de l'expression génique dans les cellules B (ex. : BCL11A, PAX5, EBF1 et MTA3).

Pour finir, une dernière analyse a été réalisée en s'appuyant uniquement sur l'expression différentielle des gènes qui codent pour des FT dans les différents embranchements du processus de l'hématopoïèse (**Figure 2.2**). Pour chaque embranchement du processus de différenciation (ex. : cellules myéloïdes versus cellules lymphoïdes), l'expression génique des FT est comparée entre les deux branches afin d'identifier ceux susceptibles d'être impliqués dans l'établissement et le maintien d'un sous-groupe de cellules immunitaires primaires (**Figure S2.5 et Tableau S2.5**). Sur les 74 FT différentiellement exprimés dans un type cellulaire ou un sous-groupe de cellules immunitaires, 22 d'entre eux sont connus pour être impliqués dans l'étape de différenciation à laquelle il est associé (**Figure S2.5 et Tableau S2.5**). Par exemple, FOXP3 est impliqué dans la différenciation des cellules T CD4+ (294-296), IKZF3 joue un rôle important dans le développement des lymphocytes (288), et PAX5 est indispensable dans la différenciation et la fonction des cellules B matures (11). D'autres FT ont été associés à un type cellulaire ou sous-groupe mais leurs rôles ne sont pas encore connus dans ces populations (ZNF860, ZNF385A et MSC). Ils semblent être des candidats intéressants à étudier dans l'établissement et le maintien de la différenciation au sein de l'hématopoïèse.

Cette étude a permis d'identifier des groupes de gènes co-exprimés ainsi que des FT impliqués dans la régulation de l'expression génique dans un sous-ensemble de cellules immunitaires ainsi que dans l'établissement et le maintien de la différenciation cellulaire. Ces résultats sont en accord avec la littérature, mais il reste encore un certain nombre de gènes associés à un sous-groupe de cellules immunitaires dont la fonction n'est pas connue ou leur rôle n'a pas été étudié dans les cellules immunitaires primaires. Ainsi par *guilt by association*, ces gènes sembleraient avoir un rôle important dans certains des types cellulaires immunitaires présentés et leur étude permettrait possiblement, de mieux comprendre la fonction cellulaire de ces populations. Par exemple, les gènes DIRAS1 et RASGRF1 seraient potentiellement impliqués dans la signalisation des cellules B (234), alors que NECAB2 jouerait un rôle important dans les neutrophiles, au niveau de la régulation de l'apoptose, l'autophagie et la formation des NETs.

4.1.2 Les isoformes offrent à la protéine des fonctions spécifiques dans certaines cellules immunitaires primaires

Certains gènes humains ont la capacité d'exprimer plusieurs transcrits qui vont coder pour des isoformes distinctes de la protéine. Par exemple, une étude menée par X. Yang et al. (45) montre que sur 1 492 gènes humains étudiés, 506 d'entre eux expriment des transcrits qui codent pour au moins deux isoformes de la protéine. Ils ont regardé la différence d'isoformes encodées par ces 506 gènes dans 5 tissus humains et ont découvert que plus de 200 gènes avaient une expression plus élevée de l'isoforme alternative de la protéine par rapport à l'isoforme canonique. Cela suggère qu'une isoforme, et indirectement le transcrit qui l'encode, pourrait être spécifiquement exprimée dans un tissu ou un type cellulaire. Il serait aussi intéressant de regarder si la fonction de l'isoforme alternative de la protéine est distincte de l'isoforme canonique.

Dans le second article présenté dans cette thèse, les données de séquençage d'ARNm dans les cellules immunitaires provenant de 12 individus sains ont été analysées par transcrit pour étudier les profils d'expression des transcrits d'un même gène dans nos cellules immunitaires primaires. Pour cela nous avons appliqué notre approche dans deux sous-ensembles de populations cellulaires. Le premier sous-ensemble d'analyse est les cellules lymphoïdes (cellules B, NK, T $\gamma\delta$, T CD8+ et T CD4+) et les monocytes qui regroupent 6 985 gènes avec au moins deux transcrits exprimés dans nos données. Nous avons décidé de ne pas inclure les neutrophiles car, son transcriptome a un profil d'expression quantitativement différent des autres types cellulaires primaires (**Figure S3.1**). Le second sous-ensemble de cellules immunitaires correspond aux monocytes, macrophages et macrophages activés par LPS. Ce sous-groupe contient 5 773 gènes avec au moins deux transcrits exprimés dans nos données. Pour identifier les gènes qui possèdent des transcrits avec des profils d'expression distincts dans nos sous-ensembles de populations, dans la comparaison de deux types cellulaires nous avons : (1) défini comme « *query* » tout transcrit d'un gène avec une expression élevée (moyenne géométrique > 60 CPM) dans le premier type cellulaire (ex : cellule T CD4+) et une faible expression (moyenne géométrique < 5 CPM) dans le second type cellulaire de la comparaison (ex : cellules T CD8+), puis (2) nous identifions tous les transcrits du même gène qui ont une expression élevée dans le second type cellulaire, ici les cellules T CD8+, que nous appellerons « *alternate* ». Dans le cas où nous ne comparons pas deux types cellulaires entre eux mais deux groupes de populations, pour les transcrits de type « *query* » l'expression doit être élevée dans au moins un des types cellulaires du premier groupe (ex : si le 1^{er} groupe est les cellules lymphoïdes, l'expression doit être > 60 CPM dans au moins un de ces types cellulaires). De même pour les transcrits « *alternate* », ils doivent avoir une expression élevée dans un des types cellulaires du second groupe de populations de cellules. Les gènes qui possèdent au moins un transcrit de type « *query* » et un de type « *alternate* » sont appelés des gènes candidats. Un premier tri a été réalisé parmi ces gènes. Nous avons conservé les gènes dont les transcrits « *query* » et « *alternate* » codent pour des isoformes protéiques différentes expliqués par des SIT distincts ou l'épissage ou les deux. Par la suite, une matrice de corrélation et la valeur p associée sont calculées pour sélectionner le sous-ensemble de gènes candidats qui ont été étudiés et présentés dans l'article. Pour être sélectionné, les gènes candidats devaient avoir une valeur de corrélation et une valeur p entre un transcrit « *query* » et un transcrit « *alternate* » inférieures à 0 et 0,05, respectivement. En utilisant cette approche nous avons sélectionné 11 gènes candidats (*CD300A*, *FYB1*, *GPI*, *LITAF*, *PSMA1*, *PTMA*, *PTPRC*, *RPL32*, *SEPTIN9*,

SH3BP2 et *SH3KBP1*) dans l'analyse des cellules lymphoïdes et monocytes (**Tableau 3.1**) et cinq gènes candidats (*CD44*, *FNBP1*, *KLF6*, *OAZ2* et *SEPTIN9*) pour la lignée monocytes/macrophages (**Tableau 3.2**). Nous avons considéré que ces 15 gènes étaient des candidats idéaux pour avoir plusieurs transcrits avec des profils d'expression différents et qui codent pour des isoformes protéiques avec des fonctions potentiellement différentes selon le type cellulaire où le transcrit est exprimé.

Afin de déterminer si notre approche permet de répondre à notre hypothèse qui est qu'un même gène exprime une isoforme protéique différente selon le type cellulaire car ses isoformes ont des fonctions différentes, nous avons étudié la littérature pour savoir lesquels de nos gènes candidats sélectionnés répondaient à cette hypothèse. Le gène *PTPRC* est un très bon exemple pour confirmer notre hypothèse et ainsi valider notre approche. Ce gène code pour la protéine CD45, dont les différentes isoformes sont bien étudiées. CD45 est connue pour avoir un niveau de sensibilité différent pour chaque isoforme et ses isoformes sont spécifiquement exprimées dans certains types cellulaires immunitaires (305, 321). Dans nos données nous observons trois transcrits (*PTPRC-201*, *PTPRC-203* et *PTPRC-209*) avec des profils d'expression différents et codant pour différentes isoformes de la protéine (**Figure 3.2**). Ces transcrits diffèrent à leur extrémité 5' UTR mais aussi dans la distribution des exons 4, 5 et 6 qui codent pour la partie C-terminale de la protéine correspondant à la région extracellulaire. La variabilité de cette région a une conséquence directe sur l'affinité pour la liaison à ses ligands. Un second gène aussi bien connu valide notre approche dans la lignée monocytes/macrophages. Le gène *CD44* est connu pour exprimer des isoformes protéiques spécifiques lors de la différenciation des monocytes en macrophages et au moment de l'activation des macrophages par LPS.

L'étude des 10 autres gènes, dans l'analyse des cellules lymphoïdes et monocytes, a mis l'emphase sur les gènes *FYB1*, *PTMA* et *SH3KBP1* qui ont des transcrits avec des profils d'expression différents et qui codent pour des protéines ayant des fonctions distinctes (**Annexes B; Box S3.1**). De plus, le gène *SH3BP2* code pour une protéine interagissant avec *SH3KBP1* dans les cellules T ce qui va permettre à ce complexe, de réguler l'endocytose et le réarrangement du cytosquelette (333).

Un autre gène intéressant est *SEPTIN9* car présent dans nos deux sous-ensembles d'analyses. Il a trois isoformes protéiques qui sont prédominantes dans les lymphocytes (*SEPTIN9-202*; alias i5), les monocytes (*SEPTIN9-201*; alias i2) et dans les macrophages (*SEPTIN9-203*; alias i1), respectivement. La protéine SEPT9, encodée par *SEPTIN9*, lie les microtubules dans les macrophages par son motif MAP qui est encodé par l'exon 2. Cet exon est uniquement présent dans le transcrit *SEPTIN9-203* ce qui laisse croire que cette isoforme est importante dans les macrophages en comparaison avec les autres isoformes du gène *SEPTIN9* car en liant les microtubules, SEPT9 les stabilisent et permet ainsi l'activation des macrophages (323).

L'étude de ces gènes augmente la confiance concernant la validité de notre approche sur l'analyse des profils d'expression des transcrits d'un gène pour obtenir un niveau supplémentaire sur la compréhension de l'établissement et le maintien des cellules immunitaires ainsi que leurs fonctions spécifiques et partagées. Cependant nous savons que les paramètres utilisés pourraient

être modifier afin d'être plus stricte ou plus permissif ce qui affectera directement le nombre de gènes candidats.

4.2 Limites et perspectives des approches d'analyses de la co-expression génique, la régulation par des facteurs de transcription et des isoformes protéiques

Dans les deux études présentées, les données analysées étaient de type séquençage d'ARN. Les marqueurs de surface choisis pour identifier les types cellulaires immunitaires ont permis d'obtenir des populations de cellules hétérogènes validées par immunophénotypage (**Figure S2.1**). Ainsi, les données de séquençage d'ARN proviennent de population de cellules immunitaires hétérogènes, ce qui ajoute un niveau de bruit supplémentaire lors de l'étude des données. Aujourd'hui il existe un certain nombre de méthodes de séquençage de l'ARN dont les lectures courtes d'ADNc, qui a été utilisé dans notre étude, les lectures longues d'ADNc et les lectures longues directes d'ARN (**Figure 4.1** et **Tableau 4.1**) (334). Les méthodes de séquençage à lectures longues sont plus récentes que celle à lectures courtes. Comme nous pouvons l'observer dans la **Figure 6.1**, les modèles expérimentaux et le pré-traitement des données pour ces trois méthodes sont relativement différentes. La méthode à lectures longues directes d'ARN ne réalise aucun pré-traitement du matériel biologique alors que la lectures courtes d'ADNc fragmente l'ARN et le transcrit en ADNc. De plus, les méthodes à lectures longues de l'ADNc et direct de l'ARN diminue fortement l'ambiguïté sur les isoformes de transcrits obtenus, ce qui amène une méthode mieux adaptée pour l'analyse de profils d'expression des transcrits d'un même gène.

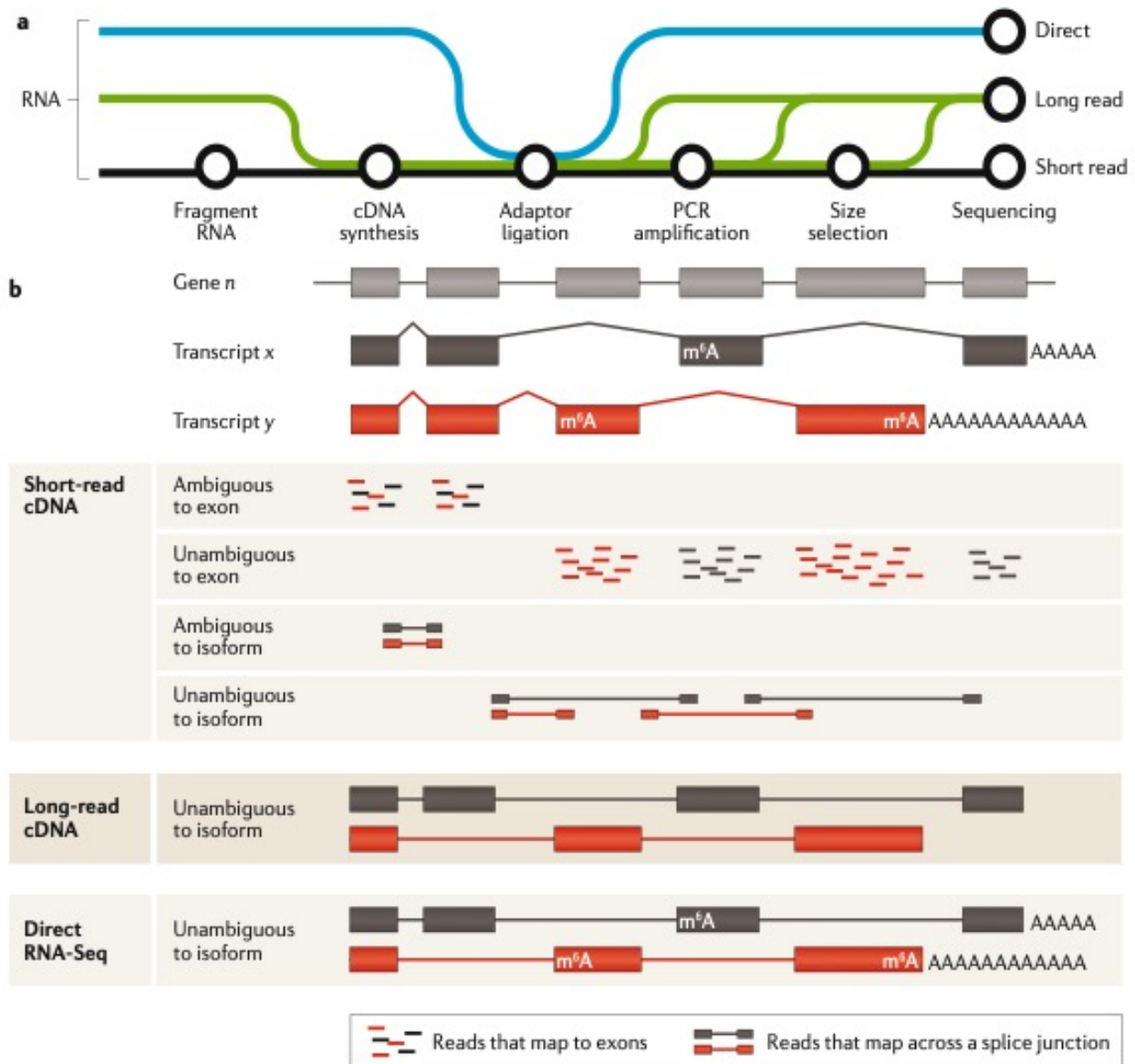


Figure 4.1. Présentation de trois méthodes de séquençage d'ARN. **a.** Schéma des méthodes de préparation des bibliothèque (appelé en anglais *library*) pour les méthodes de séquençage d'ARN de types séquençages à lectures courtes (appelé en anglais *short-read*), à lectures longues (appelé en anglais *long-read*) et à lecture longue direct. Ces types sont respectivement représentés par les lignes noire, verte et bleu. La complexité et les biais de la préparation des bibliothèques varient en fonction de l'approche spécifique utilisée. Les méthodes à lecture courte et longue d'ADNc partagent plusieurs étapes dans leurs protocoles, mais toutes les méthodes nécessitent une étape de ligature adaptée et toutes sont affectées par la qualité de l'échantillon et les problèmes de calcul en amont et en aval de la préparation des bibliothèques. **b.** Comparaison des analyses à lectures courtes, lectures longues et lecture longue direct d'ADNc. Plus de 90% des gènes humains (ex : *gene n* de la figure) sont alternativement épissés pour former deux ou plusieurs isoformes distinctes et exprimées (ex : *transcripts x* et *y* dans la figure). La complexité des informations obtenues augmente pour le

séquençage à lectures courtes de l'ADNc, où la détection des isoformes de transcrits peut être compromise par des lectures qui ne peuvent pas être cartographiées sans ambiguïté, en comparaison aux méthodes à lectures longues qui séquentent directement les isoformes de transcrits. Dans le séquençage à lectures courtes d'ADNc, une proportion importante de lectures est cartographiée de manière ambiguë lorsqu'un exon est partagé entre des isoformes de transcrits; les lectures qui couvrent les jonctions exon-exon peuvent être utilisées pour améliorer l'analyse des isoformes de transcrits, mais peuvent également être cartographiées de manière ambiguë lorsqu'une jonction est partagée entre les isoformes de transcrits. Ces problèmes compliquent l'analyse et l'interprétation des résultats. La méthode à lecture longue d'ADNc peut générer des lectures d'isoformes de transcrits de longueur complète qui suppriment ou réduisent considérablement ces artefacts et améliorent l'analyse de l'expression différentielle des isoformes de transcrits. Cependant, cette méthode repose sur la conversion de l'ADNc, qui supprime les informations sur les modifications de base de l'ARN et ne peut faire que des estimations brutes de la longueur de la queue de polyadénylation (poly(A)). Le séquençage à lecture longue direct de l'ARN permet l'analyse des isoformes de transcrits sur toute la longueur de l'ARN, la détection des modifications de base (comme la N6-méthyladénosine (m6A)) et l'estimation de la longueur de la queue poly(A). Tiré de (334).

Tableau 4.1. Comparaison des plateformes de séquençage de l'ARN à lectures courtes et lecture longues. Tiré de (334).

Sequencing technology	Platform	Advantages	Disadvantages	Key applications
Short-read cDNA	Illumina, Ion Torrent	<ul style="list-style-type: none"> • Technology features very high throughput: currently 100–1,000 times more reads per run than long-read platforms • Biases and error profiles are well understood (homopolymers are still an issue for Ion Torrent) • A huge catalogue of compatible methods and computational workflows are available • Analysis works with degraded RNA 	<ul style="list-style-type: none"> • Sample preparation includes reverse transcription, PCR and size selection adding biases to all methods • Isoform detection and quantitation can be limited • Transcript discovery methods require a de novo transcriptome alignment and/or assembly step 	Nearly all RNA-seq methods have been developed for short-read cDNA sequencing: DGE, WTA, small RNA, single-cell, spatialomics, nascent RNA, transcriptome, structural and RNA-protein interaction analysis, and more are all possible
Long-read cDNA	PacBio, ONT	<ul style="list-style-type: none"> • Long reads of 1–50 kb capture many full-length transcripts • Computational methods for de novo transcriptome analysis are simplified 	<ul style="list-style-type: none"> • Technology features low-to-medium throughput: currently only 500,000 to 10 million reads per run • Sample preparation includes reverse transcription, PCR and size selection (for some protocols), adding biases to many methods • Degraded RNA analysis is not recommended 	Sequencing is particularly suited to isoform discovery, de novo transcriptome analysis, fusion transcript discovery, and MHC, HLA or other complex transcript analysis
Long-read RNA	ONT	<ul style="list-style-type: none"> • Long reads of 1–50 kb capture many full-length transcripts • Computational methods for de novo transcriptome analysis are simplified • Sample preparation does not require reverse transcription or PCR-reducing biases • RNA base modifications can be detected • Poly(A) tail lengths can be directly estimated from single-molecule sequencing 	<ul style="list-style-type: none"> • Technology features low throughput: currently only 500,000 to 1 million reads per run • Sample preparation and sequencing biases are not well understood • Degraded RNA analysis is not recommended 	<ul style="list-style-type: none"> • Sequencing is particularly suited to isoform discovery, de novo transcriptome analysis, fusion transcript discovery, and MHC, HLA or other complex transcript analysis • Ribonucleotide modifications can be detected

Légende : Le tableau fournit un aperçu de haut niveau des avantages et des inconvénients des trois principales technologies de séquençage pour le séquençage de l'ARN. DGE, differential gene expression; HLA, human leukocyte antigen; MHC, major histocompatibility complex; ONT, Oxford Nanopore; PacBio, Pacific Biosciences; poly(A), polyadenylation; WTA, whole-transcriptome analysis.

Pour la préparation des bibliothèques (appelé en anglais *library*), il existe deux principales approches pour les trois types de séquençages présentés précédemment : (1) la méthode *oligo-dT* qui ne récupère pas les ARN courts non-codant et (2) la méthode d'analyse du transcriptome entier (appelé en anglais *whole-transcriptome analysis (WTA)*) qui elle, récupère les courts ARN non-codant (ex : ARN mitochondrial et ARN ribosomal) (334). Dans notre étude, nous avons choisi la première méthode, *oligo-dT*, car nous souhaitons uniquement étudier l'expression génique dans nos cellules immunitaires primaires. Cette méthode, contrairement à la WTA, nous permet de faire moins de nettoyage des données de séquençage, car elle détecte moins de transcrits.

Un second paramètre est important pour minimiser le nettoyage des données après séquençage. Ce paramètre est le choix de détection des duplicatas obtenus par Amplification en Chaîne par Polymérase (APC appelé en anglais PCR). Dans notre cas nous avons choisi le séquençage en paire (appelé en anglais *paired-end sequencing*) au lieu du séquençage seul (appelé en anglais *single-end sequencing*). Ce choix nous permet d'obtenir des lectures brutes qui ont pu former une paire avec un ADNc de référence et l'élimination des duplicatas qui ont un mauvais alignement (334). De plus, les ADNc qui forment des paires doivent avoir le même identifiant moléculaire unique (IMU appelé en anglais UMI) pour être validé. Un IMU est distribué à chaque ADNc obtenu lors de la construction des bibliothèques et peut être utilisé pour calculer la fréquence des allèles dans les données finales. Cette méthodologie permet de réduire les taux de variance et de faux positifs ainsi que de supprimer les artefacts.

Afin d'obtenir des résultats de meilleures qualités, le nombre de réplicas par conditions est important. Il est préconisé, d'après la littérature, d'avoir minimum 6 réplicas par condition pour identifier au mieux les variabilités biologiques entre les échantillons. Pour notre étude nous avons choisi d'avoir 1 réplica de chaque échantillon pour 12 individus dans neuf types cellulaires immunitaires. Ainsi nous avons 12 échantillons de chaque condition, provenant de 12 individus différents. Cette analyse étant produite pour définir un protocole d'analyse, nous avons fait le choix de ne pas produire une grande quantité de données. Les analyses de qualités (**Tableau S3.3**) ont montré que, malgré l'absence de réplicas pour chaque individu, nous avons obtenu une bonne profondeur de lectures brutes qui est en moyenne supérieur à 20 millions de lectures à travers nos échantillons. La profondeur de lectures requise est entre 10 et 30 millions de lectures brutes (334).

À cette étape, les données de séquençages d'ARN sont des lectures brutes contenues dans un fichier au format FASTQ. Le choix des outils pour estimer l'expression des gènes et des transcrits provenant des données de séquençage vont dépendre : (1) du type de données obtenus, identifié par le choix des paramètres précédemment présentés, (2) du type d'analyse qui va être réalisé sur l'expression par gènes ou transcrits (ex: comparaison de l'expression entre les échantillon, analyse de la distribution de l'expression à travers les échantillon ou découverte de novo transcrits) et (3) des méthodes d'analyse des outils existants. Les plus courantes voies d'analyse utilisées avec des données de séquençage sont présentées dans la **Figure 4.2**.

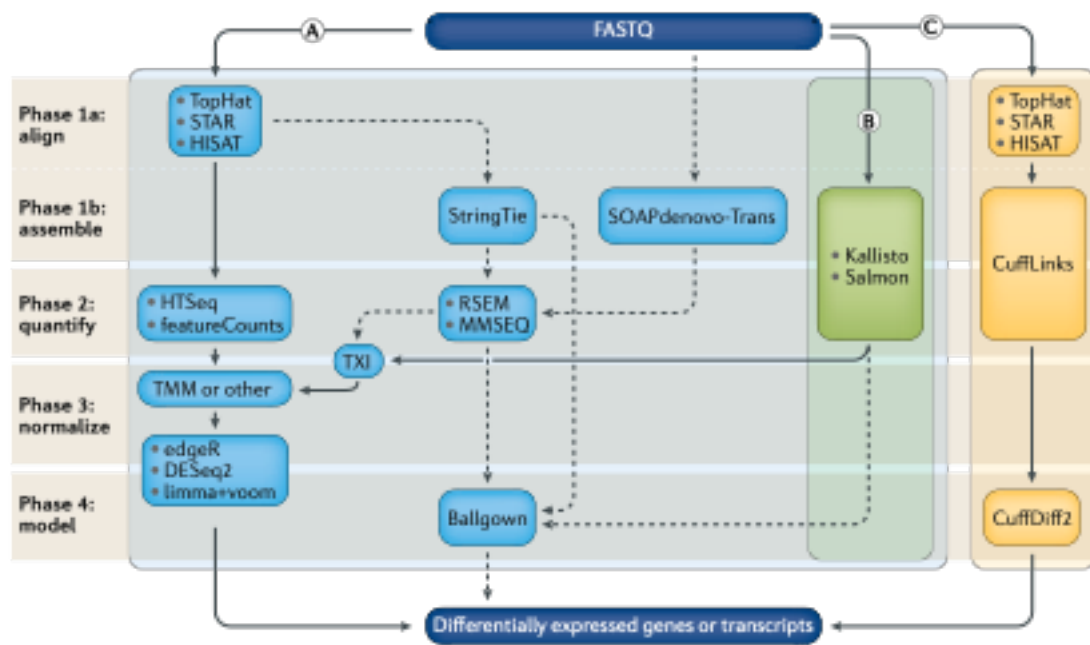


Figure 4.2 **Schématisation des étapes d'analyse de données de séquençage d'ARN pour étudier l'expression différentielle des gènes.** L'analyse computationnelle de l'expression différentielle des gènes (EDG) commence par le séquençage de l'ARN brut présenté dans un fichier au format FASTQ. Ce fichier peut suivre un certain nombre de voies d'analyse différentes. Trois voies d'analyse populaires (A, B et C, représentés par les lignes pleines) sont données à titre d'exemple, et certains des outils alternatifs les plus courants (représentés par les lignes pointillées) sont indiqués. Dans la voie d'analyse A, les outils d'alignement tels que TopHat (124), STAR (335) ou HISAT2 (336) utilisent un génome de référence pour cartographier les lectures aux emplacements génomiques, puis les outils de quantification, tels que HTSeq (337) et featureCounts (338), attribuent les lectures aux bons emplacements génomiques. Après normalisation (généralement à l'aide de méthodes intégrées dans les outils de quantification ou de modélisation de l'expression, tels que la moyenne tronquée des valeurs M (MTM appelé en anglais TMM (339)), l'expression génique est modélisée à l'aide d'outils tels que edgeR (193), DESeq2 (340) et limma + voom (303), puis une liste de variables exprimées de manière différentielle des gènes ou des transcrits est obtenue pour une visualisation et une interprétation ultérieure. Dans la voie d'analyse B, des outils plus récents et sans alignement, tels que Kallisto (341) et Salmon (129), assemblent un transcriptome et quantifient l'abondance en une seule étape. La sortie de ces outils est généralement convertie en estimations de comptage (à l'aide de tximport (TXI) (342)) et exécutée par la même normalisation et modélisation utilisée dans la voie d'analyse A, pour produire une liste de gènes ou de transcriptions exprimés de manière différentielle. Alternativement, la voie d'analyse C commence par aligner les lectures (généralement effectuées avec TopHat, bien que STAR et HISAT puissent également être utilisés), suivi de l'utilisation de CuffLinks (343) pour traiter les lectures brutes et du package CuffDiff2 pour produire des estimations d'abondance de transcrits et une liste de gènes ou transcrits exprimés de manière différentielle. D'autres outils couramment utilisés

incluent StringTie (128), qui assemble un modèle de transcriptome à partir de TopHat (ou d'outils similaires) avant que les résultats ne soient transmis à RSEM (344) ou MMSEQ (345) pour estimer l'abondance des transcrits, puis à Ballgown (346) pour identifier les gènes ou les transcrits exprimés de manière différentielle, et SOAPdenovo-trans (347), qui aligne et assemble simultanément les lectures pour analyse via une voie alternative. Tiré de (334).

Une méthode alternative ou complémentaire à l'étude de ces données serait d'étudier des données provenant de séquençage d'ARN de cellule unique (appelé en anglais *scRNA-Seq*), permettant ainsi d'étudier le transcriptome d'une unique cellule (184, 187, 348). Cela permet d'obtenir des estimations de l'expression spécifique au type cellulaire et à l'état de la cellule (ex : naïve, mature ou activée) soit son phénotype. Cette méthode permet d'identifier des nouvelles sous-populations cellulaires et d'identifier leur niveau de développement au sein du tissu. Cela offre des informations sur la dynamique des processus de différenciation cellulaire du tissu étudié. Le séquençage par ARN de cellule unique, utilise les méthodes à lectures courtes et lecture longues d'ADNc sur le transcriptome d'une unique cellule (**Figure 4.3**) (349). Ces approches, comme nous avons pu le voir précédemment, demandent un certain nombre de manipulations des données d'ARN puis d'ADNc et, contrairement au séquençage d'ARN qui utilise des échantillons avec des milliers de cellules, dans ce cas, le transcriptome d'une unique cellule est utilisé. Cela amène une plus grande sensibilité pour la dégradation, la perte ou la contamination de l'ARN. De plus l'APC doit être plus profonde que pour le séquençage d'ARN ce qui produit plus de bruits et demande un nettoyage post-production des données plus rigoureux (350).

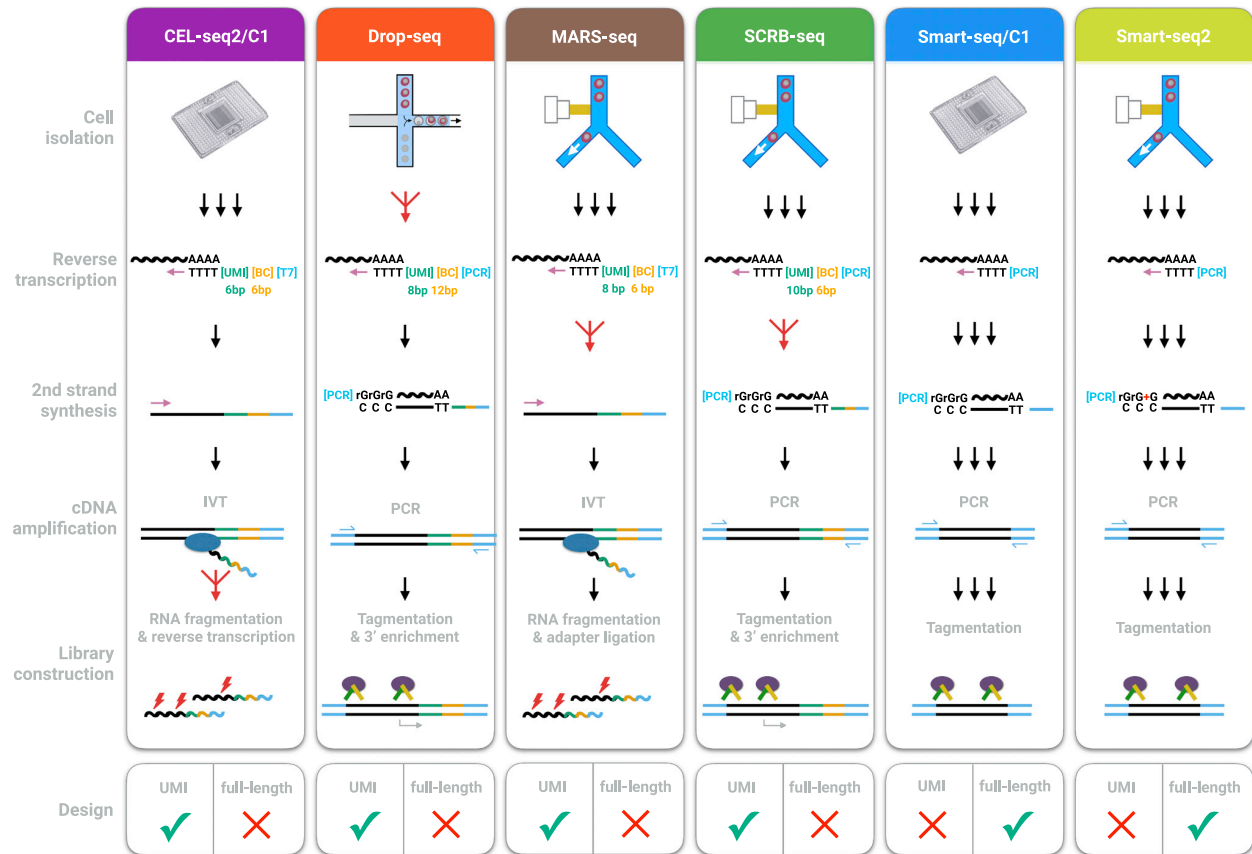


Figure 4.3 Schématisation complète des étapes de préparations des bibliothèques pour le séquençage d'ARN par cellule unique. Tiré de (349).

Dans notre premier article, nous avons analysé la régulation de l'expression génique par les FT ainsi que leur implication dans la différenciation et activation des types cellulaires de l'hématopoïèse. Pour identifier les gènes qui codent pour des facteurs de transcription, nous avons utilisé la base de données présentée dans l'article de Lambert et al. (26). Ils ont identifié au total 1 639 FT humains qu'ils ont identifié avec les deux caractéristiques suivantes : (1) être capable de lier l'ADN sur une séquence spécifique et (2) pouvoir réguler la transcription génique. Cette base de données a été obtenue en 2018 et depuis de nombreuses bases de données de FT humains ont vu le jour avec des critères de sélection différentes. Par exemple le catalogue GO (*Gene Ontology*) a identifié 1 457 FT humains qui lient l'ADN (351). Ces FT proviennent de la base de données TFcheckpoint (352), contient les FT associés à l'ARN Pol II et la base de données de FT et cofacteurs, HumanTFDB (353). Pour valider les FT présents dans ces bases de données, le catalogue GO a validé par la littérature les FT provenant de TFcheckpoint et pour HumanTFDB, ils ont utilisé une validation avec la base de données UniProt avec trois termes GO : (1) *DNA-binding TF activity*, (2) *general TF initiation factor activity* et (3) *transcription coregulator activity*.

Certaines bases de données se sont intéressées à la régulation du transcriptome par les FT humains. Parmi elles, se trouvent KnockTF (354) qui a développé une approche d'analyse par *knockdown* et *knockout* de FT provenant des bases de données NCBI GEO (355) et ENCODE

(356). Au total, ils ont étudié 308 FT dans plus de 500 types cellulaires et tissus. ChIPSummitDB (357) a développé une méthode pour analyser les données ChIP-Seq provenant de la base de données NCBI SRA (358) pour 292 FT ainsi qu'une approche d'analyse pour identifier les relations topologiques entre les liaisons et co-liaisons de protéines sur les sites de liaison de FT. hTFtarget (359) est la base de données, parmi celles présentées dans cette thèse, qui a étudié les données ChIP-Seq pour le plus grand nombre de FT, soit 659 FT humains dans 569 conditions (399 lignées cellulaires, 129 tissus ou cellules et 141 traitements). Les données ont été récoltées de NCBI GEO, NCBI SRA et ENCODE. Cette base de données permet d'identifier les gènes cibles de ces FT humains ainsi que les FT qui sont possiblement impliqués dans leur régulation, et cela à travers les 569 conditions.

La méthode de séquençage d'ARNm a permis l'amplification de lecture de minimum 100 pb. Cela a permis de détecter des longs ARN non-codants (lncRNA) dans nos données d'expression génique qui une longueur d'environ 200 nucléotides. Certains de ces lncRNA sont présents dans des modules de co-expression génique très fortement associés à un ou plusieurs types cellulaires et se retrouvent même dans les listes de TEGs et MRGs. Par exemple, dans le module 38 fortement associé aux cellules B, le lncRNA LOC102724714 appartient à la liste des MRGs. Au sein d'une cellule, les lncRNA peuvent avoir de nombreuses fonctions dont la régulation de la transcription et la traduction en protéine (**Figure 4.4**). Par exemple dans les macrophages, lncRNA-MMP2 régule la polarisation en maintenant l'activation de STAT6 (360). Chez la souris, il a été observé que le lncRNA Neat1 est impliqué dans l'activation des inflammasomes NLRP3, NLRC4 et AIM2, dans les macrophages, ainsi que l'activation de la caspase 1, de la production de cytokine et de la mort cellulaire pyroptotique (361). Les lncRNA peuvent aussi jouer un rôle dans certaines maladies dont le lncRNA-MIAT qui est impliqué dans le diabète sucré. Le lncRNA-MIAT joue un rôle dans la pathogenèse de la dysfonction microvasculaire (362).

Une des fonctions des lncRNA est d'interagir avec les micro-ARN, et ainsi, limiter ou inhiber leur action. Dans nos données, certains micro-ARN ont pu être détectés malgré leur très petite taille, moins de 30 nucléotides, et sont présents dans les listes de TEGs ou MRGs des modules de co-expression. Dans le module 22, deux des quatre micro-ARN du module appartiennent à la liste des MRGs (MIR8071-1 et MIR8071-2). Une des fonctions des micro-ARN est d'interagir avec certaines protéines et de réduire ou d'inhiber leur fonction. Par exemple, des micro-ARN (miR-17, miR-93, miR-20a, miR-20b, miR-106a, and miR-106b) ont la capacité de se lier à l'extrémité 3'UTR de BCL11B pour réprimer son expression. Le FT BCL11B est impliqué dans l'expression de molécule du CMH de classe I, MICA et MICB, qui sont des ligands du récepteur NKG2D et permettent ainsi aux cellules NK de reconnaître les cellules du soi (97). La réduction en nombre de ces récepteurs à la surface cellulaire entraîne la mort cellulaire par les cellules NK, entre autres. Il serait donc intéressant d'étudier des données RNA-Seq détectant les lncRNA et micro-ARN dans nos cellules immunitaires primaires et d'identifier leurs fonctions dans nos différentes populations cellulaires (363-365).

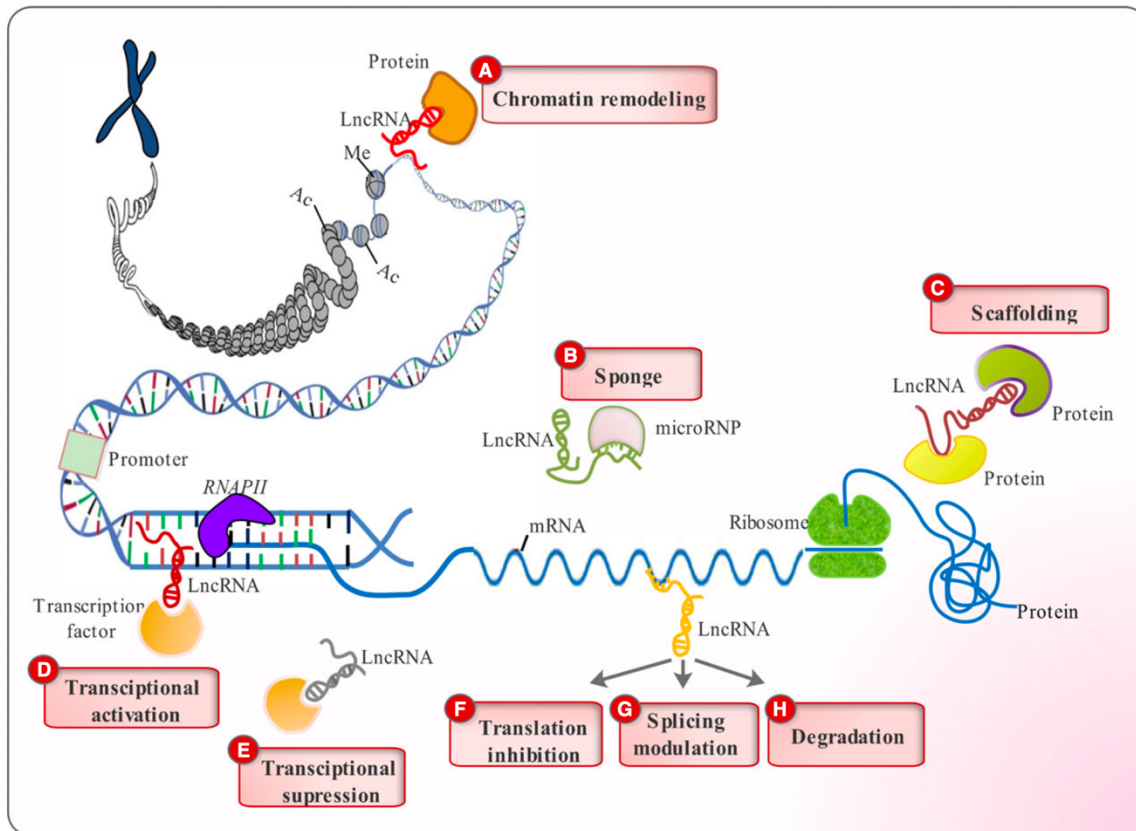


Figure 4.4 Classification des fonctions des lncRNA. A. Les lncRNA peuvent recruter différents composants protéiques du complexe de remodelage de la chromatine pour changer l'organisation de la chromatine. B. Les lncRNA peuvent agir comme des « éponges » en se liant avec un micro-ARN complémentaire pour réduire son effet. C. Les lncRNA peuvent jouer un rôle d'encrage en donnant un site d'amarrage pour les protéines qui ont une fonction dans la même voie biologique. D. Les lncRNA activent la transcription de certains gènes en guidant les FT vers leurs promoteurs cibles. E. Les lncRNA peuvent être des suppresseurs de transcription en séquestrant les FT pour les garder loin de leur promoteurs cibles. Les lncRNA peuvent moduler les ARNm avec lesquelles ils se lient et F. inhiber leur translation, G. altérer leurs modèles d'épissage et H. les amener vers les voies de dégradation d'ARNm. Tiré de (366).

Dans la seconde approche développée dans cette thèse, nous avons pu observer que certains transcrits étaient de type *nonsense mediated decay* (NMD) c'est-à-dire que ce sont des ARNm non-sens et qu'ils sont dégradés par la cellule. Ces ARNm non-sens possèdent un codon stop prématuré qui donne une protéine tronquée (367-369). Cette isoforme tronquée de la protéine peut avoir un effet délétère pour la cellule et mener à son dysfonctionnement, en étant une isoforme dominante négative par exemple, et peut mener à la mort cellulaire ou dans le pire des cas, être à l'origine d'une maladie (368, 370). La dégradation par les nucléases des ARNm non-sens est dépendante de l'erreur observée. Ces erreurs peuvent être dues à une erreur de transcription, une mutation ou une erreur d'épissage (367-369, 371).

Trois modes de surveillance de l'ARNm, exercés dans le cytoplasme lors de la traduction en protéine, peuvent déclencher la dégradation de l'ARNm et sont appelés en anglais *nonsense mediated mRNA Decay* (NMD), *Nonstop mediated mRNA Decay* (NSD) et *No-Go Decay* (NGD) (371). Ils sont respectivement provoqués par l'identification d'un codon stop prématuré (NMD), l'absence de codon stop (NSD) ou l'arrêt de la traduction d'un ARNm par un ribosome (NGD), principalement. Ces modes de surveillance possèdent des voies de dégradations spécifiques présentées dans la **Figure 4.5**. D'autres erreurs peuvent aussi être à l'origine de l'activation de ces voies de dégradation telles qu'une queue polyA anormalement longue, un codon sélénocystéine, un cadre de lecture ouvert dans la région 5' non traduite (uORF, *upstream ORF*) ou encore un décalage programmé du cadre de lecture (appelé en anglais *frameshift*).

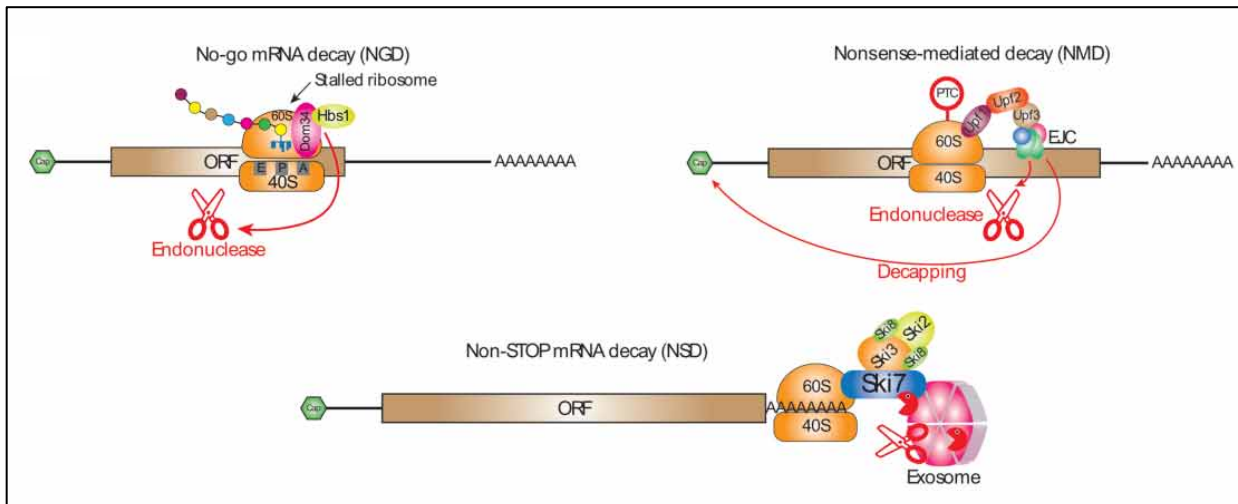


Figure 4.5 Schéma représentatif du lien entre la traduction et les trois modes de surveillance pour la dégradation de l'ARNm. Pour les trois voies connues de surveillance de l'ARNm, la traduction est nécessaire pour signaler la dégradation de l'ARNm. NGD dégrade les ARNm qui provoquent le blocage des ribosomes dans la traduction de l'ARNm. NGD nécessite une interaction entre Dom34/Hbs1 (Pelota/Hbs1 chez les mammifères) et le site A du ribosome, conduisant à une coupure endonucléolytique de l'ARNm. NSD se produit lorsque le ribosome ne rencontre jamais de codon stop et atteint à la place l'extrémité 3' d'un transcrite. NSD implique des interactions entre le ribosome et Ski7 chez *S. cerevisiae*, ou Pelota(Dom34)/Hbs1 chez les mammifères, entraînant une dégradation par les activités exonucléolytiques et endonucléolytiques de l'exosome. NMD dégrade les ARNm qui terminent la traduction au niveau d'un codon stop prématuré (PTC). NMD nécessite Upf1, qui interagit avec eRF3 (eukaryotic release factor 3) sur le ribosome, et implique également les partenaires de liaison Upf1 Upf2 et Upf3. Les PTC déclencheraient un événement de terminaison de la traduction aberrante en raison de leur distance par rapport à l'extrémité 3' du transcrite, de la composition du complexe ARNm-protéines en aval du PTC, ou des deux. Chez les mammifères, la reconnaissance d'un PTC est renforcée par la présence en aval d'un complexe de jonction d'exons (EJC). Ce complexe lie l'ARNm 20 à 24 nucléotides en amont des jonctions exon-exon et le forme le complexe ARNm-protéine avec les protéines Upf1 Upf2 et Upf3 en présence de PTC. Chez *S. cerevisiae*, les substrats de NMD sont éliminés par décapitation et dégradation 5'–3'. Chez les mammifères, le décapitation se produit également,

mais la coupure endonucléolytique par Smg6 est la forme prédominante de dégradation. Tiré de (371)

Ainsi NMD, NSD et NGD ont une fonction importante dans la régulation de l'expression génique et du maintien de l'homéostasie du transcriptome (370, 372). L'ampleur de la réponse NMD est variable selon les individus et peut être à l'origine de maladies (370). Par exemple, la mutation du gène UPF3B amène des troubles neurodéveloppementaux et des déficiences intellectuelles si la voie NMD est inefficace (370, 372). Dans certains cas, l'ARNm avec un codon stop prématuré n'est pas dégradé par la cellule ce qui est le cas pour l'isoforme de la β -globuline qui a un codon stop prématuré dans le premier exon (373-375). Cette exception laisse supposer que la proximité entre le codon AUG et le codon stop prématuré ne permet pas l'activation de la voie de dégradation de NMD. Dans le cas de la β -globuline, cette isoforme tronquée est stable dans la cellule et n'entraîne pas de maladie.

Dans ces deux approches, de nouveaux gènes et de nouveaux isoformes ont été associés à un sous-ensemble de cellules immunitaires, ce qui laisserait penser qu'ils auraient un rôle important dans la fonction cellulaire. Pour confirmer ces hypothèses, il serait intéressant d'étudier leur fonction au sein de ces cellules mais surtout de connaître : (1) les interactions protéine-protéine pour chacune de leurs isoformes, (2) la localisation cellulaire de ces protéines et leur isoformes et (3) surtout de valider que la protéine ou l'isoforme encodée est bien présente dans nos types cellulaires. Pour ce dernier point, il faudrait s'intéresser particulièrement aux marqueurs de l'épigénétique dans les cellules immunitaires. Par exemple, nous avons présenté précédemment que les lncRNA pouvaient impacter la traduction des ARN en protéine en altérant leurs modèles d'épissage, les amener vers les voies de dégradation d'ARNm ou en inhibant simplement la traduction (**Figure 4.4**). L'épigénétique elle, s'intéresse aux modifications réversibles qui ne touchent pas l'ADN et qui peuvent être transmises lors de la division cellulaire (376, 377). Ces modifications s'effectuent sur les histones et les groupes méthyles de l'ADN. La méthylation de l'ADN empêche les protéines de s'y lier et d'activer ou inactiver la transcription des gènes avoisinants. Concernant les histones il existe plusieurs modifications qui vont favoriser ou inhiber la transcription génique (la méthylation, l'acétylation, la phosphorylation, l'ubiquitination et la sumoylation). L'ensemble des données, sur les trois points présentés précédemment, permettaient d'éliminer certaines hypothèses et de mieux définir un protocole d'analyse de la fonction protéique ou de son isoforme. Ainsi, cela pourrait expliquer pourquoi, dans l'étude des isoformes de transcrits, pour les données dans les cellules lymphoïdes et monocytes, le gène candidat *RPL32* exprime des isoformes différentes selon le type cellulaire, et surtout si, ces isoformes ont un impact sur le fonctionnement de la machinerie d'épissage. Pour finir, des données complémentaires sur la régulation de l'expression génique dans nos neuf types cellulaires serviraient à valider l'hypothèse que les FT essentiels pour la fonction cellulaire sont co-exprimés avec les gènes qu'ils régulent (366, 378, 379).

Les deux approches d'analyses décrites dans les articles pourraient être utilisées dans des contextes différents d'études des cellules immunitaires avec un matériel biologique mieux adapté. Par exemple, elle pourrait provenir d'analyse de données de séquençage à lectures longues pour l'analyse de comparaison des profils d'expression des transcrits d'un gène, que ce

soit par séquençage d'ARN par échantillon ou séquençage d'ARN par cellule unique. Cela permettrait d'avoir une plus grande confiance en notre ensemble de données, et un meilleur choix des paramètres, rendraient l'analyse plus robuste (ex : choix de la méthode de séquençage et des outils d'analyses, un plus grand nombre de répliques). Elles pourraient être utilisées avec des données comparant le transcriptome de cellules saines et de cellules malades et ainsi identifier les gènes principalement touchés par la maladie étudiée, et aussi observer le gain ou la perte d'un isoforme protéique. Cette analyse pourrait être accompagnée de données épigénétiques, sur la régulation par les FT et sur les protéines (ex : isoformes, localisation cellulaire, interaction protéique). Dans d'autres cas, la comparaison peut se faire entre deux conditions expérimentales ou deux états d'une maladie telle que la progression de la maladie ou la réponse à un médicament. Dans les cas de comparaison énuméré précédemment, certains paramètres vont devoir être ajustés car l'analyse de populations cellulaires malades ou impliqués dans une expérimentation ajoute un niveau de complexité aux données. Ces comparaisons pourraient aussi être accompagnées de données épigénétiques, sur la régulation par les FT et sur les protéines (ex : isoformes, localisation cellulaire, interaction protéique). Dans les 10 dernières années, les outils de séquençage ont été peaufinés et de nouvelles méthodes de séquençage ont été développées et a été accompagnées par le développement de nouveaux outils adaptés (**Figures 4.2 et 4.3**). Aujourd'hui, la bio-informatique étant démocratisés et très bien développée, nous nous attendons à l'apparition de méthodes encore plus performantes pour nous aider à mieux comprendre le transcriptome humain.

Références

Uncategorized References

1. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics*. 2006;7:115. Epub 2006/05/18.
2. Gazit R, Garrison BS, Rao TN, Shay T, Costello J, Ericson J, et al. Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Reports*. 2013;1(3):266-80. Epub 2013/12/10.
3. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015;348(6235):660-5. Epub 2015/05/09.
4. Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science*. 2015;350(6264):1092-6. Epub 2015/10/17.
5. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, et al. Identification and characterization of essential genes in the human genome. *Science*. 2015;350(6264):1096-101. Epub 2015/10/17.
6. Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. 2015;163(6):1515-26. Epub 2015/12/03.
7. Boller S, Grosschedl R. The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. *Immunol Rev*. 2014;261(1):102-15. Epub 2014/08/16.
8. Lindner JM, Wong CS, Moller A, Nielsen PJ. A C-terminal acidic domain regulates degradation of the transcriptional coactivator Bob1. *Mol Cell Biol*. 2013;33(23):4628-40. Epub 2013/09/26.
9. Corcoran L, Emslie D, Kratina T, Shi W, Hirsch S, Taubenheim N, et al. Oct2 and Obf1 as Facilitators of B:T Cell Collaboration during a Humoral Immune Response. *Front Immunol*. 2014;5:108. Epub 2014/04/02.
10. Bordon A, Bosco N, Du Roure C, Bartholdy B, Kohler H, Matthias G, et al. Enforced expression of the transcriptional coactivator OBF1 impairs B cell differentiation at the earliest stage of development. *PLoS One*. 2008;3(12):e4007. Epub 2008/12/24.
11. Medvedovic J, Ebert A, Tagoh H, Busslinger M. Pax5: a master regulator of B cell development and leukemogenesis. *Adv Immunol*. 2011;111:179-206. Epub 2011/10/06.
12. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: the guardian of B cell identity and function. *Nat Immunol*. 2007;8(5):463-70. Epub 2007/04/19.
13. Mendez A, Mendoza L. A Network Model to Describe the Terminal Differentiation of B Cells. *PLoS Comput Biol*. 2016;12(1):e1004696. Epub 2016/01/12.
14. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, et al. MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nat Immunol*. 2013;14(10):1084-92. Epub 2013/08/27.
15. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*. 2006;7:29-59. Epub 2006/05/25.
16. Danino YM, Even D, Ideses D, Juven-Gershon T. The core promoter: At the heart of gene expression. *Biochim Biophys Acta*. 2015;1849(8):1116-31. Epub 2015/05/03.
17. Pang B, Snyder MP. Systematic identification of silencers in human cells. *Nat Genet*. 2020;52(3):254-63. Epub 2020/02/26.

18. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol.* 2015;16(3):144-54. Epub 2015/02/05.
19. Pott S, Lieb JD. What are super-enhancers? *Nat Genet.* 2015;47(1):8-12. Epub 2014/12/31.
20. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell.* 2013;153(2):307-19. Epub 2013/04/16.
21. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell.* 2013;155(4):934-47. Epub 2013/10/15.
22. Onufriev AV, Schiessel H. The nucleosome: from structure to function through physics. *Curr Opin Struct Biol.* 2019;56:119-30. Epub 2019/02/03.
23. McGinty RK, Tan S. Nucleosome structure and function. *Chem Rev.* 2015;115(6):2255-73. Epub 2014/12/17.
24. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev.* 2011;25(10):1010-22. Epub 2011/05/18.
25. Vavouri T, Lehner B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol.* 2012;13(11):R110. Epub 2012/11/29.
26. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. *Cell.* 2018;172(4):650-65. Epub 2018/02/10.
27. Zhao Y, Garcia BA. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol.* 2015;7(9):a025064. Epub 2015/09/04.
28. Stillman B. Histone Modifications: Insights into Their Influence on Gene Expression. *Cell.* 2018;175(1):6-9. Epub 2018/09/16.
29. Stryer L. *Biochemistry.* Freeman and Company^e éd. New York - Basingstoke: W. H. : Freeman and Company; 1995.
30. Eggeling R. Disentangling transcription factor binding site complexity. *Nucleic Acids Res.* 2018;46(20):e121. Epub 2018/08/08.
31. Lis M, Walther D. Erratum to: 'The orientation of transcription factor binding site motifs in gene promoter regions: does it matter?'. *BMC Genomics.* 2016;17:310. Epub 2016/04/30.
32. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet.* 2012;13(9):613-26. Epub 2012/08/08.
33. Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. *Mol Genet Genomics.* 2017;292(6):1175-95. Epub 2017/07/15.
34. Mozaffari-Jovin S, Wandersleben T, Santos KF, Will CL, Luhrmann R, Wahl MC. Novel regulatory principles of the spliceosomal Brr2 RNA helicase and links to retinal disease in humans. *RNA Biol.* 2014;11(4):298-312. Epub 2014/03/20.
35. Dredge BK, Stefani G, Engelhard CC, Darnell RB. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* 2005;24(8):1608-20. Epub 2005/06/04.
36. Sanidas I, Polytarchou C, Hatziaepostolou M, Ezell SA, Kottakis F, Hu L, et al. Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Mol Cell.* 2014;53(4):577-90. Epub 2014/01/28.
37. Rahhal R, Seto E. Emerging roles of histone modifications and HDACs in RNA splicing. *Nucleic Acids Res.* 2019;47(10):4911-26. Epub 2019/06/05.
38. Davis-Turak JC, Allison K, Shokhirev MN, Ponomarenko P, Tsimring LS, Glass CK, et al. Considering the kinetics of mRNA synthesis in the analysis of the genome and epigenome

reveals determinants of co-transcriptional splicing. *Nucleic Acids Res.* 2015;43(2):699-707. Epub 2014/12/30.

39. Dever TE, Green R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol.* 2012;4(7):a013706. Epub 2012/07/04.
40. Merrick WC, Pavitt GD. Protein Synthesis Initiation in Eukaryotic Cells. *Cold Spring Harb Perspect Biol.* 2018;10(12). Epub 2018/05/08.
41. Mace K, Giudice E, Gillet R. [Protein synthesis by the ribosome: a pathway full of pitfalls]. *Med Sci (Paris).* 2015;31(3):282-90. Epub 2015/04/10. La synthèse des protéines par le ribosome - Un chemin semé d'embûches.
42. Yael Avissar (Cell Biology) RIC, Jung Choi (Genetics) GIoT, Jean DeSaix (Evolution) UoNCaCH, Vladimir Jurukovski (Animal Physiology) SCCC, Robert Wise (Plant Biology) UoW, Oshkosh, Connie Rye (General Content Lead) EMCC. *Biology: OpenStax*; 2016.
43. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. *Gene.* 2013;514(1):1-30. Epub 2012/08/23.
44. Aguiar D, Cheng LF, Dumitrascu B, Mordelet F, Pai AA, Engelhardt BE. Bayesian nonparametric discovery of isoforms and individual specific quantification. *Nat Commun.* 2018;9(1):1681. Epub 2018/04/29.
45. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell.* 2016;164(4):805-17. Epub 2016/02/13.
46. Stastna M, Van Eyk JE. Analysis of protein isoforms: can we do it better? *Proteomics.* 2012;12(19-20):2937-48. Epub 2012/08/14.
47. Schwerk C, Schulze-Osthoff K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell.* 2005;19(1):1-13. Epub 2005/07/02.
48. Tanuma N, Nakamura K, Shima H, Kikuchi K. Protein-tyrosine phosphatase PTPepsilon C inhibits Jak-STAT signaling and differentiation induced by interleukin-6 and leukemia inhibitory factor in M1 leukemia cells. *J Biol Chem.* 2000;275(36):28216-21. Epub 2000/06/22.
49. Lapointe F, Turcotte S, Roy J, Bissonnette E, Rola-Pleszczynski M, Stankova J. RPTepsilon promotes M2-polarized macrophage migration through ROCK2 signaling and podosome formation. *J Cell Sci.* 2020;133(5). Epub 2019/11/15.
50. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol.* 2010;125(2 Suppl 2):S3-23. Epub 2010/03/05.
51. Parkin J, Cohen B. An overview of the immune system. *Lancet.* 2001;357(9270):1777-89. Epub 2001/06/14.
52. Yatim KM, Lakkis FG. A brief journey through the immune system. *Clin J Am Soc Nephrol.* 2015;10(7):1274-81. Epub 2015/04/08.
53. Cruz-Migoni S, Caamano J. Fat-Associated Lymphoid Clusters in Inflammation and Immunity. *Front Immunol.* 2016;7:612. Epub 2017/01/10.
54. Mola S, Foisy S, Boucher G, Major F, Beauchamp C, Karaky M, et al. A transcriptome-based approach to identify functional modules within and across primary human immune cells. *PLoS One.* 2020;15(5):e0233543. Epub 2020/05/30.
55. Labrecque N, Cermakian N. Circadian Clocks in the Immune System. *J Biol Rhythms.* 2015;30(4):277-90. Epub 2015/04/23.
56. Dogra P, Rancan C, Ma W, Toth M, Senda T, Carpenter DJ, et al. Tissue Determinants of Human NK Cell Development, Function, and Residence. *Cell.* 2020;180(4):749-63 e13. Epub 2020/02/16.

57. Janeway C. Immunobiology: The Immune System in Health and Disease. 5th ed.^e éd. Garland Science: ew York : Garland Pub., c2001; 2001.
58. Paul S, Shilpi, Lal G. Role of gamma-delta (gammadelta) T cells in autoimmunity. *J Leukoc Biol.* 2015;97(2):259-71. Epub 2014/12/17.
59. Peters C, Kabelitz D, Wesch D. Regulatory functions of gammadelta T cells. *Cell Mol Life Sci.* 2018;75(12):2125-35. Epub 2018/03/10.
60. He Y, Wu K, Hu Y, Sheng L, Tie R, Wang B, et al. gammadelta T cell and other immune cells crosstalk in cellular immunity. *J Immunol Res.* 2014;2014:960252. Epub 2014/04/18.
61. Hou L, Wang T, Sun J. gammadelta T cells in infection and autoimmunity. *Int Immunopharmacol.* 2015;28(2):887-91. Epub 2015/04/14.
62. Stegelmeier AA, van Vloten JP, Mould RC, Klafuric EM, Minott JA, Wootton SK, et al. Myeloid Cells during Viral Infections and Inflammation. *Viruses.* 2019;11(2). Epub 2019/02/23.
63. Commins SP, Borish L, Steinke JW. Immunologic messenger molecules: cytokines, interferons, and chemokines. *J Allergy Clin Immunol.* 2010;125(2 Suppl 2):S53-72. Epub 2009/11/26.
64. Lunney JK. Cytokines orchestrating the immune response. *Rev Sci Tech.* 1998;17(1):84-94. Epub 1998/06/25.
65. Sokol CL, Luster AD. The chemokine system in innate immunity. *Cold Spring Harb Perspect Biol.* 2015;7(5). Epub 2015/01/31.
66. Arango Duque G, Descoteaux A. Macrophage cytokines: involvement in immunity and infectious diseases. *Front Immunol.* 2014;5:491. Epub 2014/10/24.
67. Maghazachi AA, Al-Aoukaty A, Schall TJ. CC chemokines induce the generation of killer cells from CD56+ cells. *Eur J Immunol.* 1996;26(2):315-9. Epub 1996/02/01.
68. Jones J, Stevens CE, Rubinstein P, Robertazzi RR, Kerr A, Cabbad MF. Obstetric predictors of placental/umbilical cord blood volume for transplantation. *Am J Obstet Gynecol.* 2003;188(2):503-9. Epub 2003/02/20.
69. Tecchio C, Cassatella MA. Neutrophil-derived chemokines on the road to immunity. *Semin Immunol.* 2016;28(2):119-28. Epub 2016/05/07.
70. Griffith JW, Sokol CL, Luster AD. Chemokines and chemokine receptors: positioning cells for host defense and immunity. *Annu Rev Immunol.* 2014;32:659-702. Epub 2014/03/25.
71. Vignola AM, Gjomarkaj M, Arnoux B, Bousquet J. Monocytes. *J Allergy Clin Immunol.* 1998;101(2 Pt 1):149-52. Epub 1998/03/21.
72. Gann AW, Amoroso JW, Einck VJ, Rice WP, Chambers JJ, Schnarr NA. A photoinduced, benzyne click reaction. *Org Lett.* 2014;16(7):2003-5. Epub 2014/03/15.
73. Cipriani B, Borsellino G, Poccia F, Placido R, Tramonti D, Bach S, et al. Activation of C-C beta-chemokines in human peripheral blood gammadelta T cells by isopentenyl pyrophosphate and regulation by cytokines. *Blood.* 2000;95(1):39-47. Epub 1999/12/23.
74. Borish LC, Steinke JW. 2. Cytokines and chemokines. *J Allergy Clin Immunol.* 2003;111(2 Suppl):S460-75. Epub 2003/02/20.
75. Xin P, Xu X, Deng C, Liu S, Wang Y, Zhou X, et al. The role of JAK/STAT signaling pathway and its inhibitors in diseases. *Int Immunopharmacol.* 2020;80:106210. Epub 2020/01/24.
76. Morris R, Kershaw NJ, Babon JJ. The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Sci.* 2018;27(12):1984-2009. Epub 2018/09/30.
77. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD Nomenclature 2015: Human Leukocyte Differentiation Antigen Workshops as a Driving Force in Immunology. *J Immunol.* 2015;195(10):4555-63. Epub 2015/11/08.

78. Clark G, Stockinger H, Balderas R, van Zelm MC, Zola H, Hart D, et al. Nomenclature of CD molecules from the Tenth Human Leucocyte Differentiation Antigen Workshop. *Clin Transl Immunology*. 2016;5(1):e57. Epub 2016/02/24.
79. Gabius HJ, Kaltner H, Kopitz J, Andre S. The glycobiology of the CD system: a dictionary for translating marker designations into glycan/lectin structure and function. *Trends Biochem Sci*. 2015;40(7):360-76. Epub 2015/05/20.
80. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol*. 2012;12(3):191-200. Epub 2012/02/22.
81. Orru V, Steri M, Sole G, Sidore C, Virdis F, Dei M, et al. Genetic variants regulating immune cell levels in health and disease. *Cell*. 2013;155(1):242-56. Epub 2013/10/01.
82. Lakschevitz FS, Hassanpour S, Rubin A, Fine N, Sun C, Glogauer M. Identification of neutrophil surface marker changes in health and inflammation using high-throughput screening flow cytometry. *Exp Cell Res*. 2016;342(2):200-9. Epub 2016/03/13.
83. Scapini P, Marini O, Tecchio C, Cassatella MA. Human neutrophils in the saga of cellular heterogeneity: insights and open questions. *Immunol Rev*. 2016;273(1):48-60. Epub 2016/08/26.
84. Depince-Berger AE, Aanei C, Iobagiu C, Jeraiby M, Lambert C. New tools in cytometry. *Morphologie*. 2016;100(331):199-209. Epub 2016/10/30.
85. McKinnon KM. Flow Cytometry: An Overview. *Curr Protoc Immunol*. 2018;120:5 1 -5 1 11. Epub 2018/03/08.
86. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J*. 2007;48(1):11-23. Epub 2007/02/28.
87. da Silva MB, da Cunha FF, Terra FF, Camara NO. Old game, new players: Linking classical theories to new trends in transplant immunology. *World J Transplant*. 2017;7(1):1-25. Epub 2017/03/11.
88. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol*. 2016;37(11):724-37. Epub 2016/09/12.
89. Leon Machado JA, Steimle V. The MHC Class II Transactivator CIITA: Not (Quite) the Odd-One-Out Anymore among NLR Proteins. *Int J Mol Sci*. 2021;22(3). Epub 2021/01/28.
90. Jongsma MLM, Guarda G, Spaapen RM. The regulatory network behind MHC class I expression. *Mol Immunol*. 2019;113:16-21. Epub 2017/12/12.
91. Reith W, LeibundGut-Landmann S, Waldburger JM. Regulation of MHC class II gene expression by the class II transactivator. *Nat Rev Immunol*. 2005;5(10):793-806. Epub 2005/10/04.
92. Krawczyk M, Reith W. Regulation of MHC class II expression, a unique regulatory system identified by the study of a primary immunodeficiency disease. *Tissue Antigens*. 2006;67(3):183-97. Epub 2006/04/01.
93. Bahram S, Bresnahan M, Geraghty DE, Spies T. A second lineage of mammalian major histocompatibility complex class I genes. *Proc Natl Acad Sci U S A*. 1994;91(14):6259-63. Epub 1994/07/05.
94. Wu J, Groh V, Spies T. T cell antigen receptor engagement and specificity in the recognition of stress-inducible MHC class I-related chains by human epithelial gamma delta T cells. *J Immunol*. 2002;169(3):1236-40. Epub 2002/07/23.
95. Cox ST, Madrigal JA, Saudemont A. Diversity and characterization of polymorphic 5' promoter haplotypes of MICA and MICB genes. *Tissue Antigens*. 2014;84(3):293-303. Epub 2014/06/26.

96. Cox ST, Hernandez D, Danby R, Turner TR, Madrigal JA. Diversity and characterisation of polymorphic 3' untranslated region haplotypes of MICA and MICB genes. *HLA*. 2018;92(6):392-402. Epub 2018/11/25.
97. Qian M, Geng J, Luo K, Huang Z, Zhang Q, Zhang JA, et al. BCL11B regulates MICA/B-mediated immune response by acting as a competitive endogenous RNA. *Oncogene*. 2020;39(7):1514-26. Epub 2019/11/02.
98. Van Rhijn I, Moody DB. CD1 and mycobacterial lipids activate human T cells. *Immunol Rev*. 2015;264(1):138-53. Epub 2015/02/24.
99. Chancellor A, Gadola SD, Mansour S. The versatility of the CD1 lipid antigen presentation pathway. *Immunology*. 2018;154(2):196-203. Epub 2018/02/21.
100. Kaczmarek R, Pasciak M, Szymczak-Kulus K, Czerwinski M. CD1: A Singed Cat of the Three Antigen Presentation Systems. *Arch Immunol Ther Exp (Warsz)*. 2017;65(3):201-14. Epub 2017/04/08.
101. Moody DB, Cotton RN. Four pathways of CD1 antigen presentation to T cells. *Curr Opin Immunol*. 2017;46:127-33. Epub 2017/08/02.
102. Eckhardt E, Bastian M. Animal models for human group 1 CD1 protein function. *Mol Immunol*. 2021;130:159-63. Epub 2021/01/02.
103. Hauptmann G, Bahram S. Genetics of the central MHC. *Curr Opin Immunol*. 2004;16(5):668-72. Epub 2004/09/03.
104. Colten HR. Expression of the MHC class III genes. *Philos Trans R Soc Lond B Biol Sci*. 1984;306(1129):355-66. Epub 1984/09/06.
105. Ling M, Murali M. Analysis of the Complement System in the Clinical Immunology Laboratory. *Clin Lab Med*. 2019;39(4):579-90. Epub 2019/11/02.
106. Belizario JE, Neyra JM, Setubal Destro Rodrigues MF. When and how NK cell-induced programmed cell death benefits immunological protection against intracellular pathogen infection. *Innate Immun*. 2018;24(8):452-65. Epub 2018/09/22.
107. Lanier LL. NKG2D Receptor and Its Ligands in Host Defense. *Cancer Immunol Res*. 2015;3(6):575-82. Epub 2015/06/05.
108. Jiang J, Natarajan K, Margulies DH. MHC Molecules, T cell Receptors, Natural Killer Cell Receptors, and Viral Immuno-evasins-Key Elements of Adaptive and Innate Immunity. *Adv Exp Med Biol*. 2019;1172:21-62. Epub 2019/10/20.
109. Dotiwala F, Lieberman J. Granulysin: killer lymphocyte safeguard against microbes. *Curr Opin Immunol*. 2019;60:19-29. Epub 2019/05/22.
110. de Saint Basile G, Menasche G, Fischer A. Molecular mechanisms of biogenesis and exocytosis of cytotoxic granules. *Nat Rev Immunol*. 2010;10(8):568-79. Epub 2010/07/17.
111. Osinska I, Popko K, Demkow U. Perforin: an important player in immune response. *Cent Eur J Immunol*. 2014;39(1):109-15. Epub 2014/01/01.
112. Tanaka S, Baba Y. B Cell Receptor Signaling. *Adv Exp Med Biol*. 2020;1254:23-36. Epub 2020/04/24.
113. Smulski CR, Eibel H. BAFF and BAFF-Receptor in B Cell Selection and Survival. *Front Immunol*. 2018;9:2285. Epub 2018/10/24.
114. Cowland JB, Borregaard N. Granulopoiesis and granules of human neutrophils. *Immunol Rev*. 2016;273(1):11-28. Epub 2016/08/26.
115. Liew PX, Kubes P. The Neutrophil's Role During Health and Disease. *Physiol Rev*. 2019;99(2):1223-48. Epub 2019/02/14.
116. Mortaz E, Alipoor SD, Adcock IM, Mumby S, Koenderman L. Update on Neutrophil Function in Severe Inflammation. *Front Immunol*. 2018;9:2171. Epub 2018/10/26.

117. Ravindran M, Khan MA, Palaniyar N. Neutrophil Extracellular Trap Formation: Physiology, Pathology, and Pharmacology. *Biomolecules*. 2019;9(8). Epub 2019/08/17.
118. Winterbourn CC, Kettle AJ, Hampton MB. Reactive Oxygen Species and Neutrophil Function. *Annu Rev Biochem*. 2016;85:765-92. Epub 2016/04/07.
119. Gao H, Dai W, Zhao L, Min J, Wang F. The Role of Zinc and Zinc Homeostasis in Macrophage Function. *J Immunol Res*. 2018;2018:6872621. Epub 2019/01/10.
120. Maares M, Haase H. Zinc and immunity: An essential interrelation. *Arch Biochem Biophys*. 2016;611:58-65. Epub 2016/03/30.
121. Freeman SA, Grinstein S. Phagocytosis: receptors, signal integration, and the cytoskeleton. *Immunol Rev*. 2014;262(1):193-215. Epub 2014/10/17.
122. Jain N, Moeller J, Vogel V. Mechanobiology of Macrophages: How Physical Factors Coregulate Macrophage Plasticity and Phagocytosis. *Annu Rev Biomed Eng*. 2019;21:267-97. Epub 2019/06/06.
123. Schumann J. It is all about fluidity: Fatty acids and macrophage phagocytosis. *Eur J Pharmacol*. 2016;785:18-23. Epub 2015/05/20.
124. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36. Epub 2013/04/27.
125. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323. Epub 2011/08/06.
126. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012;22(10):2008-17. Epub 2012/06/23.
127. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290-5. Epub 2015/02/19.
128. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650-67. Epub 2016/08/26.
129. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-9. Epub 2017/03/07.
130. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol*. 2014;32(5):462-4. Epub 2014/04/23.
131. Kim H, Bi Y, Pal S, Gupta R, Davuluri RV. IsoformEx: isoform level gene expression estimation using weighted non-negative least squares from mRNA-Seq data. *BMC Bioinformatics*. 2011;12:305. Epub 2011/07/29.
132. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. Epub 2008/12/31.
133. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A*. 2014;111(51):E5593-601. Epub 2014/12/07.
134. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr Protoc Hum Genet*. 2015;87:11 6 1-6 4. Epub 2015/10/07.
135. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods*. 2010;7(12):1009-15. Epub 2010/11/09.

136. Wu J, Akerman M, Sun S, McCombie WR, Krainer AR, Zhang MQ. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics*. 2011;27(21):3010-6. Epub 2011/09/08.
137. Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ, et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol*. 2017;18(1):51. Epub 2017/03/17.
138. Ergun A, Doran G, Costello JC, Paik HH, Collins JJ, Mathis D, et al. Differential splicing across immune system lineages. *Proc Natl Acad Sci U S A*. 2013;110(35):14324-9. Epub 2013/08/13.
139. Schaub A, Glasmacher E. Splicing in immune cells-mechanistic insights and emerging topics. *Int Immunol*. 2017;29(4):173-81. Epub 2017/05/13.
140. Byun S, Han S, Zheng Y, Planelles V, Lee Y. The landscape of alternative splicing in HIV-1 infected CD4 T-cells. *BMC Med Genomics*. 2020;13(Suppl 5):38. Epub 2020/04/04.
141. Ullrich S, Guigo R. Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during B-cell development. *Nucleic Acids Res*. 2020;48(3):1327-40. Epub 2019/12/28.
142. West KO, Scott HM, Torres-Odio S, West AP, Patrick KL, Watson RO. The Splicing Factor hnRNP M Is a Critical Regulator of Innate Immune Gene Expression in Macrophages. *Cell Rep*. 2019;29(6):1594-609 e5. Epub 2019/11/07.
143. Rotival M, Quach H, Quintana-Murci L. Defining the genetic and evolutionary architecture of alternative splicing in response to infection. *Nat Commun*. 2019;10(1):1671. Epub 2019/04/13.
144. Liu H, Lorenzini PA, Zhang F, Xu S, Wong MSM, Zheng J, et al. Alternative splicing analysis in human monocytes and macrophages reveals MBNL1 as major regulator. *Nucleic Acids Res*. 2018;46(12):6069-86. Epub 2018/05/18.
145. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*. 2013;498(7453):236-40. Epub 2013/05/21.
146. Kalam H, Fontana MF, Kumar D. Correction: Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection. *PLoS Pathog*. 2018;14(1):e1006833. Epub 2018/01/05.
147. Sinha S, van Nimwegen E, Siggia ED. A probabilistic method to detect regulatory modules. *Bioinformatics*. 2003;19 Suppl 1:i292-301. Epub 2003/07/12.
148. Liu H, Liu M, You H, Li X, Li X. Oncogenic Network and Hub Genes for Natural Killer/T-Cell Lymphoma Utilizing WGCNA. *Front Oncol*. 2020;10:223. Epub 2020/03/21.
149. Zhai X, Xue Q, Liu Q, Guo Y, Chen Z. Colon cancer recurrence-associated genes revealed by WGCNA coexpression network analysis. *Mol Med Rep*. 2017;16(5):6499-505. Epub 2017/09/14.
150. Chen S, Yang D, Lei C, Li Y, Sun X, Chen M, et al. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *PeerJ*. 2019;7:e7873. Epub 2019/10/15.
151. Barabási AL, Ravasza, E., Vicsek, T. Deterministic scale-free networks. *Elsevier*. 2001;299(3-4):559-64.
152. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. *Nature*. 2000;406(6794):378-82. Epub 2000/08/10.
153. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*. 2009;10:327. Epub 2009/07/22.

154. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol.* 2007;1:54. Epub 2007/11/23.
155. Shay T, Jovic V, Zuk O, Rothamel K, Puyraimond-Zemmour D, Feng T, et al. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci U S A.* 2013;110(8):2946-51. Epub 2013/02/06.
156. Heng TS, Painter MW, Immunological Genome Project C. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol.* 2008;9(10):1091-4. Epub 2008/09/19.
157. Immunological Genome P. ImmGen at 15. *Nat Immunol.* 2020;21(7):700-3. Epub 2020/06/25.
158. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature.* 2007;447(7146):799-816. Epub 2007/06/16.
159. Consortium EP. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 2011;9(4):e1001046. Epub 2011/04/29.
160. Consortium EP, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature.* 2020;583(7818):699-710. Epub 2020/07/31.
161. Consortium EP, Snyder MP, Gingeras TR, Moore JE, Weng Z, Gerstein MB, et al. Perspectives on ENCODE. *Nature.* 2020;583(7818):693-8. Epub 2020/07/31.
162. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57-74. Epub 2012/09/08.
163. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. *Nature.* 2012;489(7414):91-100. Epub 2012/09/08.
164. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature.* 2020;583(7818):711-9. Epub 2020/07/31.
165. Quinones-Valdez G, Tran SS, Jun HI, Bahn JH, Yang EW, Zhan L, et al. Regulation of RNA editing by RNA-binding proteins in human cells. *Commun Biol.* 2019;2:19. Epub 2019/01/18.
166. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet.* 2009;10(10):669-80. Epub 2009/09/09.
167. Rye M, Saetrom P, Handstad T, Drablos F. Clustered ChIP-Seq-defined transcription factor binding sites and histone modifications map distinct classes of regulatory elements. *BMC Biol.* 2011;9:80. Epub 2011/11/26.
168. Wang J, Zhuang J, Iyer S, Lin XY, Greven MC, Kim BH, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res.* 2013;41(Database issue):D171-6. Epub 2012/12/04.
169. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 2012;22(9):1798-812. Epub 2012/09/08.
170. Clark DN, Read RD, Mayhew V, Petersen SC, Argueta LB, Stutz LA, et al. Four Promoters of IRF5 Respond Distinctly to Stimuli and are Affected by Autoimmune-Risk Polymorphisms. *Front Immunol.* 2013;4:360. Epub 2013/11/14.

171. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference mapping of human transcription factor footprints. *Nature*. 2020;583(7818):729-36. Epub 2020/07/31.
172. Xiong T, Turner JE. Innate lymphoid cells in autoimmunity and chronic inflammatory diseases. *Semin Immunopathol*. 2018;40(4):393-406. Epub 2018/03/24.
173. Pockley AG, Henderson B. Extracellular cell stress (heat shock) proteins-immune responses and disease: an overview. *Philos Trans R Soc Lond B Biol Sci*. 2018;373(1738). Epub 2017/12/06.
174. Ribatti D. The concept of immune surveillance against tumors. The first theories. *Oncotarget*. 2017;8(4):7175-80. Epub 2016/10/21.
175. Conway EM, Pikor LA, Kung SH, Hamilton MJ, Lam S, Lam WL, et al. Macrophages, Inflammation, and Lung Cancer. *Am J Respir Crit Care Med*. 2016;193(2):116-30. Epub 2015/11/20.
176. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. *Nat Immunol*. 2017;18(7):725-32. Epub 2017/06/21.
177. Abbas AK, Lichtman AH, Pillai S. *Cellular and Molecular Immunology*. edition t, rédacteur2017 11th May 2017.
178. Murphy KM, Weaver C. *Janeway's Immunobiology*. edition t, rédacteur2016.
179. Underhill DM, Gordon S, Imhof BA, Nunez G, Bousso P. Elie Metchnikoff (1845-1916): celebrating 100 years of cellular immunology and beyond. *Nat Rev Immunol*. 2016;16(10):651-6. Epub 2016/08/02.
180. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999;286(5439):531-7. Epub 1999/10/16.
181. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med*. 2003;197(6):711-23. Epub 2003/03/19.
182. Lowes MA, Suarez-Farinas M, Krueger JG. Immunology of psoriasis. *Annu Rev Immunol*. 2014;32:227-55. Epub 2014/03/25.
183. Kim CC, Lanier LL. Beyond the transcriptome: completion of act one of the Immunological Genome Project. *Curr Opin Immunol*. 2013;25(5):593-7. Epub 2013/10/31.
184. Proserpio V, Mahata B. Single-cell technologies to study the immune system. *Immunology*. 2016;147(2):133-40. Epub 2015/11/10.
185. Shay T, Kang J. Immunological Genome Project and systems immunology. *Trends Immunol*. 2013;34(12):602-9. Epub 2013/05/02.
186. Bjorklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nat Immunol*. 2016;17(4):451-60. Epub 2016/02/16.
187. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. 2017;356(6335). Epub 2017/04/22.
188. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*. 2013;14:632. Epub 2013/09/24.
189. van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 2018;19(4):575-92. Epub 2017/01/13.

190. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20. Epub 2014/04/04.
191. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9. Epub 2009/06/10.
192. Delhomme N, Padiou I, Furlong EE, Steinmetz LM. easyRNASeq: a bioconductor package for processing RNA-Seq data. *Bioinformatics*. 2012;28(19):2532-3. Epub 2012/08/01.
193. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40. Epub 2009/11/17.
194. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40(10):4288-97. Epub 2012/01/31.
195. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. Epub 2015/01/22.
196. Jolliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. *Stat Methods Med Res*. 1992;1(1):69-95. Epub 1992/01/01.
197. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374(2065):20150202. Epub 2016/03/10.
198. Sainani KL. Introduction to principal components analysis. *PM R*. 2014;6(3):275-8. Epub 2014/02/26.
199. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17. Epub 2006/05/02.
200. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*. 2008;24(5):719-20. Epub 2007/11/21.
201. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. Epub 2009/01/10.
202. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. The Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2005;33(Database issue):D154-9. Epub 2004/12/21.
203. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*. 2015;43(Database issue):D213-21. Epub 2014/11/28.
204. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015;43(Database issue):D1049-56. Epub 2014/11/28.
205. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457-62. Epub 2015/10/18.
206. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. *Database (Oxford)*. 2010;2010:baq020. Epub 2010/08/07.
207. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036. Epub 2011/01/20.
208. Anders ME, Evans DP. Comparison of PubMed and Google Scholar literature searches. *Respir Care*. 2010;55(5):578-83. Epub 2010/04/28.

209. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* 2006;34(Database issue):D108-10. Epub 2005/12/31.
210. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2016;44(D1):D110-5. Epub 2015/11/05.
211. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. *Cell.* 2013;152(1-2):327-39. Epub 2013/01/22.
212. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature.* 2015;527(7578):384-8. Epub 2015/11/10.
213. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science.* 2017;356(6337). Epub 2017/05/06.
214. Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 2015;43(Database issue):D117-22. Epub 2014/11/08.
215. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell.* 2014;158(6):1431-43. Epub 2014/09/13.
216. Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, et al. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol.* 2009;10(3):R29. Epub 2009/03/17.
217. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet.* 2009;10(4):252-63. Epub 2009/03/11.
218. Wingender E, Schoeps T, Haubrock M, Donitz J. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* 2015;43(Database issue):D97-102. Epub 2014/11/02.
219. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford).* 2011;2011:bar030. Epub 2011/07/26.
220. Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, et al. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol.* 2005;23(1):137-44. Epub 2005/01/08.
221. Veerla S, Ringner M, Hoglund M. Genome-wide transcription factor binding site/promoter databases for the analysis of gene sets and co-occurrence of transcription factor binding motifs. *BMC Genomics.* 2010;11:145. Epub 2010/03/03.
222. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform.* 2013;14(2):144-61. Epub 2012/08/22.
223. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carre C, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* 2019;26(6):1627-40 e7. Epub 2019/02/07.
224. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A.* 2003;100(2):605-10. Epub 2003/01/17.

225. Dylke J, Lopes J, Dang-Lawson M, Machtaler S, Matsuuchi L. Role of the extracellular and transmembrane domain of Ig-alpha/beta in assembly of the B cell antigen receptor (BCR). *Immunol Lett.* 2007;112(1):47-57. Epub 2007/08/07.
226. Nutt SL, Kee BL. The transcriptional regulation of B cell lineage commitment. *Immunity.* 2007;26(6):715-25. Epub 2007/06/22.
227. Zandi S, Mansson R, Tsapogas P, Zetterblad J, Bryder D, Sigvardsson M. EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors. *J Immunol.* 2008;181(5):3364-72. Epub 2008/08/21.
228. Blackburn TE, Santiago T, Burrows PD. FCRLA-A Resident Endoplasmic Reticulum Protein that Associates with Multiple Immunoglobulin Isotypes in B Lineage Cells. *Curr Top Microbiol Immunol.* 2017;408:47-65. Epub 2017/09/08.
229. Muller J, Nitschke L. The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease. *Nat Rev Rheumatol.* 2014;10(7):422-8. Epub 2014/04/26.
230. Rinaldi A, Chiaravalli AM, Mian M, Zucca E, Tibiletti MG, Capella C, et al. Serotonin receptor 3A expression in normal and neoplastic B cells. *Pathobiology.* 2010;77(3):129-35. Epub 2010/06/03.
231. Li S, Liu J, Min Q, Ikawa T, Yasuda S, Yang Y, et al. Kelch-like protein 14 promotes B-1a but suppresses B-1b cell development. *Int Immunol.* 2018;30(7):311-8. Epub 2018/06/26.
232. Hagiwara S. Transgenic expression of VpreB-3 under the control of the immunoglobulin heavy chain enhancer and SV40 promoter. *Kobe J Med Sci.* 1996;42(1):43-59. Epub 1996/02/01.
233. Ebeling O, Duczmal A, Aigner S, Geiger C, Schollhammer S, Kemshead JT, et al. L1 adhesion molecule on human lymphocytes and monocytes: expression and involvement in binding to alpha v beta 3 integrin. *Eur J Immunol.* 1996;26(10):2508-16. Epub 1996/10/01.
234. Tabrizi SJ, Niuro H, Masui M, Yoshimoto G, Iino T, Kikushige Y, et al. T cell leukemia/lymphoma 1 and galectin-1 regulate survival/cell death pathways in human naive and IgM+ memory B cells through altering balances in Bcl-2 family proteins. *J Immunol.* 2009;182(3):1490-9. Epub 2009/01/22.
235. Imamura Y, Oda A, Katahira T, Bundo K, Pike KA, Ratcliffe MJ, et al. BLNK binds active H-Ras to promote B cell receptor-mediated capping and ERK activation. *J Biol Chem.* 2009;284(15):9804-13. Epub 2009/02/17.
236. Kabak S, Skaggs BJ, Gold MR, Affolter M, West KL, Foster MS, et al. The direct recruitment of BLNK to immunoglobulin alpha couples the B-cell antigen receptor to distal signaling pathways. *Mol Cell Biol.* 2002;22(8):2524-35. Epub 2002/03/23.
237. Klein J, Sato A. The HLA system. Second of two parts. *N Engl J Med.* 2000;343(11):782-6. Epub 2000/09/14.
238. Fortin JS, Cloutier M, Thibodeau J. Exposing the Specific Roles of the Invariant Chain Isoforms in Shaping the MHC Class II Peptidome. *Front Immunol.* 2013;4:443. Epub 2014/01/01.
239. Nguyen TA, Smith BRC, Tate MD, Belz GT, Barrios MH, Elgass KD, et al. SIDT2 Transports Extracellular dsRNA into the Cytoplasm for Innate Immune Recognition. *Immunity.* 2017;47(3):498-509 e6. Epub 2017/09/17.
240. Huber C, Martensson A, Bokoch GM, Nemazee D, Gavin AL. FGD2, a CDC42-specific exchange factor expressed by antigen-presenting cells, localizes to early endosomes and active membrane ruffles. *J Biol Chem.* 2008;283(49):34002-12. Epub 2008/10/08.
241. English WR, Puente XS, Freije JM, Knauper V, Amour A, Merryweather A, et al. Membrane type 4 matrix metalloproteinase (MMP17) has tumor necrosis factor-alpha convertase

- activity but does not activate pro-MMP2. *J Biol Chem.* 2000;275(19):14046-55. Epub 2000/05/09.
242. Medley QG, Kedersha N, O'Brien S, Tian Q, Schlossman SF, Streuli M, et al. Characterization of GMP-17, a granule membrane protein that moves to the plasma membrane of natural killer cells following target cell recognition. *Proc Natl Acad Sci U S A.* 1996;93(2):685-9. Epub 1996/01/23.
243. Chang GW, Hsiao CC, Peng YM, Vieira Braga FA, Kragten NA, Remmerswaal EB, et al. The Adhesion G Protein-Coupled Receptor GPR56/ADGRG1 Is an Inhibitory Receptor on Human NK Cells. *Cell Rep.* 2016;15(8):1757-70. Epub 2016/05/18.
244. Pazina T, Shemesh A, Brusilovsky M, Porgador A, Campbell KS. Regulation of the Functions of Natural Cytotoxicity Receptors by Interactions with Diverse Ligands and Alterations in Splice Variant Expression. *Front Immunol.* 2017;8:369. Epub 2017/04/21.
245. Perez-Quintero LA, Roncagalli R, Guo H, Latour S, Davidson D, Veillette A. EAT-2, a SAP-like adaptor, controls NK cell activation through phospholipase Cgamma, Ca⁺⁺, and Erk, leading to granule polarization. *J Exp Med.* 2014;211(4):727-42. Epub 2014/04/02.
246. Le Bouteiller P, Tabiasco J, Polgar B, Kozma N, Giustiniani J, Siewiera J, et al. CD160: a unique activating NK cell receptor. *Immunol Lett.* 2011;138(2):93-6. Epub 2011/02/18.
247. Cai G, Freeman GJ. The CD160, BTLA, LIGHT/HVEM pathway: a bidirectional switch regulating T-cell activation. *Immunol Rev.* 2009;229(1):244-58. Epub 2009/05/12.
248. Maher K, Konjar S, Watts C, Turk B, Kopitar-Jerala N. Cystatin F regulates proteinase activity in IL-2-activated natural killer cells. *Protein Pept Lett.* 2014;21(9):957-65. Epub 2014/04/08.
249. Stoeckle C, Gouttefangeas C, Hammer M, Weber E, Melms A, Tolosa E. Cathepsin W expressed exclusively in CD8⁺ T cells and NK cells, is secreted during target cell killing but is not essential for cytotoxicity in human CTLs. *Exp Hematol.* 2009;37(2):266-75. Epub 2008/12/23.
250. Yamazaki C, Miyamoto R, Hoshino K, Fukuda Y, Sasaki I, Saito M, et al. Conservation of a chemokine system, XCR1 and its ligand, XCL1, between human and mice. *Biochem Biophys Res Commun.* 2010;397(4):756-61. Epub 2010/06/15.
251. Lu D, Liu L, Ji X, Gao Y, Chen X, Liu Y, et al. The phosphatase DUSP2 controls the activity of the transcription activator STAT3 and regulates TH17 differentiation. *Nat Immunol.* 2015;16(12):1263-73. Epub 2015/10/20.
252. Zhu S, Phatarpekar PV, Denman CJ, Senyukov VV, Somanchi SS, Nguyen-Jackson HT, et al. Transcription of the activating receptor NKG2D in natural killer cells is regulated by STAT3 tyrosine phosphorylation. *Blood.* 2014;124(3):403-11. Epub 2014/06/04.
253. Lotem J, Levanon D, Negreanu V, Leshkowitz D, Friedlander G, Groner Y. Runx3-mediated transcriptional program in cytotoxic lymphocytes. *PLoS One.* 2013;8(11):e80467. Epub 2013/11/16.
254. Ohno S, Sato T, Kohu K, Takeda K, Okumura K, Satake M, et al. Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation. *Int Immunol.* 2008;20(1):71-9. Epub 2007/11/16.
255. Picozza M, Battistini L, Borsellino G. Mononuclear phagocytes and marker modulation: when CD16 disappears, CD38 takes the stage. *Blood.* 2013;122(3):456-7. Epub 2013/07/23.
256. Sierra-Filardi E, Nieto C, Dominguez-Soto A, Barroso R, Sanchez-Mateos P, Puig-Kroger A, et al. CCL2 shapes macrophage polarization by GM-CSF and M-CSF: identification of CCL2/CCR2-dependent gene expression profile. *J Immunol.* 2014;192(8):3858-67. Epub 2014/03/19.

257. Verreck FA, de Boer T, Langenberg DM, van der Zanden L, Ottenhoff TH. Phenotypic and functional profiling of human proinflammatory type-1 and anti-inflammatory type-2 macrophages in response to microbial antigens and IFN-gamma- and CD40L-mediated costimulation. *J Leukoc Biol.* 2006;79(2):285-93. Epub 2005/12/07.
258. Shinagawa S, Okazaki T, Ikeda M, Yudoh K, Kisanuki YY, Yanagisawa M, et al. T cells upon activation promote endothelin 1 production in monocytes via IFN-gamma and TNF-alpha. *Sci Rep.* 2017;7(1):14500. Epub 2017/11/05.
259. Subramanian Vignesh K, Deepe GS, Jr. Metallothioneins: Emerging Modulators in Immunity and Infection. *Int J Mol Sci.* 2017;18(10). Epub 2017/10/27.
260. Huang Y, Krein PM, Muruve DA, Winston BW. Complement factor B gene regulation: synergistic effects of TNF-alpha and IFN-gamma in macrophages. *J Immunol.* 2002;169(5):2627-35. Epub 2002/08/24.
261. Metz R, Duhadaway JB, Kamasani U, Laury-Kleintop L, Muller AJ, Prendergast GC. Novel tryptophan catabolic enzyme IDO2 is the preferred biochemical target of the antitumor indoleamine 2,3-dioxygenase inhibitory compound D-1-methyl-tryptophan. *Cancer Res.* 2007;67(15):7082-7. Epub 2007/08/03.
262. Kawata K, Kubota S, Eguchi T, Aoyama E, Moritani NH, Oka M, et al. A Tumor Suppressor Gene Product, Platelet-Derived Growth Factor Receptor-Like Protein Controls Chondrocyte Proliferation and Differentiation. *J Cell Biochem.* 2017;118(11):4033-44. Epub 2017/04/14.
263. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A. Neutrophil function: from mechanisms to disease. *Annu Rev Immunol.* 2012;30:459-89. Epub 2012/01/10.
264. Jaillon S, Peri G, Delneste Y, Fremaux I, Doni A, Moalli F, et al. The humoral pattern recognition receptor PTX3 is stored in neutrophil granules and localizes in extracellular traps. *J Exp Med.* 2007;204(4):793-804. Epub 2007/03/29.
265. Vorobjeva NV, Pinegin BV. Neutrophil extracellular traps: mechanisms of formation and role in health and disease. *Biochemistry (Mosc).* 2014;79(12):1286-96. Epub 2015/02/27.
266. Yu Y, Wang J, Khaled W, Burke S, Li P, Chen X, et al. Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med.* 2012;209(13):2467-83. Epub 2012/12/12.
267. Liu P, Keller JR, Ortiz M, Tessarollo L, Rachel RA, Nakamura T, et al. Bcl11a is essential for normal lymphoid development. *Nat Immunol.* 2003;4(6):525-32. Epub 2003/04/30.
268. Lee BS, Lee BK, Iyer VR, Sleckman BP, Shaffer AL, 3rd, Ippolito GC, et al. Corrected and Republished from: BCL11A Is a Critical Component of a Transcriptional Network That Activates RAG Expression and V(D)J Recombination. *Mol Cell Biol.* 2018;38(1). Epub 2017/10/19.
269. Garvie CW, Stagno JR, Reid S, Singh A, Harrington E, Boss JM. Characterization of the RFX complex and the RFX5(L66A) mutant: implications for the regulation of MHC class II gene expression. *Biochemistry.* 2007;46(6):1597-611. Epub 2007/02/07.
270. Lochamy J, Rogers EM, Boss JM. CREB and phospho-CREB interact with RFX5 and CIITA to regulate MHC class II genes. *Mol Immunol.* 2007;44(5):837-47. Epub 2006/05/30.
271. Thakker S, Purushothaman P, Gupta N, Challa S, Cai Q, Verma SC. Kaposi's Sarcoma-Associated Herpesvirus Latency-Associated Nuclear Antigen Inhibits Major Histocompatibility Complex Class II Expression by Disrupting Enhanceosome Assembly through Binding with the Regulatory Factor X Complex. *J Virol.* 2015;89(10):5536-56. Epub 2015/03/06.
272. Boyd NH, Morgan JE, Greer SF. Polycomb recruitment at the Class II transactivator gene. *Mol Immunol.* 2015;67(2 Pt B):482-91. Epub 2015/08/19.

273. Mehta NT, Truax AD, Boyd NH, Greer SF. Early epigenetic events regulate the adaptive immune response gene CIITA. *Epigenetics*. 2011;6(4):516-25. Epub 2011/01/27.
274. Guo M, Price MJ, Patterson DG, Barwick BG, Haines RR, Kania AK, et al. EZH2 Represses the B Cell Transcriptional Program and Regulates Antibody-Secreting Cell Metabolism and Antibody Production. *J Immunol*. 2018;200(3):1039-52. Epub 2017/12/31.
275. Hashwah H, Schmid CA, Kasser S, Bertram K, Stelling A, Manz MG, et al. Inactivation of CREBBP expands the germinal center B cell compartment, down-regulates MHCII expression and promotes DLBCL growth. *Proc Natl Acad Sci U S A*. 2017;114(36):9701-6. Epub 2017/08/24.
276. Lin JH, Lin JY, Chou YC, Chen MR, Yeh TH, Lin CW, et al. Epstein-Barr virus LMP2A suppresses MHC class II expression by regulating the B-cell transcription factors E47 and PU.1. *Blood*. 2015;125(14):2228-38. Epub 2015/01/30.
277. Fujita N, Jaye DL, Geigerman C, Akyildiz A, Mooney MR, Boss JM, et al. MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell*. 2004;119(1):75-86. Epub 2004/09/30.
278. Peng SL, Gerth AJ, Ranger AM, Glimcher LH. NFATc1 and NFATc2 together control both T and B cell activation and differentiation. *Immunity*. 2001;14(1):13-20. Epub 2001/02/13.
279. Tamura A, Hirai H, Yokota A, Kamio N, Sato A, Shoji T, et al. C/EBPbeta is required for survival of Ly6C(-) monocytes. *Blood*. 2017;130(16):1809-18. Epub 2017/08/16.
280. Murn J, Mlinaric-Rascan I, Vaigot P, Alibert O, Frouin V, Gidrol X. A Myc-regulated transcriptional network controls B-cell fate in response to BCR triggering. *BMC Genomics*. 2009;10:323. Epub 2009/07/18.
281. Habib T, Park H, Tsang M, de Alboran IM, Nicks A, Wilson L, et al. Myc stimulates B lymphocyte differentiation and amplifies calcium signaling. *J Cell Biol*. 2007;179(4):717-31. Epub 2007/11/14.
282. Rudolf R, Busch R, Patra AK, Muhammad K, Avots A, Andrau JC, et al. Architecture and expression of the nfatc1 gene in lymphocytes. *Front Immunol*. 2014;5:21. Epub 2014/02/20.
283. Muhammad K, Rudolf R, Pham DAT, Klein-Hessling S, Takata K, Matsushita N, et al. Induction of Short NFATc1/alphaA Isoform Interferes with Peripheral B Cell Differentiation. *Front Immunol*. 2018;9:32. Epub 2018/02/09.
284. Samitas K, Lotvall J, Bossios A. B cells: from early development to regulating allergic diseases. *Arch Immunol Ther Exp (Warsz)*. 2010;58(3):209-25. Epub 2010/05/12.
285. Miyazaki K, Miyazaki M, Murre C. The establishment of B versus T cell identity. *Trends Immunol*. 2014;35(5):205-10. Epub 2014/04/01.
286. Somasundaram R, Prasad MA, Ungerback J, Sigvardsson M. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood*. 2015;126(2):144-52. Epub 2015/05/21.
287. Hu Y, Yoshida T, Georgopoulos K. Transcriptional circuits in B cell transformation. *Curr Opin Hematol*. 2017;24(4):345-52. Epub 2017/05/04.
288. Schmitt C, Tonnelle C, Dalloul A, Chabannon C, Debre P, Rebollo A. Aiolos and Ikaros: regulators of lymphocyte development, homeostasis and lymphoproliferation. *Apoptosis*. 2002;7(3):277-84. Epub 2002/05/09.
289. Itoh-Nakadai A, Matsumoto M, Kato H, Sasaki J, Uehara Y, Sato Y, et al. A Bach2-Cebp Gene Regulatory Network for the Commitment of Multipotent Hematopoietic Progenitors. *Cell Rep*. 2017;18(10):2401-14. Epub 2017/03/09.
290. Ji Y, Lee HJ, Goodman C, Uskokovic M, Liby K, Sporn M, et al. The synthetic triterpenoid CDDO-imidazole induces monocytic differentiation by activating the Smad and

- ERK signaling pathways in HL60 leukemia cells. *Mol Cancer Ther.* 2006;5(6):1452-8. Epub 2006/07/05.
291. Wang C, Petriello MC, Zhu B, Hennig B. PCB 126 induces monocyte/macrophage polarization and inflammation through AhR and NF-kappaB pathways. *Toxicol Appl Pharmacol.* 2019;367:71-81. Epub 2019/02/16.
292. Hume DA. The Many Alternative Faces of Macrophage Activation. *Front Immunol.* 2015;6:370. Epub 2015/08/11.
293. Pierson E, Consortium GT, Koller D, Battle A, Mostafavi S, Ardlie KG, et al. Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *PLoS Comput Biol.* 2015;11(5):e1004220. Epub 2015/05/15.
294. Zhou L, Chong MM, Littman DR. Plasticity of CD4+ T cell lineage differentiation. *Immunity.* 2009;30(5):646-55. Epub 2009/05/26.
295. Kyaw T, Toh BH, Bobik A. Foxp3+CD4+ Regulatory T-Cell Subtypes and Atherosclerosis. *Circ Res.* 2016;119(11):1151-3. Epub 2017/01/05.
296. Kraj P, Ignatowicz L. The mechanisms shaping the repertoire of CD4(+) Foxp3(+) regulatory T cells. *Immunology.* 2018;153(3):290-6. Epub 2017/11/07.
297. de Klerk E, t Hoen PA. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet.* 2015;31(3):128-39. Epub 2015/02/05.
298. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev.* 1999;63(2):405-45. Epub 1999/06/05.
299. Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* 2018;46(2):582-92. Epub 2017/12/05.
300. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A.* 2003;100(1):189-92. Epub 2002/12/28.
301. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357-60. Epub 2015/03/10.
302. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80. Epub 2004/10/06.
303. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15(2):R29. Epub 2014/02/04.
304. Akerman M, Fregoso OI, Das S, Ruse C, Jensen MA, Pappin DJ, et al. Differential connectivity of splicing activators and repressors to the human spliceosome. *Genome Biol.* 2015;16:119. Epub 2015/06/07.
305. Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol.* 2003;21:107-37. Epub 2002/11/05.
306. Rheinlander A, Schraven B, Bommhardt U. CD45 in human physiology and clinical medicine. *Immunol Lett.* 2018;196:22-32. Epub 2018/01/26.
307. Sun J, Zheng MY, Li YW, Zhang SW. Structure and function of Septin 9 and its role in human malignant tumors. *World J Gastrointest Oncol.* 2020;12(6):619-31. Epub 2020/07/24.
308. Marcus J, Bejerano-Sagie M, Patterson N, Bagchi S, Verkhusha VV, Connolly D, et al. Septin 9 isoforms promote tumorigenesis in mammary epithelial cells by increasing migration and ECM degradation through metalloproteinase secretion at focal adhesions. *Oncogene.* 2019;38(30):5839-59. Epub 2019/07/10.

309. Bai X, Bowen JR, Knox TK, Zhou K, Pendziwiat M, Kuhlenbaumer G, et al. Novel septin 9 repeat motifs altered in neuralgic amyotrophy bind and bundle microtubules. *J Cell Biol.* 2013;203(6):895-905. Epub 2013/12/18.
310. Kuzmic M, Castro Linares G, Leischner Fialova J, Iv F, Salaun D, Llewellyn A, et al. Septin-microtubule association via a motif unique to isoform 1 of septin 9 tunes stress fibers. *J Cell Sci.* 2022;135(1). Epub 2021/12/03.
311. Hassn Mesrati M, Syafruddin SE, Mohtar MA, Syahir A. CD44: A Multifunctional Mediator of Cancer Progression. *Biomolecules.* 2021;11(12). Epub 2021/12/25.
312. Xu H, Manivannan A, Crane I, Dawson R, Liversidge J. Critical but divergent roles for CD62L and CD44 in directing blood monocyte trafficking in vivo during inflammation. *Blood.* 2008;112(4):1166-74. Epub 2008/04/09.
313. Qadri M, Almadani S, Jay GD, Elsaid KA. Role of CD44 in Regulating TLR2 Activation of Human Macrophages and Downstream Expression of Proinflammatory Cytokines. *J Immunol.* 2018;200(2):758-67. Epub 2017/12/03.
314. Johnson P, Ruffell B. CD44 and its role in inflammation and inflammatory diseases. *Inflamm Allergy Drug Targets.* 2009;8(3):208-20. Epub 2009/07/16.
315. Andreoli V, Gehrau RC, Bocco JL. Biology of Kruppel-like factor 6 transcriptional regulator in cell life and death. *IUBMB Life.* 2010;62(12):896-905. Epub 2010/12/15.
316. Syafruddin SE, Mohtar MA, Wan Mohamad Nazarie WF, Low TY. Two Sides of the Same Coin: The Roles of KLF6 in Physiology and Pathophysiology. *Biomolecules.* 2020;10(10). Epub 2020/10/02.
317. Date D, Das R, Narla G, Simon DI, Jain MK, Mahabeleshwar GH. Kruppel-like transcription factor 6 regulates inflammatory macrophage polarization. *J Biol Chem.* 2014;289(15):10318-29. Epub 2014/01/05.
318. Zhang Y, Lei CQ, Hu YH, Xia T, Li M, Zhong B, et al. Kruppel-like factor 6 is a co-activator of NF-kappaB that mediates p65-dependent transcription of selected downstream genes. *J Biol Chem.* 2014;289(18):12876-85. Epub 2014/03/19.
319. Kim GD, Das R, Goduni L, McClellan S, Hazlett LD, Mahabeleshwar GH. Kruppel-like Factor 6 Promotes Macrophage-mediated Inflammation by Suppressing B Cell Leukemia/Lymphoma 6 Expression. *J Biol Chem.* 2016;291(40):21271-82. Epub 2016/08/20.
320. Kim GD, Ng HP, Chan ER, Mahabeleshwar GH. Kruppel-like factor 6 promotes macrophage inflammatory and hypoxia response. *FASEB J.* 2020;34(2):3209-23. Epub 2020/01/08.
321. Vang T, Miletic AV, Arimura Y, Tautz L, Rickert RC, Mustelin T. Protein tyrosine phosphatases in autoimmunity. *Annu Rev Immunol.* 2008;26:29-55. Epub 2008/02/29.
322. Kopf A, Kiermaier E. Dynamic Microtubule Arrays in Leukocytes and Their Role in Cell Migration and Immune Synapse Formation. *Front Cell Dev Biol.* 2021;9:635511. Epub 2021/02/27.
323. Akira S, Misawa T, Satoh T, Saitoh T. Macrophages control innate inflammation. *Diabetes Obes Metab.* 2013;15 Suppl 3:10-8. Epub 2013/09/18.
324. Verdier-Pinard P, Salaun D, Bouguenina H, Shimada S, Pophillat M, Audebert S, et al. Septin 9_i2 is downregulated in tumors, impairs cancer cell migration and alters subnuclear actin filaments. *Sci Rep.* 2017;7:44976. Epub 2017/03/25.
325. Hanania R, Sun HS, Xu K, Pustyl'nik S, Jeganathan S, Harrison RE. Classically activated macrophages use stable microtubules for matrix metalloproteinase-9 (MMP-9) secretion. *J Biol Chem.* 2012;287(11):8468-83. Epub 2012/01/25.

326. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell*. 2014;156(4):812-24. Epub 2014/02/18.
327. Koh WS, Porter JR, Batchelor E. Tuning of mRNA stability through altering 3'-UTR sequences generates distinct output expression in a synthetic circuit driven by p53 oscillations. *Sci Rep*. 2019;9(1):5976. Epub 2019/04/14.
328. Jia L, Mao Y, Ji Q, Dersh D, Yewdell JW, Qian SB. Decoding mRNA translatability and stability from the 5' UTR. *Nat Struct Mol Biol*. 2020;27(9):814-21. Epub 2020/07/29.
329. UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480-D9. Epub 2020/11/26.
330. Vallier P, Bata J, Colobert L. [Optimal conditions of alpha-amylase production by *Aspergillus oryzae* in liquid media]. *Ann Microbiol (Paris)*. 1977;128(3):359-71. Epub 1977/10/01. Conditions optimales de production d'alpha-amylase en milieu liquide par *Aspergillus oryzae*.
331. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13(4):227-32. Epub 2012/03/14.
332. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci Rep*. 2015;5:10775. Epub 2015/06/09.
333. Le Bras S, Moon C, Foucault I, Breittmayer JP, Deckert M. Abl-SH3 binding protein 2, 3BP2, interacts with CIN85 and HIP-55. *FEBS Lett*. 2007;581(5):967-74. Epub 2007/02/20.
334. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*. 2019;20(11):631-56. Epub 2019/07/26.
335. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. Epub 2012/10/30.
336. Singh ON, Burgess J. Characterization of albumin-alginic acid complex coacervation. *J Pharm Pharmacol*. 1989;41(10):670-3. Epub 1989/10/01.
337. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-9. Epub 2014/09/28.
338. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-30. Epub 2013/11/15.
339. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11(3):R25. Epub 2010/03/04.
340. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. Epub 2014/12/18.
341. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525-7. Epub 2016/04/05.
342. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521. Epub 2016/03/01.
343. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7(3):562-78. Epub 2012/03/03.
344. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13. Epub 2016/01/28.

345. Turro E, Su SY, Goncalves A, Coin LJ, Richardson S, Lewin A. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol.* 2011;12(2):R13. Epub 2011/02/12.
346. Frazee AC, Pertea G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol.* 2015;33(3):243-6. Epub 2015/03/10.
347. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics.* 2014;30(12):1660-6. Epub 2014/02/18.
348. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol.* 2018;19(1):85-97. Epub 2017/11/24.
349. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell.* 2017;65(4):631-43 e4. Epub 2017/02/19.
350. Shintaku H, Nishikii H, Marshall LA, Kotera H, Santiago JG. On-chip separation and analysis of RNA and DNA from single cells. *Anal Chem.* 2014;86(4):1953-7. Epub 2014/02/07.
351. Lovering RC, Gaudet P, Acencio ML, Ignatchenko A, Jolma A, Fornes O, et al. A GO catalogue of human DNA-binding transcription factors. *Biochim Biophys Acta Gene Regul Mech.* 2021;1864(11-12):194765. Epub 2021/10/22.
352. Chawla K, Tripathi S, Thommesen L, Laegreid A, Kuiper M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics.* 2013;29(19):2519-20. Epub 2013/08/13.
353. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* 2019;47(D1):D33-D8. Epub 2018/09/12.
354. Feng C, Song C, Liu Y, Qian F, Gao Y, Ning Z, et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res.* 2020;48(D1):D93-D100. Epub 2019/10/11.
355. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 2013;41(Database issue):D991-5. Epub 2012/11/30.
356. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004;306(5696):636-40. Epub 2004/10/23.
357. Czipa E, Schiller M, Nagy T, Kontra L, Steiner L, Koller J, et al. ChIPSummitDB: a ChIP-seq-based database of human transcription factor binding sites and the topological arrangements of the proteins bound to them. *Database (Oxford).* 2020;2020. Epub 2020/01/17.
358. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C. The sequence read archive. *Nucleic Acids Res.* 2011;39(Database issue):D19-21. Epub 2010/11/11.
359. Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, et al. hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genomics Proteomics Bioinformatics.* 2020;18(2):120-8. Epub 2020/08/29.
360. Cao J, Dong R, Jiang L, Gong Y, Yuan M, You J, et al. LncRNA-MM2P Identified as a Modulator of Macrophage M2 Polarization. *Cancer Immunol Res.* 2019;7(2):292-305. Epub 2018/11/22.

361. Zhang P, Cao L, Zhou R, Yang X, Wu M. The lncRNA Neat1 promotes activation of inflammasomes in macrophages. *Nat Commun.* 2019;10(1):1495. Epub 2019/04/04.
362. Yan B, Yao J, Liu JY, Li XM, Wang XQ, Li YJ, et al. lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. *Circ Res.* 2015;116(7):1143-56. Epub 2015/01/15.
363. Iwakiri J, Hamada M, Asai K. Bioinformatics tools for lncRNA research. *Biochim Biophys Acta.* 2016;1859(1):23-30. Epub 2015/08/15.
364. Zhu J, Fu H, Wu Y, Zheng X. Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci China Life Sci.* 2013;56(10):876-85. Epub 2013/10/05.
365. Allantaz F, Cheng DT, Bergauer T, Ravindran P, Rossier MF, Ebeling M, et al. Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PLoS One.* 2012;7(1):e29979. Epub 2012/01/26.
366. Salehi S, Taheri MN, Azarpira N, Zare A, Behzad-Behbahani A. State of the art technologies to explore long non-coding RNAs in cancer. *J Cell Mol Med.* 2017;21(12):3120-40. Epub 2017/06/21.
367. Dailey UG. *Current Medical Thought.* *J Natl Med Assoc.* 1916;8(3):166-7. Epub 1916/07/01.
368. Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci.* 2016;129(3):461-7. Epub 2016/01/21.
369. Hug N, Longman D, Caceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 2016;44(4):1483-95. Epub 2016/01/17.
370. Nguyen LS, Wilkinson MF, Gecz J. Nonsense-mediated mRNA decay: inter-individual variability and human disease. *Neurosci Biobehav Rev.* 2014;46 Pt 2:175-86. Epub 2013/11/19.
371. Bicknell AA, Ricci EP. When mRNA translation meets decay. *Biochem Soc Trans.* 2017;45(2):339-51. Epub 2017/04/15.
372. Smith JE, Baker KE. Nonsense-mediated RNA decay--a switch and dial for regulating gene expression. *Bioessays.* 2015;37(6):612-23. Epub 2015/03/31.
373. Silva AL, Ribeiro P, Inacio A, Liebhaber SA, Romao L. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA.* 2008;14(3):563-76. Epub 2008/01/31.
374. Silva AL, Pereira FJ, Morgado A, Kong J, Martins R, Faustino P, et al. The canonical UPF1-dependent nonsense-mediated mRNA decay is inhibited in transcripts carrying a short open reading frame independent of sequence context. *RNA.* 2006;12(12):2160-70. Epub 2006/11/02.
375. Inacio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, et al. Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. *J Biol Chem.* 2004;279(31):32170-80. Epub 2004/05/27.
376. Zhao LY, Song J, Liu Y, Song CX, Yi C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell.* 2020;11(11):792-808. Epub 2020/05/23.
377. Gjaltema RAF, Rots MG. Advances of epigenetic editing. *Curr Opin Chem Biol.* 2020;57:75-81. Epub 2020/07/04.
378. Santoso CS, Li Z, Lal S, Yuan S, Gan KA, Agosto LM, et al. Comprehensive mapping of the human cytokine gene regulatory network. *Nucleic Acids Res.* 2020;48(21):12055-73. Epub 2020/11/13.
379. Zhou L, Ma X, Sun F. The effects of protein interactions, gene essentiality and regulatory regions on expression variation. *BMC Syst Biol.* 2008;2:54. Epub 2008/06/28.

Annexe A.

Annexe A. Informations supplémentaires pour l'article « A transcriptome-based approach to identify functional modules within and across primary human immune cells »

Annexe de l'article présenté dans le chapitre 2 et publié dans PlosOne

A.1. Supplementary analyses

S2.1 Text. Analysis of all no-TEG and all no-MRG genes in *modules 22, 38 and 41*.

<https://doi.org/10.1371/journal.pone.0233543.s001>

To validate our approach of global and targeted functional annotation of co-expression modules, the entire list of genes within *modules 22, 38 and 41*, were analyzed using the GeneCards, PubMed and Google Scholar databases (see Materials and Methods section).

- **B lymphocyte-specific modules 22 and 38; cell activation and BCR receptor engagements.** As described in the main text, global analysis of *module 22* showed gene annotation enrichment for specific B cell functions and target analyses revealed genes involved in B cell receptor (BCR) structure and/or signaling and B cell activation, development and maturation. For *module 38*, no specific functions were identified via gene annotation enrichment analysis, but the targeted analyses identified genes involved in BCR activation, Ig assembly, and B cell signaling. Extending the annotation analyses to the entire set of genes in *module 22* identified encoding genes participating in the functions identified by the global and targeted analyses presented above. For example, the *BLK* gene, which is part of the complete list of genes in *module 22* but was not highlighted in the targeted analyses, is known to interact with BANK1 and CD79A in order to link BCR-mediated signaling to the formation of intracellular second messengers through PLCG2 (380). In addition, many other genes were involved in BCR stimulation (*CD19, FCRL1, FCRL2, MS5A1, and TEAD2*). These functions are presented in **Figure 4**. Some genes in *module 22* were implicated in apoptosis (*CD24, CDCA7L, DNASE1L3, RASSF6, SYVN1, and TP63*), Ig presentation and expression (*FCER2, IL4R, and MZB1*), B cell differentiation (*CXCR5* and *FCER2*) and Wnt signaling pathway (*CSNK1G3* and *ROR1*). These analyses suggest that one of the primary functions of the genes in *module 22* is B cell activation via the BCR signaling pathway. Extended analysis of *module 38* genes identified the classical B-cell genes *EBF1*, a transcriptional activator, and *IL6*, an inflammatory cytokine. Moreover, this module contains numerous genes involved in endosome/vesicle trafficking (*ACBD3, AP1S3, MARCH3, and USP6NL*), actin/cytoskeleton remodeling (*ARHGAP20, DMD, and ENAH*), and lipid metabolism (*ABCB4, FA2H, and SMPD2*). Although these have not been studied in B cells, they could potentially play a role in events pre- and post- engagement of the BCR such as endocytosis, signaling, and formation of lipid rafts. Taken together, these analyses suggest that *module 38* genes associated with the events that surround BCR receptor engagement and related pathways.
- **Module 41 shared by B lymphocytes and Monocytes; a gene expression program enabling antigen processing and presentation.** Global analysis of the 39 genes within this module revealed enrichments significantly enriched, with the three most significant

(%FDR <10⁻⁸) being “endosome”, “MHC classes I/II-like antigen recognition protein”, and “MHC class II protein complex”. Targeted analyses found a core set of genes that clearly highlights a central role in antigen processing and presentation. When exploring the remaining genes within this module, additional HLA class II paralogues can be found (*HLA-DQA1* and *HLA-DRB5*), as well as a gene that encodes the protein SCIMP, which is involved in HLA class II signal transduction and immune synapse formation (381, 382). In addition, this module contains a number of genes that encode proteins involved in vesicular traffic (*BLOC1S7*, *LAMP5*, *TBC1D5*, and *TRAK1*), potentially revealing novel players in the trafficking of endosomal vesicles involved in antigen processing and presentation. Although to a much lesser extent, there is also evidence that this module may also be involved in non-specific immune recognition (*CD180* and *CLEC4G*) and signaling (*LY86*, *GCNT1* and *IRAK1*) in response to exposure to pathogens (383, 384).

S2.2 Text. Additional analyses of genes highly-expressed in Neutrophils.

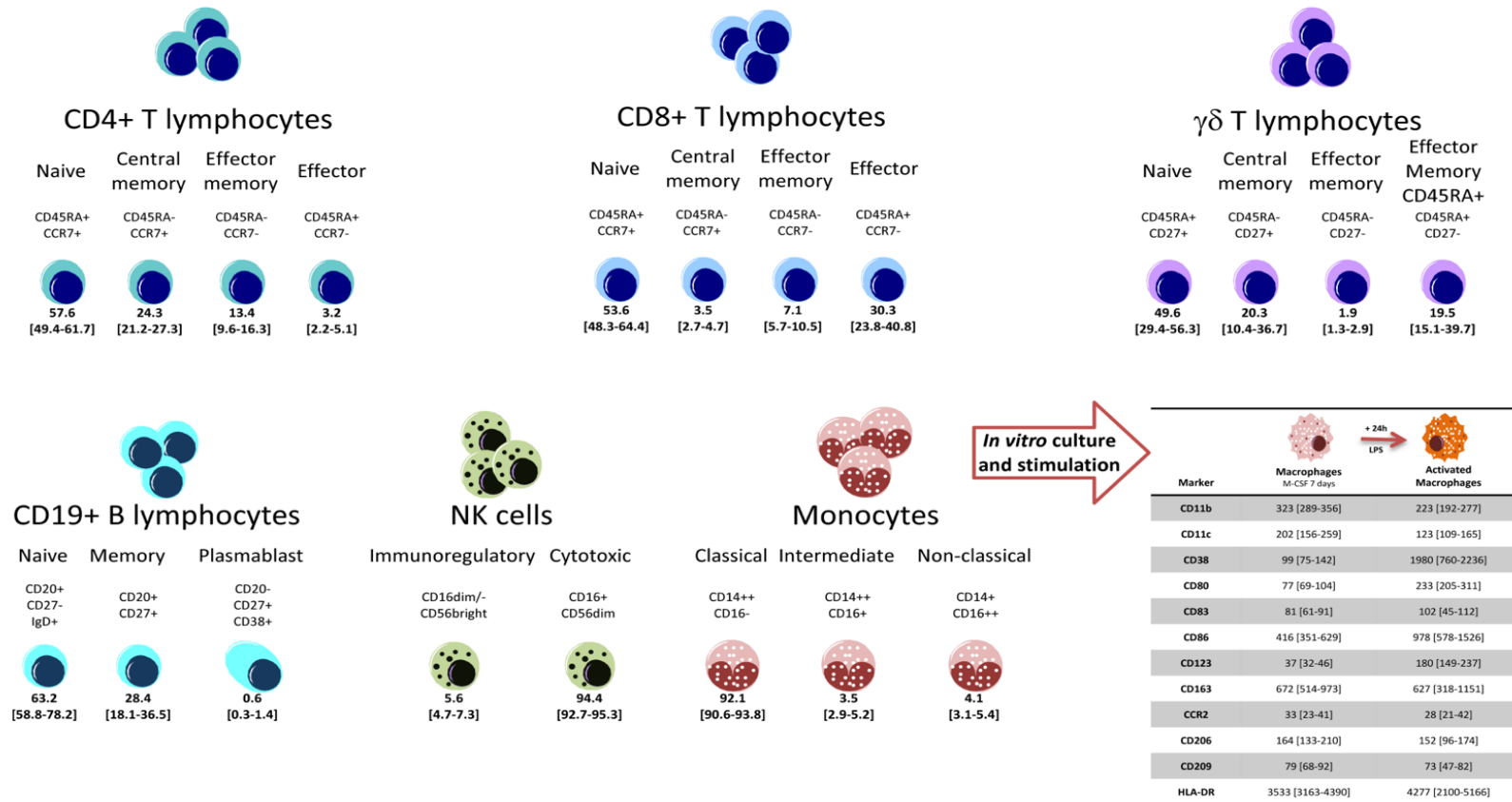
<https://doi.org/10.1371/journal.pone.0233543.s002>

The distinctive nature of the neutrophils was also observed in the co-expression analyses, with 13 of the 45 co-expression **modules (4, 7, 10-12, 14-16, 21, 25, 35, 36, 40)** highly associated with neutrophils (**Fig 3**); six of which were essentially exclusively associated with the neutrophils (**modules 11, 14-16, 21, 40**), whereas the others were also associated with other myeloid populations (monocytes and/or macrophages and/or LPS-activated macrophages). This is also reflected by a number of specific transcripts that are very different between these groups. In fact, there were 15 transcripts with a greater than 100-fold difference between the average number of reads in neutrophils compared to the average number of reads across all other cell types combined: *TRPM6*, *CA4*, *LOC101929866*, *NFE4*, *BMX*, *ALPL*, *LOC102724662*, *FCGR3B*, *CXCR1*, *CXCR2*, *TNFRSF10C*, *NECAB2*, *MME*, *MMP25*, and *CSNK1A1L*. Nine of these (underlined) are part of **module 11**, the module most associated with neutrophils. Many of these transcripts encode proteins that are hallmarks of neutrophils: *ALPL* (encoding the tissue non-specific alkaline phosphatase believed to play a critical role in the anti-microbial function of neutrophils by promoting its migration and ROS generation), *FCGR3B*, (low-affinity receptor for the Fc region of gamma immunoglobulins (IgG)) *CXCR1* and *CXCR2* (encoding for the receptor the interleukin 8), *TNFRSF10C* (encoding the receptor for TRAIL and involved in neutrophil apoptosis), *MME* (encoding CD10 which is present on the membrane of neutrophilic secretory granules), and *MMP25* (encoding Matrix Metalloproteinase 25, which is key to trans-endothelial migration of neutrophils to inflammatory sites) (385-389).

The neutrophil-related functions encoded by the other genes in this list are not as well characterized. Specifically, *TRPM6* encodes the enzyme-coupled transient receptor potential channel subfamily M member 6 that, along with its homolog *TRPM7*, are important for Mg²⁺ homeostasis but are distinct from all other ion channels in that they also have kinase domains (390). While Mg²⁺ is certainly critical to the function of multiple enzymes, these channels’ ability to conduct other divalent cations such as calcium, as well as their kinase activities, supports an even broader impact (390). For example, it has been shown that *TRPM7* regulates

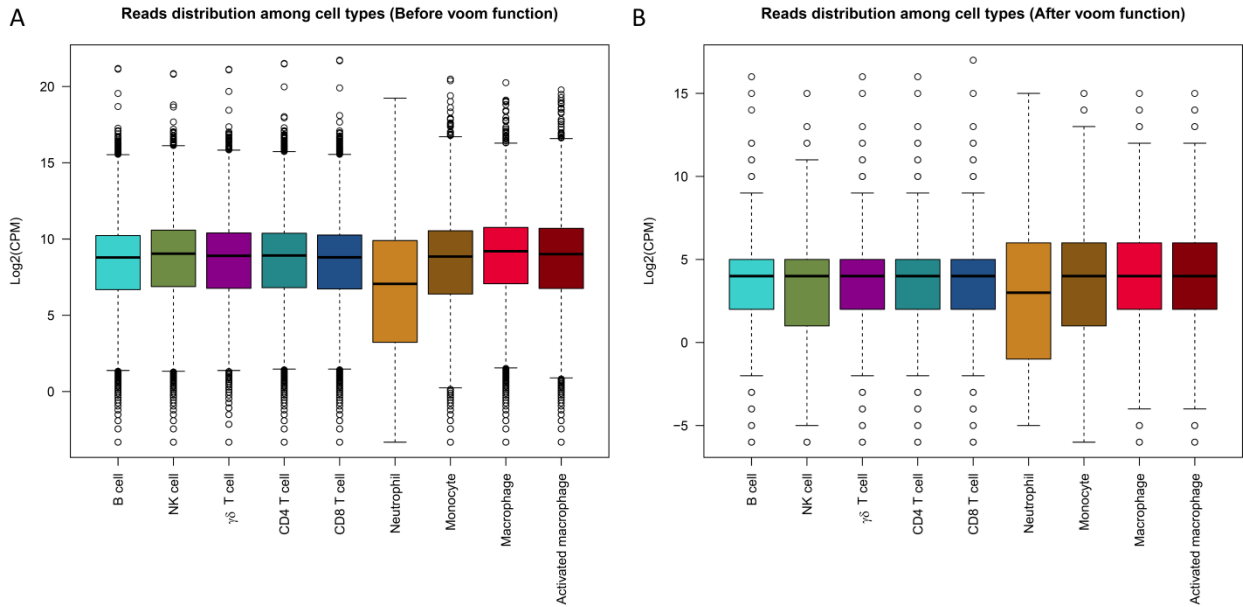
intracellular calcium signals in T cells as well as macrophage activation and polarization (391). While the role of *TRPM6* in the immune system is unknown, it's a high expression in neutrophils and absence in the other immune cells tested here suggested a neutrophil-specific role, which may include neutrophil chemotaxis given its interaction with *RACK1* (alias *GNB2L1*). *CA4* encodes the carbonic anhydrase 4 and is responsible for catalyzing the reversible hydration of carbon dioxide to form bicarbonate (HCO_3^-) and hydrogen (H^+) ions, and is suspected in having a role in the regulation of neutrophil activity and wound healing (392, 393). *NFE4* nuclear factor is involved in human globin gene switching in fetal erythroid cells (394), but may also have a role in transcriptional regulation of neutrophil genes. *BMX* encodes a non-receptor tyrosine kinase that plays central but diverse modulatory roles in various signaling processes in different cell types, with a likely role the regulation of the functional responsiveness of neutrophils to chemotactic factors (395). *NECAB2*, primarily studied as a neuronal calcium binding protein may regulate important neutrophil functions given its ability to modulate the adenosine A(2A) receptor (alias *ADORA2A*) and the latter's role in neutrophil apoptosis, autophagy and NETs (396-398). *CSNK1A1L* (Casein Kinase 1 Alpha 1 Like) is a poorly characterized member of the CK1 protein family involved in multiple functions likely mediated by its broad serine/threonine-protein kinase activity, including host defense response against infectious pathogens. For example, when CK1alpha phosphorylates type I interferon receptor 1 (*IFNAR1*) at Ser535 thereby induces the latter's ubiquitination and degradation (399).

A.2. Supplementary figures

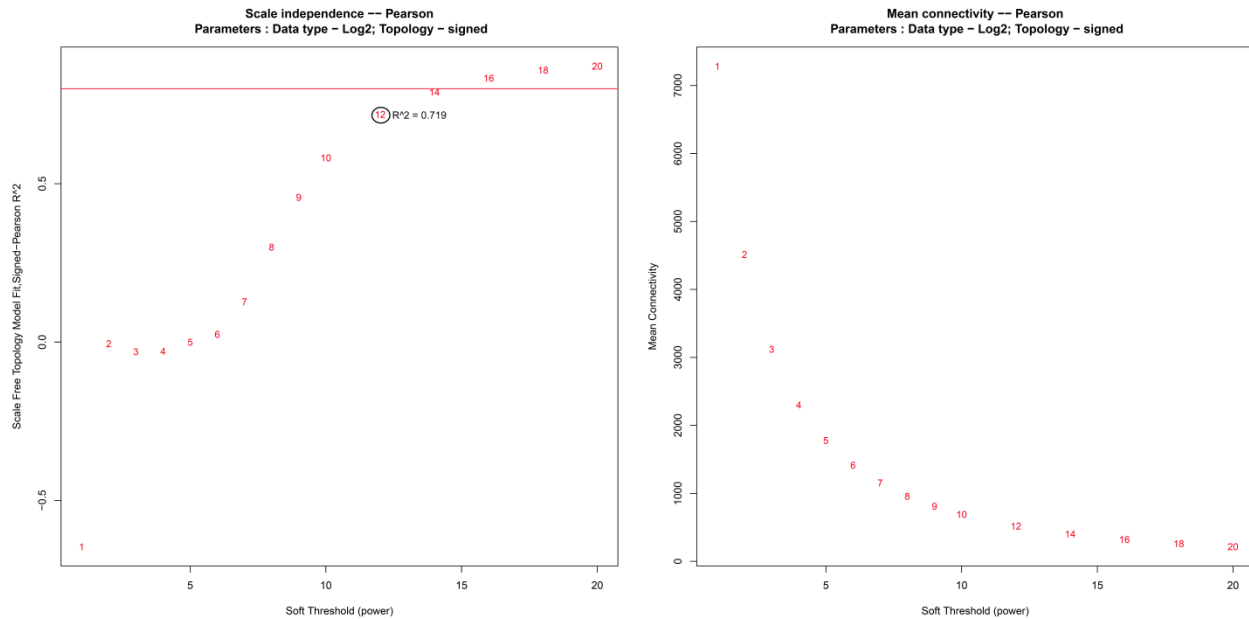


S2.1 Fig. Immunophenotyping of immune cell populations from human peripheral blood. Results are expressed as median percentage among CD3+CD4+ (n=12), CD3+CD8+ (n=12) or CD3+ TCRγδ+ (n=8) T lymphocyte populations respectively and among CD3-CD19+ B lymphocyte (n=12), CD3-CD56+ NK cell (n=12) or CD14+ monocyte (n=12) populations, with quartiles [Q1-Q3] for each subset. CD14+ enriched mononuclear cells were stimulated *in vitro* with M-CSF in order to obtain macrophages, with an additional 24h LPS

stimulation to obtain activated macrophages. Cells were stained with the mentioned surface markers and analyzed by flow cytometry. Results presented are the median MFI of specific staining minus unstained with quartiles [Q1-Q3]. (80, 257, 400-406). <https://doi.org/10.1371/journal.pone.0233543.s003>



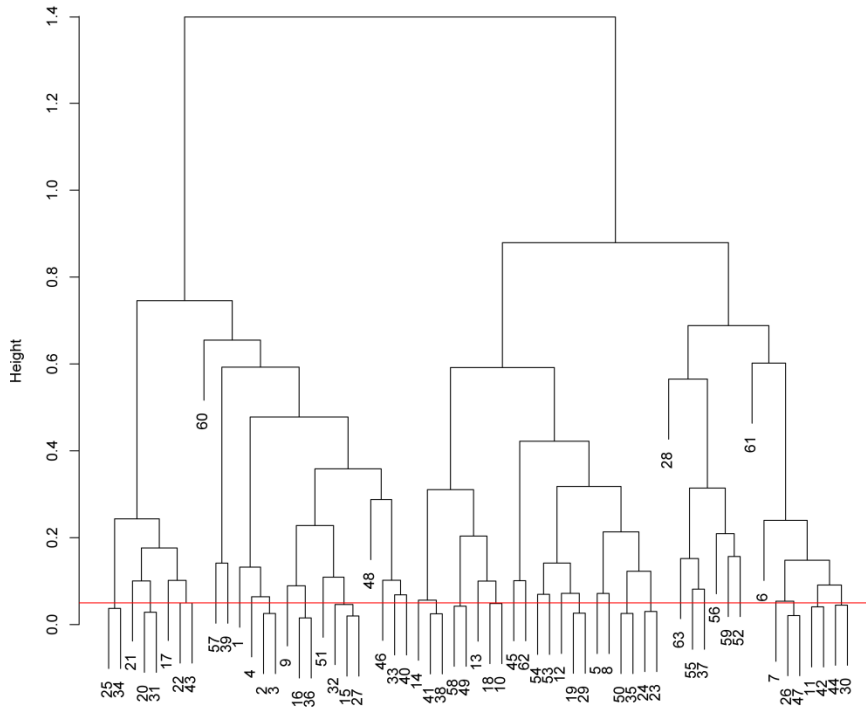
S2.2 Fig. Boxplot of gene expression distribution after normalization before and after normalization. Each boxplot represents the mean of gene expression in log₂ for a cell type in our 12 individuals. Panel A shows the log₂ of cpm normalized by library size and TMM normalization factor. Panel B shows the distribution after variance stabilization by R function *voom*. Even after normalization with *voom*, a function of Limma package, the median of neutrophil gene distribution is not aligned with other cell types because many more genes are not expressed in neutrophils than other primary immune cell types in our data (303). <https://doi.org/10.1371/journal.pone.0233543.s004>



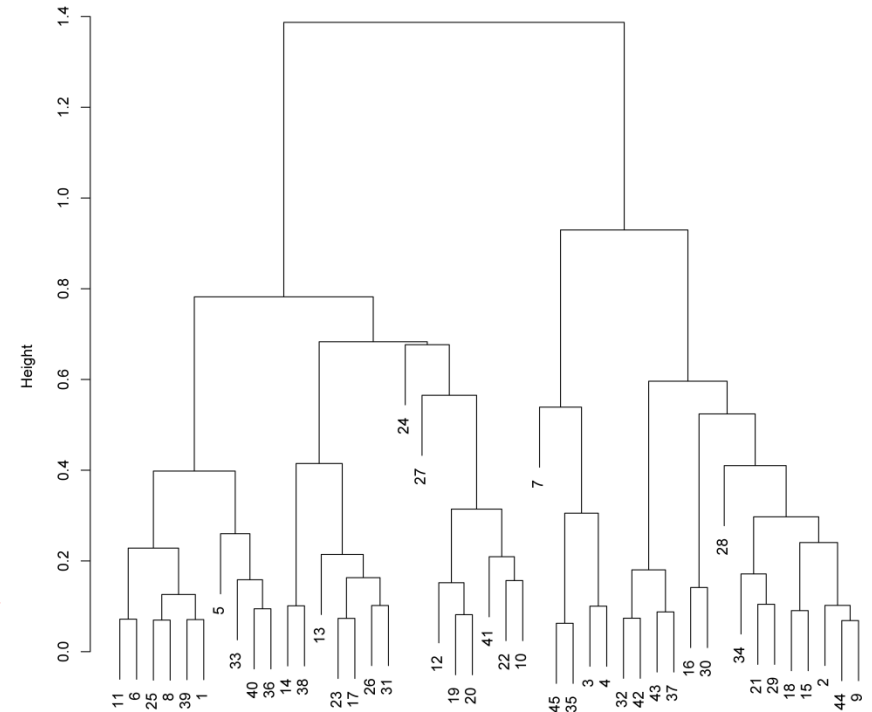
S2.3 Fig. Plot of scale independence and mean connectivity – Pearson correlation. Scale independence and mean value connectivity were used to select the β parameter, an exponent to the gene correlation matrix that determines the emphasis put on higher vs lower correlations (199). Value 12 was chosen for β because of it as a good trade-off between scale-free topology ($R^2 = 0.719$) and connectivity. Therefore, $\beta=12$ was used to compute dissimilarities between genes with the WGCNA functions, *adjacency* and *TomsimilarityFromExpr* (Topology Overlap Matrix Similarity From Expression) (132, 199). The red line on scale independence graph represents value 0.8 (suggested by the authors).

<https://doi.org/10.1371/journal.pone.0233543.s005>

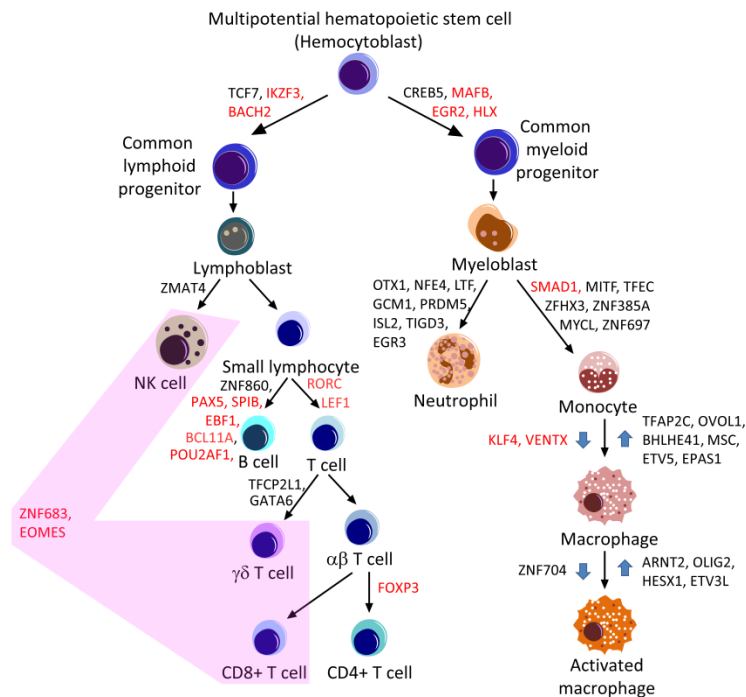
A Clustering of module eigengenes -- Pearson
 Parameters : beta - 12; minModuleSize - 20; netWorkType - signed; dataType - Log2



B Clustering of module eigengenes -- Pearson
 MEDissThres :0.05
 Parameters : beta - 12; minModuleSize - 20; netWorkType - signed; dataType - Log2

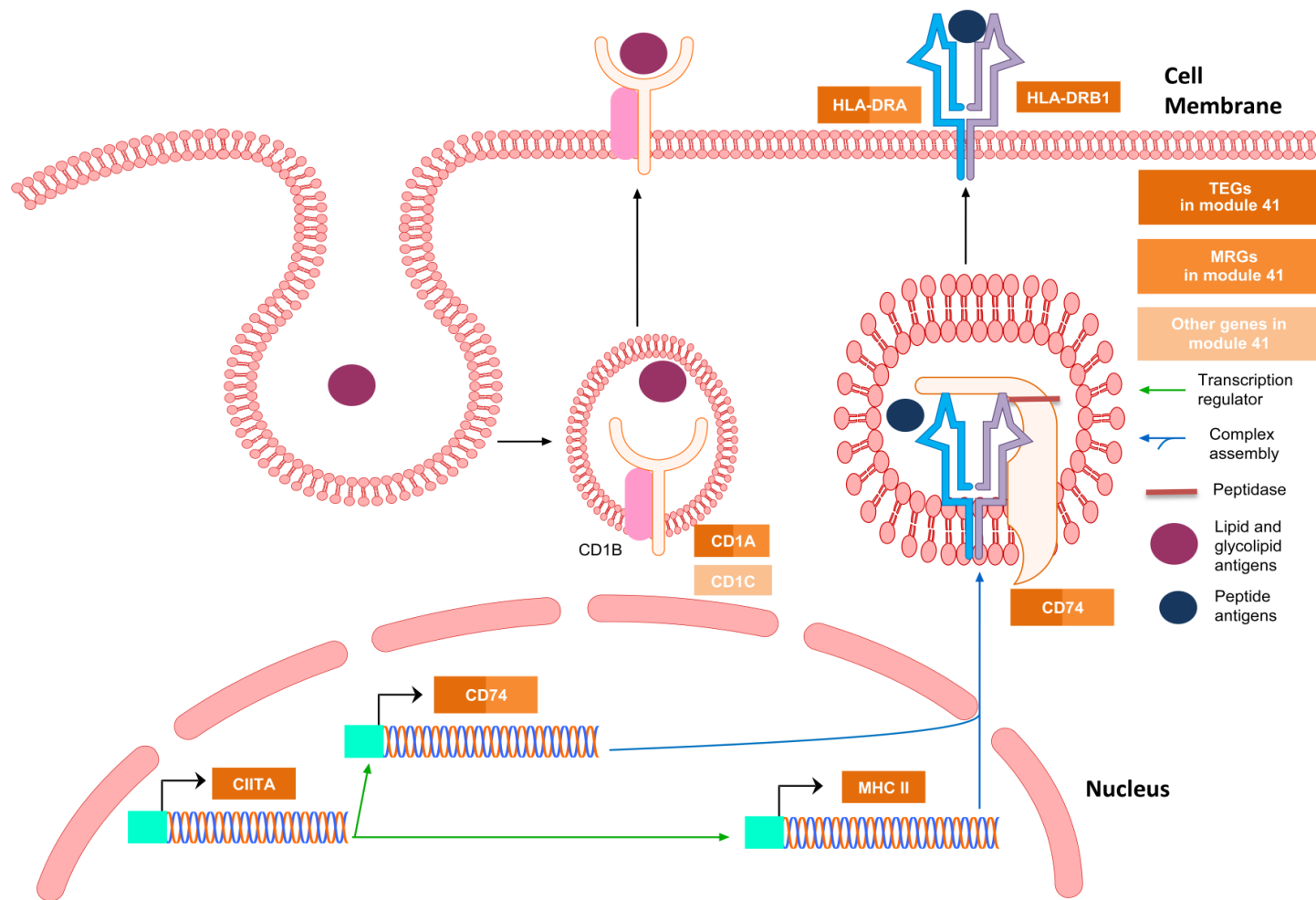


S2.4 Fig. Hierarchical trees of gene modules before and after cut tree. Graph A represents the tree of modules obtained with WGCNA tool. The red line on this graph is the value 0.05 who chooses to cut the tree to grouping similar modules in one. Graph B represents the new modules after cuts tree with new numeration. <https://doi.org/10.1371/journal.pone.0233543.s006>



S2.5 Fig. Hematopoietic differentiation scheme and associated transcription factors from differential gene expression. To identify transcription factors consistent with having a role in cell fate decisions we examined differential gene expression for all known human transcription factors (n=1638) (26). Schematic simplification is used as a representation of hematopoiesis from lymphoid and myeloid lineage. Transcription factors are in red and black. Red represents transcription factors known to be involved in the establishment and/or maintaining cell/lineage differentiation. The pink background color is used for transcription factors associated with cytotoxic cells. Blue arrows show increased or decreased expression of genes coding for transcription factors. Complete list of candidate TFs in **Table S5**. <https://doi.org/10.1371/journal.pone.0233543.s007>

S2.6 Fig. Heatmap of the correlation values (and p-values) of WGCNA modules with primary immune cell types. Columns represent modules computed with WGCNA and rows, primary immune cell types. In each square, the first number represents the correlation between a module and a given cell type and the second number in brackets is the associated p-value. <https://doi.org/10.1371/journal.pone.0233543.s008>



S2.8 Fig. Global and targeted analyses of genes within module 41, associated with B cells and monocytes, describe MHC class II and antigen processing and presentation functions. Global and targeted analyses of the genes within *module 41* were primarily associated with the presentation of peptide and lipid antigens. Genes in module 41 are represented in orange: *Top Expressed Genes* in dark orange, *Module Representative Genes* in intermediate orange and other genes in light orange. Genes from this module act together to establish Major Histocompatibility Complex class II function. To see the profile of gene expression mean of all genes of module 41 presented in this figure refers to the heatmap in **S7 Fig.** <https://doi.org/10.1371/journal.pone.0233543.s010>

A.3. *Supplementary tables*

S2.1 Table. List of antibodies used for lymphocyte/monocyte immunophenotyping.

<https://doi.org/10.1371/journal.pone.0233543.s011>

Antigen	Antibody clone	Supplier	Antigen	Antibody clone	Supplier
CD3	UCHT1	Biologend	CD27	O323	Biologend
CD4	RPA-T4	Biologend	CD38	HIT2	Biologend
CD8	RPA-T8	Biologend	CD45RA	HI100	Biologend
CD14	M5E2	Biologend	CD56	B159	BD Biosciences
CD16	3G8	Biologend	CCR7 (CD197)	150503	BD Biosciences
CD19	HIB19	Biologend	IgD	IA6-2	Biologend
CD20	2H7	Biologend	TCR $\gamma\delta$	B1	Biologend

S2.2 Table. List of antibodies used for macrophage immunophenotyping.

<https://doi.org/10.1371/journal.pone.0233543.s012>

Antigen	Antibody clone	Supplier	Antigen	Antibody clone	Supplier
CD11b	ICRF44	Biologend	CD123	6H6	Biologend
CD11c	3.9	Biologend	CD163	GHI/61	Biologend
CD38	HIT2	Biologend	CCR2 (CD192)	TG5/CCR2	Biologend
CD80	2D10	Biologend	CD206	15-2	Biologend
CD83	HB15e	Biologend	CD209	9E9A8	Biologend
CD86	IT2.2	Biologend	HLA-DR	L243	Biologend

S2.3 Table. Summary statistics of RNA-Seq data from raw reads through quality control steps. Values are reads at each step.

<https://doi.org/10.1371/journal.pone.0233543.s013>

	Monocytes		Macrophages		Activated macrophages		Neutrophils	
# of samples	12		12		12		11	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Raw FASTQ	25392853	2971647	25848399	4000125	26306327	4031458	26977675	4477712
Filtered FASTQ	24841788	2849773	25331060	3910406	25769770	3942984	26380553	4329742
Raw alignments	23926839	2633409	24215782	3732856	24610852	3742637	25275616	4183719
Properly paired	21001032	2349963	21533824	3318728	21872887	3497270	22082978	3558045

	B cell		CD4+ T cell		CD8+ T cell		γδ T cell		NK cell	
# of samples	12		12		12		8		12	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Raw FASTQ	23163724	2500054	24239648	3610972	23447063	2740439	23816789	2479157	24248847	2453890
Filtered FASTQ	22670510	2402592	23737575	3517709	22941165	2612434	23325587	2399476	23727620	2359340
Raw alignments	22010211	2252667	23180233	3572610	22493549	2803995	22597219	2371193	22942354	2253393
Properly paired	19050284	1894288	20358730	2976246	19611247	2172733	19908136	1998188	20207296	1904528

S2.4 Table. Summary of gene annotation enrichments from DAVID tool ($P \leq 0.05$).

<https://doi.org/10.1371/journal.pone.0233543.s014>

S2.5 Table. Differential gene expression and ratios of human transcription factors. First sheet: Differential gene expression and ratios of human TFs presented in [S5 Fig](#). Second sheet: Differential gene expression and ratios of all known human TFs expressed in our immune cell dataset ($n = 1112$). Third sheet: List of all known human TFs not expressed in our immune cell dataset. <https://doi.org/10.1371/journal.pone.0233543.s015>

S2.6 Table. Percentile, mean, standard deviation, median, and IQR of gene expression read counts.

First sheet: Mean of gene expression read count and percentile values. Second sheet: Standard deviation of gene expression read count. Third sheet: Median of gene expression read count. Fourth sheet: Interquartile range of gene expression read count.

<https://doi.org/10.1371/journal.pone.0233543.s016>

S2.7 Table. Summary of transcription factor binding site or TFBS enrichments from the ENCODE project. Empirical ChIP-Seq data in the GM12878 immortalized B cell line was used within the promoters of the genes within each module associated with B lymphocytes ($P \leq 0.05$).

<https://doi.org/10.1371/journal.pone.0233543.s017>

S2.8 Table. Literature review of key transcription factors involved in B-cell differentiation and maturation.

Cell types/ lineages	Known in B cell differentiation	Known in mature B cell	Known in Both	TF Gene Symbol	Literature
B-cell-lineage differentiation CLP to mature naïve B cell	✓	✓	✓	BACH2***	BACH2 inhibits myeloid differentiation in common lymphoid progenitors. (28273455, 25990863) BACH2 is involved in V(D)J recombination in pre-B cell and its expression locks differentiation of human mature B cells to plasma cells. (23852341, 28359033, 29129929, 24602812, 26751566)
	✓	✓	✓	EBF1	EBF1 and FOXO1 expression establish B cell identity and they activate the expression of PAX5 in early pro-B cells, coordinately with TCF3. (24679436) EBF1 is a key transcription factor of B cell specification and commitment, EBF1 is implicated in establishing and maintaining B-cell identity, EBF1 is also required to coordinate differentiation with cell proliferation and survival (25123279)
	✓	✓	✓	PAX5***	PAX5 is essential for the commitment of lymphoid progenitors to the B lymphocyte lineage. PAX5 represses B lineage inappropriate genes and activates B lineage-specific genes. PAX5 involved in the identity and function of B cell throughout B lymphopoiesis and in naïve and memory B cells. (21970955, 17440452, 26751566)
	✓	✓	✓	POU2AF1	With POU2F2 or POU2F1, POU2AF1 plays a role in B cell proliferation, induces surface marker expression during B-cell differentiation and enables B cells to respond normally to antigen receptor signals (24061476, 24688485, 19104664)
	✓	✓	✓	SPIB	SPIB enables B cells to appropriately respond to environmental cues and is a direct target of POU2AF1 (29127283, 16861304). SPI1 and SPIB are essential transcriptional regulators of B-cell differentiation (21768304)
	✓	✓	✓	BCLL11A	BCLL11A is essential for lymphoid development and negatively regulates p53 that inhibited apoptosis. (23230003; 24648892)
	✓	✓	✓	TCF3	TCF3 plays a particularly important role in B-lineage specification and involved in FOXO1, EBF1, and PAX5 regulation. (24679436, 27261530, 18538592) TCF3 is activated during BCR signaling (26081581) IKZF1, SPI1, and TCF3 play a role in functional lymphoid lineage priming. (25990863)
	✓	✓	✓	IKZF1	IKZF1, SPI1, and TCF3 play a role in functional lymphoid lineage priming. IKZF1 has a role in the formation of pre-BCR in the pre-B cell stage (25990863, 23303821) IKZF1 limit the response of naïve splenic B cells to BCR signals and is a negative regulator of follicular B cell activation. (26775846)
Necessary but not sufficient in B-cell-lineage differentiation	✓	✓	✓	CTCF	CTCF modifies DNA structures by enhancer-promoter interactions that involve lineage-specific transcriptional regulators such as TCF3, EBF1, and FOXO1. (24679436)
	✓	✓	✓	SPI1	IKZF1, SPI1, and TCF3 play a role in functional lymphoid lineage priming. (25990863) With IRF4 or IRF8, SPI1 involves pre-B and early-B cell development and they suppress pre-B-cell leukemia associated with reduced expression of the established B-lineage tumor suppressor genes, IKZF1 and SPIB. (26932576)
	✓	✓	✓	RUNX1	RUNX1 can regulate B-cell growth. (16584381, 25205721)
	✓	✓	✓	POU2F2	With POU2AF1, POU2F2 plays a role in B cell proliferation, induces surface marker expression during B-cell differentiation and enables B cells to respond normally to antigen receptor signals (24061476, 24688485, 19104664)
	✓	✓	✓	FOXO1	With TCF3, FOXO1 induces the expression of EBF1 in pre-pro-B cell and they activate the expression of PAX5 in early pro-B cells. FOXO1 and EBF1 establish B cell identity. (24679436) FOXO1 promotes differentiation, proliferation, survival, immunoglobulin gene rearrangement, and class switching in B cells. (29431075)
	✓	✓	✓	IRF4***	With SPI1, IRF4 involved in pre-B and early-B cell development and they suppress pre-B-cell leukemia associated with reduced expression of the established B-lineage tumor suppressor genes, IKZF1 and SPIB. (26932576) PRDM1 and IRF4 are essential for the generation of plasma cells. IRF4 is essential for the survival of plasma cells. (26779600)
	✓	✓	✓	IRF8	With SPI1, IRF8 involved in early-B-cell development and of recirculating B-cell numbers. They restore IKZF1 and SPIB gene expression in pre-B cells to inhibit leukemic cell growth. (26932576) IRF8 and SPI1 complex negatively regulates plasma cell differentiation by concurrently promoting the expression of BCL6 and PAX5 and repressing AID and PRDM1. (25288399)
Naïve B cell to plasma cell		✓		PRDM1	PRDM1 acts as a master regulator of plasma cell differentiation. (25115512) PRDM1 and IRF4 are essential for the generation of plasma cells. PRDM1 regulates the transcriptional identity of plasma cells, whose XBP1 expression. PRDM1 is also involved in antibody secretion by plasma cells. (26779600, 15345222)
		✓		XBP1	XBP1 expression is regulated by PRDM1 in plasma cells. (26779600) XBP1 is selectively and specifically required to initiate and maintaining plasma-cell differentiation from mature B cells. (11460154) XBP1 is involved in secretory plasma cell functions such as cell size, lysosome content, mitochondrial mass and function, ribosome numbers, and total protein synthesis. (15345222)

*** The TFs IRF4, PAX5, and BACH2, along with the absence of BCL6, have been reported to also play a role in the maturation of mature naïve to memory B cells. (13); Number in parenthesis is PMID number, which is the unique identifier number used in PubMed for each article.

<https://doi.org/10.1371/journal.pone.0233543.s018>

A.4. *Supplementary references*

1. Bernal-Quiros M, Wu YY, Alarcon-Riquelme ME, Castillejo-Lopez C. BANK1 and BLK act through phospholipase C gamma 2 in B-cell signaling. *PLoS One*. 2013;8(3):e59842. Epub 2013/04/05.
2. Draber P, Vonkova I, Stepanek O, Hrdinka M, Kucova M, Skopcova T, et al. SCIMP, a transmembrane adaptor protein involved in major histocompatibility complex class II signaling. *Mol Cell Biol*. 2011;31(22):4550-62. Epub 2011/09/21.
3. Kralova J, Fabisik M, Pokorna J, Skopcova T, Malissen B, Brdicka T. The Transmembrane Adaptor Protein SCIMP Facilitates Sustained Dectin-1 Signaling in Dendritic Cells. *J Biol Chem*. 2016;291(32):16530-40. Epub 2016/06/12.
4. Weekes MP, Antrobus R, Talbot S, Hor S, Simecek N, Smith DL, et al. Proteomic plasma membrane profiling reveals an essential role for gp96 in the cell surface expression of LDLR family members, including the LDL receptor and LRP6. *J Proteome Res*. 2012;11(3):1475-84. Epub 2012/02/02.
5. Dominguez-Soto A, Aragonese-Fenoll L, Martin-Gayo E, Martinez-Prats L, Colmenares M, Naranjo-Gomez M, et al. The DC-SIGN-related lectin LSECtin mediates antigen capture and pathogen binding by human myeloid cells. *Blood*. 2007;109(12):5337-45. Epub 2007/03/07.
6. Weiss MJ, Henthorn PS, Lafferty MA, Slaughter C, Raducha M, Harris H. Isolation and characterization of a cDNA encoding a human liver/bone/kidney-type alkaline phosphatase. *Proc Natl Acad Sci U S A*. 1986;83(19):7182-6. Epub 1986/10/01.
7. Bergin DA, Reeves EP, Meleady P, Henry M, McElvaney OJ, Carroll TP, et al. alpha-1 Antitrypsin regulates human neutrophil chemotaxis induced by soluble immune complexes and IL-8. *J Clin Invest*. 2010;120(12):4236-50. Epub 2010/11/10.
8. Kamohara H, Matsuyama W, Shimozato O, Abe K, Galligan C, Hashimoto S, et al. Regulation of tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) and TRAIL receptor expression in human neutrophils. *Immunology*. 2004;111(2):186-94. Epub 2004/03/19.
9. Elghetany MT. Surface antigen changes during normal neutrophilic development: a critical review. *Blood Cells Mol Dis*. 2002;28(2):260-74. Epub 2002/06/18.
10. Starr AE, Bellac CL, Dufour A, Goebeler V, Overall CM. Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *J Biol Chem*. 2012;287(16):13382-95. Epub 2012/03/01.
11. Montell C. Mg²⁺ homeostasis: the Mg²⁺-sensitive TRPM channels. *Curr Biol*. 2003;13(20):R799-801. Epub 2003/10/17.
12. Nadolni W, Zierler S. The Channel-Kinase TRPM7 as Novel Regulator of Immune System Homeostasis. *Cells*. 2018;7(8). Epub 2018/08/22.
13. Coakley RJ, Taggart C, Greene C, McElvaney NG, O'Neill SJ. Ambient pCO₂ modulates intracellular pH, intracellular oxidant generation, and interleukin-8 secretion in human neutrophils. *J Leukoc Biol*. 2002;71(4):603-10. Epub 2002/04/03.

14. Barker H, Aaltonen M, Pan P, Vahatupa M, Kaipainen P, May U, et al. Role of carbonic anhydrases in skin wound healing. *Exp Mol Med*. 2017;49(5):e334. Epub 2017/05/20.
15. Zhou W, Zhao Q, Sutton R, Cumming H, Wang X, Cerruti L, et al. The role of p22 NF-E4 in human globin gene switching. *J Biol Chem*. 2004;279(25):26227-32. Epub 2004/04/16.
16. Gilbert C, Levasseur S, Desaulniers P, Dusseault AA, Thibault N, Bourgoin SG, et al. Chemotactic factor-induced recruitment and activation of Tec family kinases in human neutrophils. II. Effects of LFM-A13, a specific Btk inhibitor. *J Immunol*. 2003;170(10):5235-43. Epub 2003/05/08.
17. Xu K, Cooney KA, Shin EY, Wang L, Deppen JN, Ginn SC, et al. Adenosine from a biologic source regulates neutrophil extracellular traps (NETs). *J Leukoc Biol*. 2019. Epub 2019/03/26.
18. Liu YW, Yang T, Zhao L, Ni Z, Yang N, He F, et al. Activation of Adenosine 2A receptor inhibits neutrophil apoptosis in an autophagy-dependent manner in mice with systemic inflammatory response syndrome. *Sci Rep*. 2016;6:33614. Epub 2016/09/21.
19. Canela L, Lujan R, Lluís C, Burgueno J, Mallol J, Canela EI, et al. The neuronal Ca(2+) - binding protein 2 (NECAB2) interacts with the adenosine A(2A) receptor and modulates the cell surface expression and function of the receptor. *Mol Cell Neurosci*. 2007;36(1):1-12. Epub 2007/08/11.
20. Jiang S, Zhang M, Sun J, Yang X. Casein kinase 1alpha: biological mechanisms and theranostic potential. *Cell Commun Signal*. 2018;16(1):23. Epub 2018/05/26.
21. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol*. 2012;12(3):191-200. Epub 2012/02/22.
22. Pitard V, Roumanes D, Lafarge X, Couzi L, Garrigue I, Lafon ME, et al. Long-term expansion of effector/memory Vdelta2-gammadelta T cells is a specific blood signature of CMV infection. *Blood*. 2008;112(4):1317-24. Epub 2008/06/10.
23. Farber DL, Yudanin NA, Restifo NP. Human memory T cells: generation, compartmentalization and homeostasis. *Nat Rev Immunol*. 2014;14(1):24-35. Epub 2013/12/18.
24. Angelo LS, Banerjee PP, Monaco-Shawver L, Rosen JB, Makedonas G, Forbes LR, et al. Practical NK cell phenotyping and variability in healthy adults. *Immunol Res*. 2015;62(3):341-56. Epub 2015/05/28.
25. Boyette LB, Macedo C, Hadi K, Elinoff BD, Walters JT, Ramaswami B, et al. Phenotype, function, and differentiation potential of human monocyte subsets. *PLoS One*. 2017;12(4):e0176460. Epub 2017/04/27.
26. Verreck FA, de Boer T, Langenberg DM, van der Zanden L, Ottenhoff TH. Phenotypic and functional profiling of human proinflammatory type-1 and anti-inflammatory type-2 macrophages in response to microbial antigens and IFN-gamma- and CD40L-mediated costimulation. *J Leukoc Biol*. 2006;79(2):285-93. Epub 2005/12/07.
27. Rey-Giraud F, Hafner M, Ries CH. In vitro generation of monocyte-derived macrophages under serum-free conditions improves their tumor promoting functions. *PLoS One*. 2012;7(8):e42656. Epub 2012/08/11.
28. Pang DJ, Neves JF, Sumaria N, Pennington DJ. Understanding the complexity of gammadelta T-cell subsets in mouse and human. *Immunology*. 2012;136(3):283-90. Epub 2012/03/06.
29. Farag SS, Caligiuri MA. Human natural killer cell development and biology. *Blood Rev*. 2006;20(3):123-37. Epub 2005/12/21.
30. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29. Epub 2014/02/04.

31. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17. Epub 2006/05/02.
32. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. Epub 2008/12/31.
33. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. *Cell*. 2018;172(4):650-65. Epub 2018/02/10.

Annexe B.

***Annexe B. Informations supplémentaires pour l'article
« Identifying transcript-level differential expression in primary human
immune cells »***

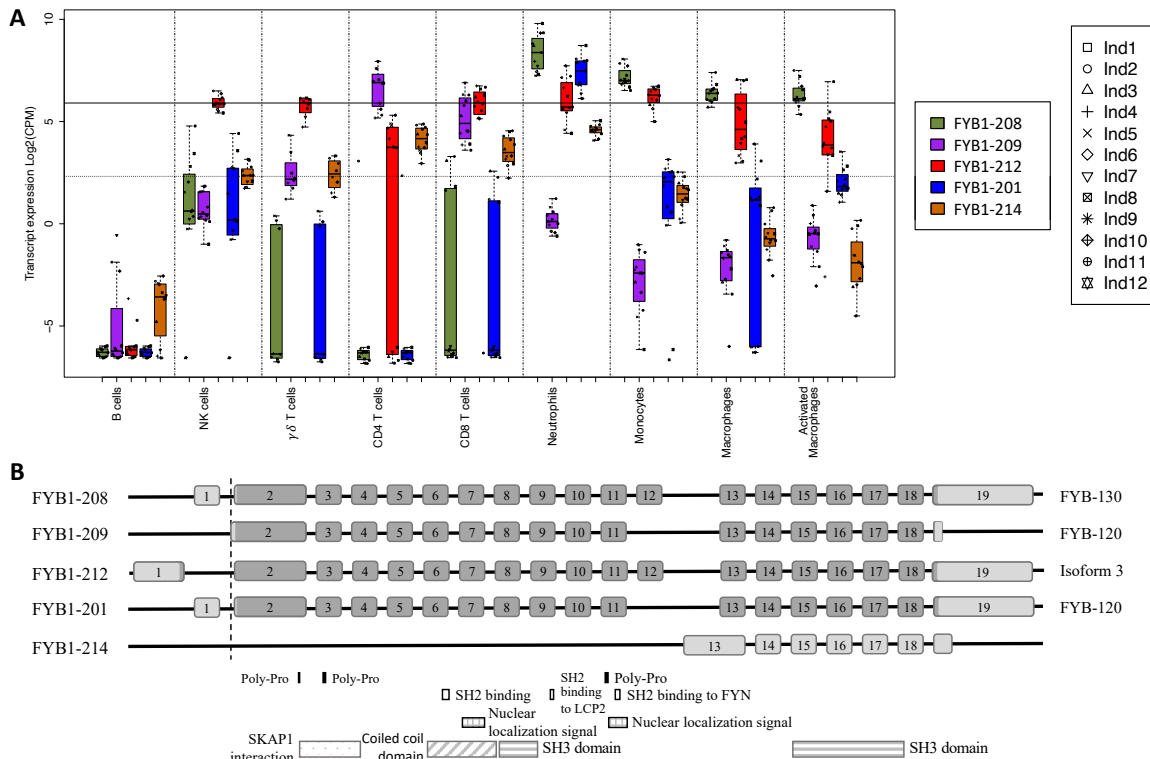
Annexe de l'article présenté dans le chapitre 3 et soumis dans Molecular Immunology

B.1. *Supplementary texts*

BOX 1. *Genes with two or more transcripts having distinct expression patterns in this dataset.*

FYB1

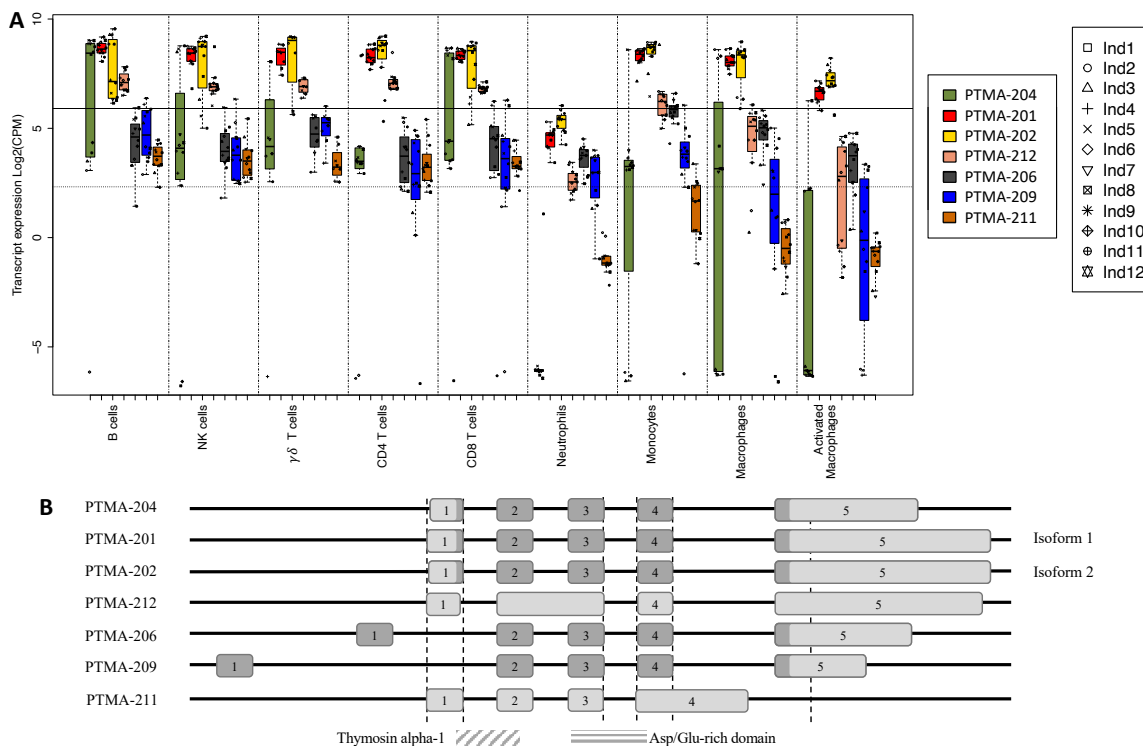
While not as extensively studied as *PTPRC*, it is also known that the *FYB1* gene encodes at least two different isoforms having different functions (407). The *FYB1* gene encodes the FYN Binding Protein 1, also known as FYB-120/130 or ADAP, and has been shown to be a key player in multiple intracellular signaling cascades in myeloid and lymphoid cells, impacting on a wide range of functions including cell activation, adhesion, and motility (408). It was first demonstrated that the two different molecular masses of FYB1 (120 kDa vs 130 kDa) were encoded by two different RNA isoforms that differed by 46 amino acids toward the carboxyl-terminal region of the protein and had differential binding to T cell antigen receptor (TCR) adaptor protein SLP-76/LCP2, with preferential expression of FYB-130 in mature T-cells (407). In our data, we detected five different transcripts, four of which are protein-coding: transcript FYB-208 encodes FYB-130, transcripts *FYB-201* and FYB-209 both encode FYB-120 (with the former having longer UTR sequences than the latter), and transcript FYB-212 encodes a protein ten amino acids longer than FYB-120 and FYB-130 that is called isoform 3. It has been shown that the amino (N)-terminal present in the two isoforms FYB-120 and FYB-130, binds mAbp1, a protein which is recruited to newly forming F-actin structures and mediates the signaling that leads to RAC activation in macrophages, although only FYB-120 was tested (409). In our data, *FYB-212* was highly expressed in all cells except B cells and CD4 T cells; in the latter it was *FYB-209* that was predominant. FYB-201, on the other hand, was only highly expressed in neutrophils. Finally, *FYB-208* was highly expressed in myeloid cells. This complex pattern suggests a greater cell-type specific regulation of FYB1-dependent mechanisms than has previously been reported.



Data summary for *FYB1* gene. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations. Dotted line represents the minimal threshold to identify low-expression transcripts (log₂(5 CPM)) in our dataset; the dark line represents the minimal threshold value used to define moderately/highly expressed transcript (log₂(60 CPM)). Each individual is represented by a different symbol. The Box illustrates the four quartiles of data, with the *middle bold* line indicating the median value; the box then includes 50% of the data points, 25% are above and 25% below. The whiskers extend up to 1.5 times the interquartile range (height of the box). Data points outside this range are generally considered outliers. **B)** Graphical representation of the different alternate transcripts, identified in our data, and their respective protein isoforms for the *PTMRC* gene. The transcript name from GENCODE is on the left while the protein name from UniProt is on the right when available. Exonic regions in dark grey are coding sequences, while regions in light grey are untranslated regions. The junctions between light and dark grey indicate the location of translation start and stop sites. FYB1-209 codes for the canonical protein isoform.

PMTA

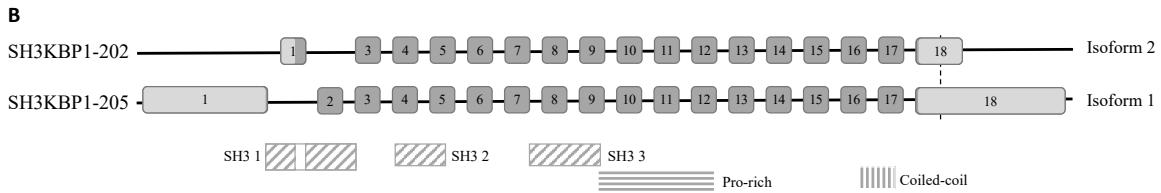
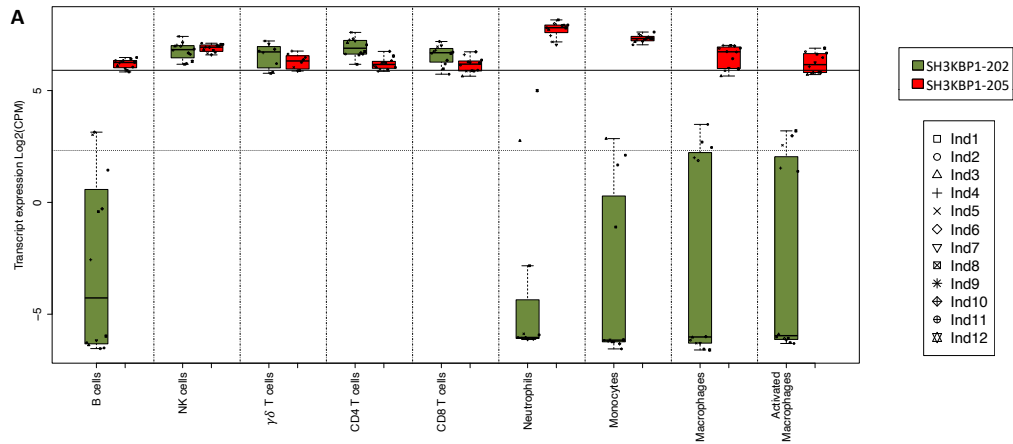
Prothymosin alpha (*PTMA*) was initially reported as a regulator and enhancer of type I IFN secretion in CD8+ T cells (410). More recently, two additional isoforms were identified in human macrophages, named isoB and p7 that were shown to upregulate multiple cytokine and chemokines in TLR4-dependent or -independent mechanisms. In our data, we detected multiple transcripts, with *PTMA-204* being predominantly expressed in lymphocytes, with weaker (and more variable) or expression in myeloid cells. Transcripts *PTMA-206*, *PTMA-209*, and *PTMA-211* are more modestly expressed, but are consistently found in all lymphocytes. Transcripts *PTMA-201* and *PTMA-202*, on the other hand, are highly expressed across all cell types examined. *PTMA-204* encode to three fewer amino acids than *PTMA-202* and they are in exon 5 in Asp/Glu-rich domain.



Data summary for *PTMA* gene. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *PTMA* transcripts, identified in our data, and their respective protein isoforms. The *PTMA* protein has a Thymosin alpha-1 and Asp/Glu rich domains. *PTMA-201* codes for the canonical protein isoform.

SH3KBP1

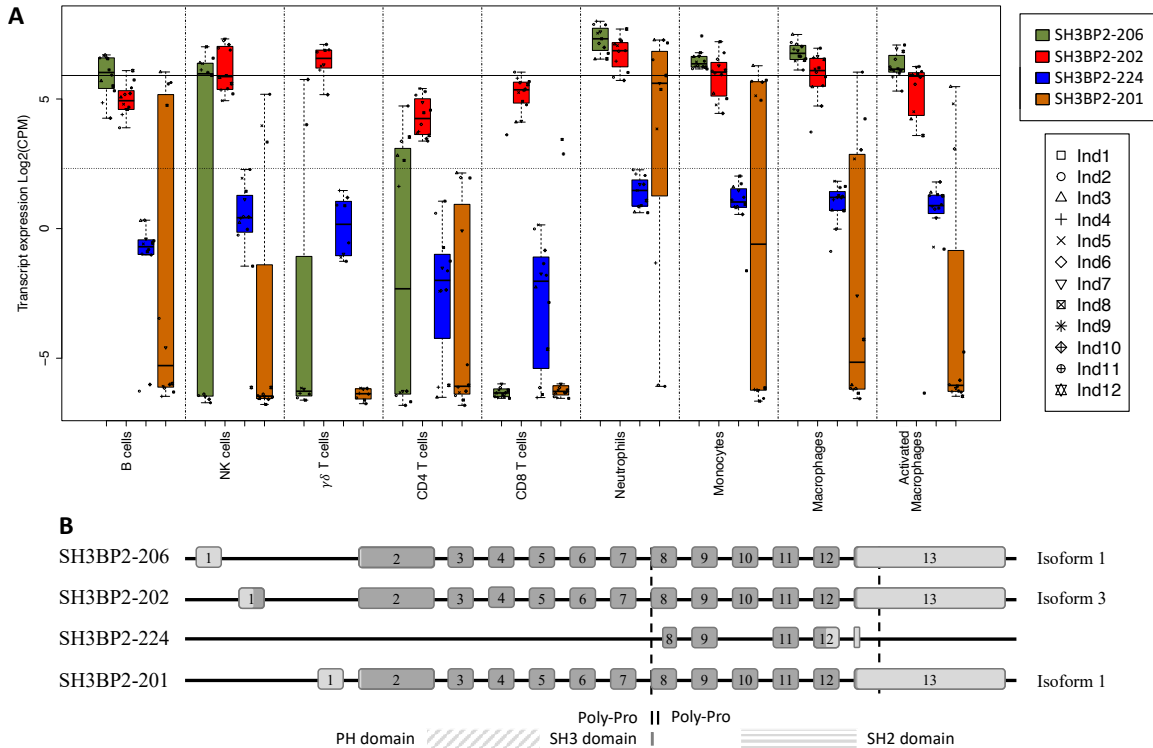
SH3KBP1 (SH3 Domain Containing Kinase Binding Protein 1) is an adaptor protein involved in numerous cellular processes including endocytosis and cytoskeletal rearrangement in T cells or effector pathways in B cells (333, 411, 412). This protein contains three SH3 domains, a proline rich domain and a coiled coil near its C-terminus. The three SH3 domains interact with proteins containing proline-rich sequences, the proline-rich region bind to SH3 domain of target proteins, and the C-terminal coiled-coil region mediates dimerization or trimerization of SH3KBP1 and binds cell membrane on phosphatidic acid (413). It has previously been reported that *SH3KBP1* has two different isoforms: a full-length isoform named CIN85 and the truncated CD2BP3 that has lost part of its first SH3 domain (414). In the current study we find that that *SH3KBP1-202* (corresponding to CIN85) is expressed in T lymphocyte populations but not B lymphocytes or myeloid populations, whereas *SH3KBP1-205* (corresponding to CD2BP3) found in all cell types.



Data summary for *SH3KBP1* gene. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *SH3KBP1* transcripts, identified in our data, and their respective protein isoforms. The SH3KBP1 protein has three SH3 domains, a pro-rich domain and a coiled-coil region. White space in SH3 1 box corresponds to coding region of *SH3KBP1-202* exon 1 that do not encode SH3 1 domain. *SH3KBP1-205* codes for the canonical protein isoform.

SH3BP2

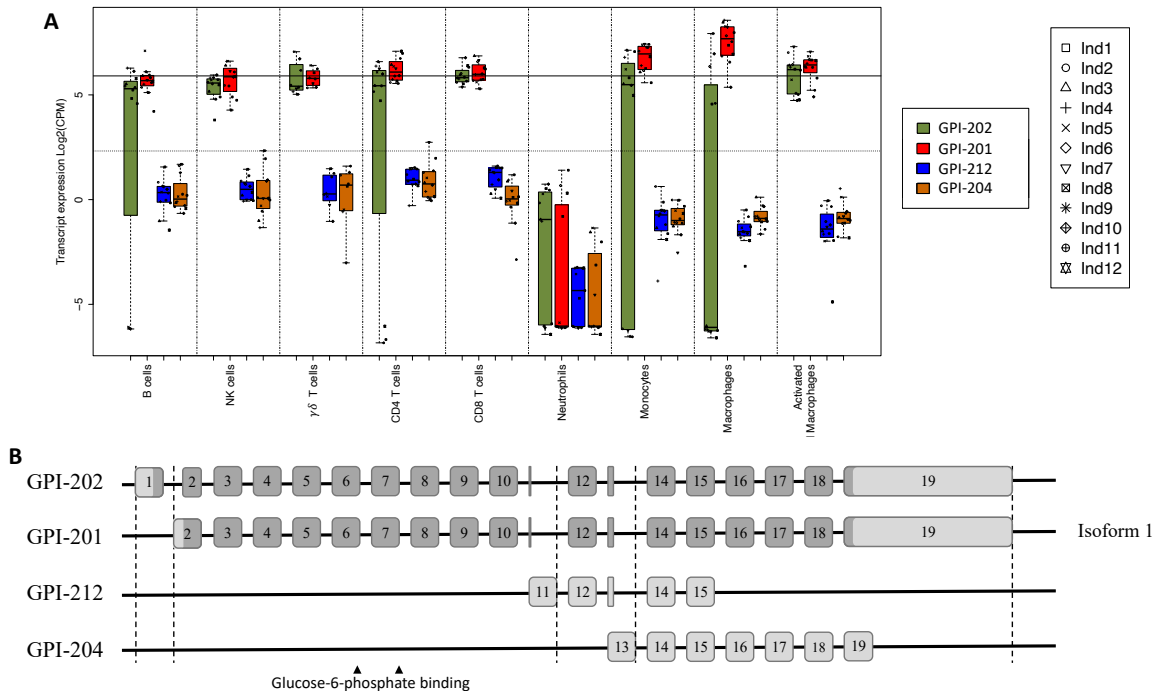
Intriguingly, however, SH3BP2 is a known SH3KBP1 interactor and has been shown to be involved in leukocyte signaling downstream of Src/Syk-kinase coupled immunoreceptors. It was shown that SH3BP2 binds to the third SH3 domain of SH3KBP1 via a PVPTPR motif within its Proline-rich domain (333). The expression pattern of *SH3BP2* transcripts appears complementary to those of *SH3KBP1* with *SH3BP2-206* being expressed in B lymphocytes but not T lymphocytes, although also expressed in myeloid populations. *SH3BP2-202*, on the other hand, appears to essentially be expressed in all populations studied.



Data summary for *SH3BP2* gene. **A)** Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *SH3BP2* transcripts, identified in our data, and their respective protein isoforms. The *SH3BP2* protein has a PH domain, one SH2 domain and two Poly-Pro motifs. One of these motifs has similar position with SH3 domain. The *SH3BP2* protein has two poly-pro *SH3BP1-206* codes for the canonical protein isoform.

GPI

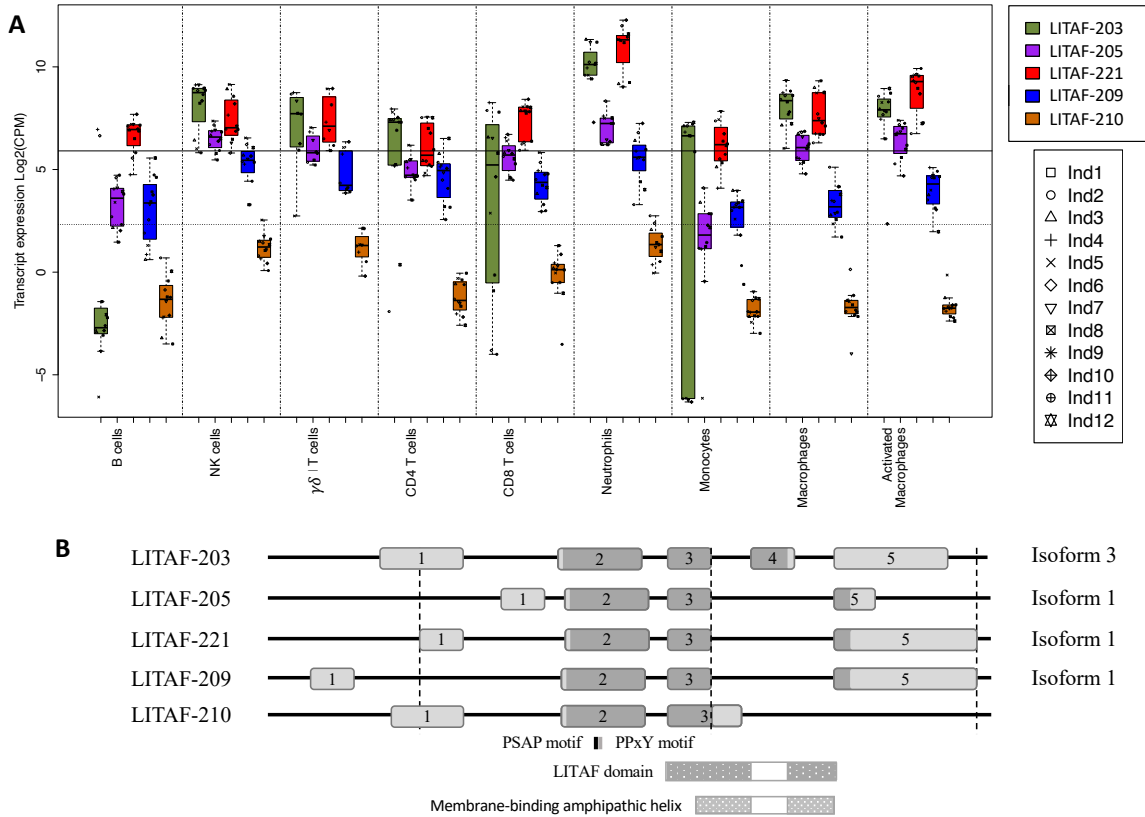
In terms of the *GPI* (alias PGI and Neuroleukin) gene which encodes Glucose-6-phosphate isomerase, two different isoforms were previously identified although differential expression of these across immune cells was not previously studied (415). Glucose-6-phosphate isomerase is known to not only catalyze the second step in the glycolysis pathway but when secreted acts as a cytokine, for example to stimulate immunoglobulin secretion (416). In our data, transcript *GPI-201* was expressed in all cell populations examined except for neutrophils. *GPI-202* was consistently and strongly expressed in cytotoxic cells (CD8+ T, $\gamma\delta$ T, and NK cells) and activated macrophages, but variable expression in other cell types. The difference between these two transcripts at the protein level is at the N-terminus. Although it is possible that there would be a potential difference in the enzymatic activity between these two isoforms, based on what is currently known about the effect of naturally occurring mutations, the cytokine activity is independent of enzymatic activity (417, 418). It has been shown that secreted GPI can bind to the AMFR extracellular receptor causing internalization, most work has been carried out within the context of solid tumors with its role within immune cells limited to our knowledge of its important role in glucose metabolism, which is nonetheless a critical factor in immune responses (419, 420).



Data summary for GPI gene. **A)** Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different GPI transcripts, identified in our data, and their respective protein isoforms. The arrowheads represent glucose-6-phosphate binding domains. GPI-201 codes for the canonical protein isoform.

LITAF

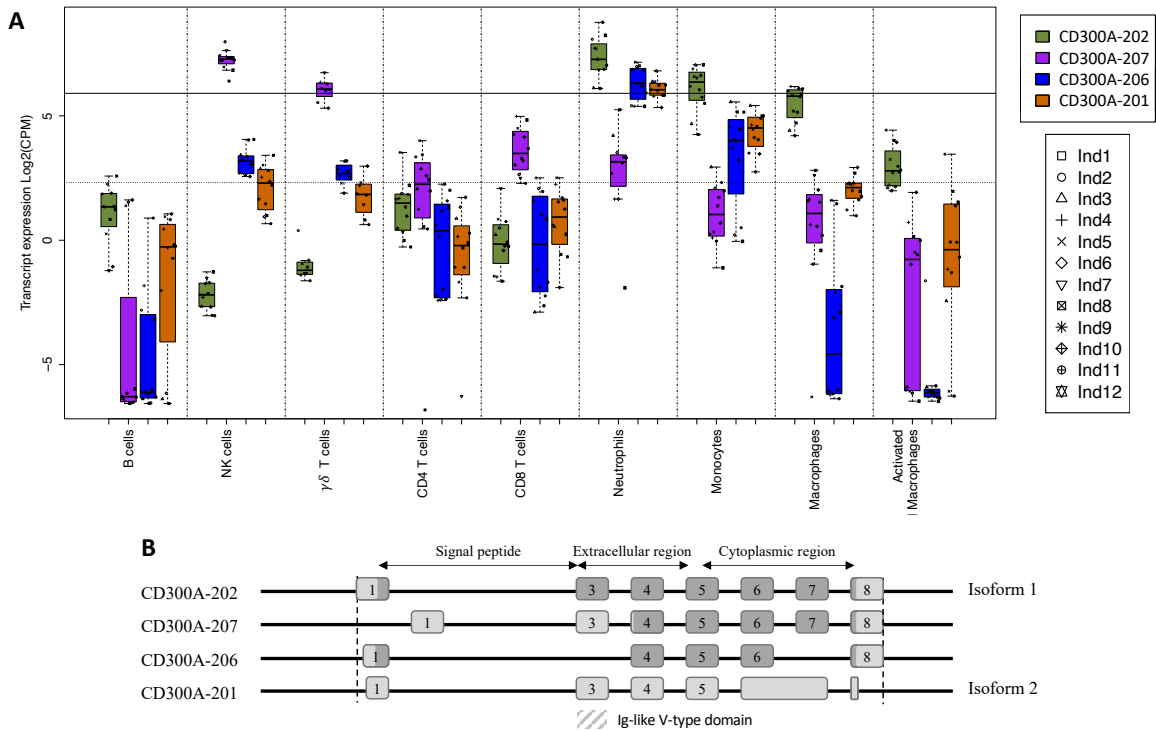
At the time of cloning of the *LITAF* (alias *SIMPLE*) gene, two RNA isoforms were detected by Northern Blot, however, differential tissue distribution of these have not been reported (421). In our analyses, we detect four different transcripts that encode a protein. Specifically, *LITAF-203* encodes protein “isoform 3”, whereas *LITAF-205*, *LITAF-221* and *LITAF-209* all encode the same protein “isoform 1”. The three transcripts that encode the isoform 1 have roughly similar expression across all cell types. *LITAF-203*, on the other hand, is essentially absent in B cells, but strongly expressed in all other cell types examined albeit, with great interindividual variability in CD8+ T cells and monocytes. The exon 4 in *LITAF-203* modifies the protein at C-terminal with an earlier stop codon and the protein isoform 3 encoded by *LITAF-203* loses its last 34 and 8 nucleic acids respectively in LITAF domain and in membrane-binding amphipathic helix. The function associated with isoform 3 is unknown but LITAF has been found to have multiple different functions including regulation of many cytokines as TNF, CCL2, CCL5, CXCL1, IL1A and IL10 (422). It is also present in late endosome/lysosome, co-localized with CD63. Some mutant forms of LITAF are responsible of Charcot-Marie-Tooth type 1C disease (CMT1C), a peripheral neuropathy. These mutants cluster within C-rich domain of the C-terminal portion, with this C-rich domain potentially being a RING finger domain with E3 ubiquitin-protein ligase activity. The mutant P135T affects LITAF subcellular localization, and this mutated LITAF is present in late endosome/lysosome but also in mitochondria (423-425).



Data summary for *LITAF* gene. **A)** Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *LITAF* transcripts, identified in our data, and their respective protein isoforms. The *LITAF* protein has a PSAP and a PPxY motifs, and a *LITAF* and a membrane-binding amphipathic helix domains. White spaces in *LITAF* and membrane-binding amphipathic helix domains correspond to coding regions of *LITAF-203* exon 4 that do not encode these domains. *LITAF-221* it codes for the canonical protein isoform.

CD300A

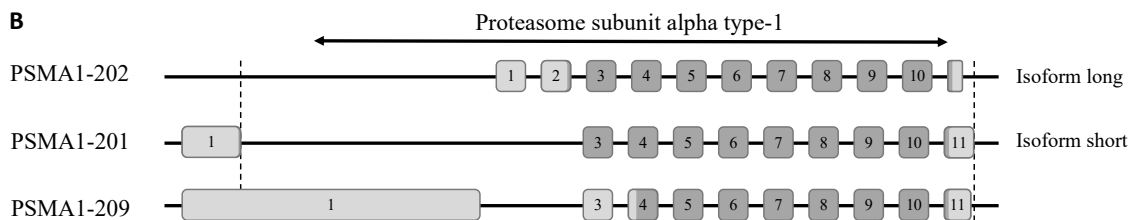
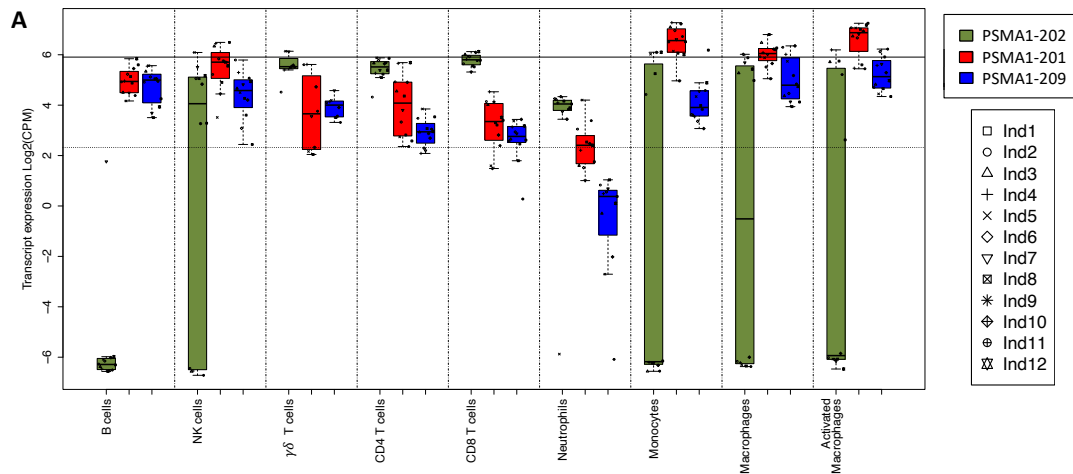
CD300A is part of the seven-member CD300 family of proteins that have multiple modulatory impacts on different cells of the immune system. Family members either have a short activating ITAM motif or an inhibitory ITIM motif within the cytoplasmic portion of the protein (426, 427). CD300A falls into the latter category and is known to modulate cellular responses to IL-2, TLR3 and TLR9 ligands, as well as to hypoxia (428). In the current study we observed that *CD300A-202* was strongly expressed in the myeloid lineage but decreased in level upon differentiation into macrophages and even more so in LPS-activated macrophages, and essentially absent in lymphocytes. *CD300A-207* was most strongly expressed in NK and $\gamma\delta$ T cells. Importantly, the protein encoded by the *CD300A-207* transcript is lacking the Ig-like V-type domain, which is responsible of the phospholipid binding capabilities of CD300A that has been demonstrated to mediate binding of apoptotic cells and certain viral pathogens (e.g.: Dengue virus, West Nile virus, yellow fever virus, Zika virus etc.), facilitating the phagocytosis of the former and recognition and entry of the latter (428, 429).



Data summary for *CD300A* gene. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *CD300A* transcripts, identified in our data, and their respective protein isoforms. The *CD300A* protein has a signal peptide region, an extracellular region, and a cytoplasmic region. Exon 5 codes to transmembrane region and exon 3 encodes the Ig-like V-type domain of the *CD300A* protein. *CD300A-202* codes for the canonical protein isoform.

PSMA1

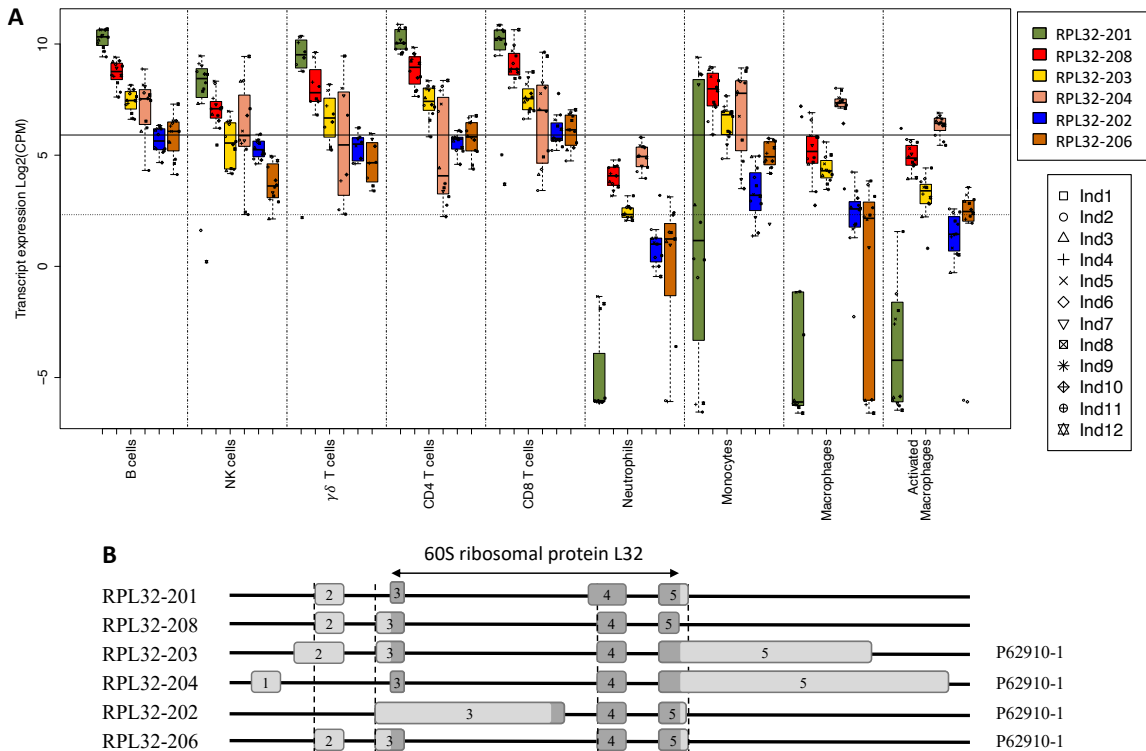
Proteasome 20S Subunit Alpha 1 (PSMA1) mediates ubiquitin-independent protein degradation and in the immunoproteasome is key to the processing of Class I MHC peptides. It is also involved in the regulation of multiple cytokines and has been shown to bind LPS (430-433). Identified in the comparison of monocytes to lymphocytes, this gene has three transcripts that are differentially expressed in our data. Indeed, the *PSMA-201* transcript is strongly expressed in monocytes and macrophages, whereas the *PSMA-202* is strongly expressed in most T cells ($\gamma\delta$ T, CD4+ T, CD8+ T). In B cells and NK cells, on the other hand, both the *PSMA-201* and *PSMA-209* transcripts are strongly expressed. The *PSMA-202*, *-201* and *-209* transcripts encode proteins of 269, 263, and 238 amino acids in length, respectively, and differ in their N-terminal sequence. The alpha subunit of the 20S proteasome is part of the N-terminal nucleophile (NTN-) hydrolase superfamily, a group of enzymes that are activated autocatalytically via an N-terminally located nucleophilic amino acid (434). It is therefore likely that differences in N-terminal sequence will affect the autocatalysis that is important for the formation of the alpha ring within the proteasome structure.



Data summary for *PSMA1* gene. **A)** Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *PSMA1* transcripts, identified in our data, and their respective protein isoforms. *PSMA1-202* codes for the canonical protein isoform.

RPL32

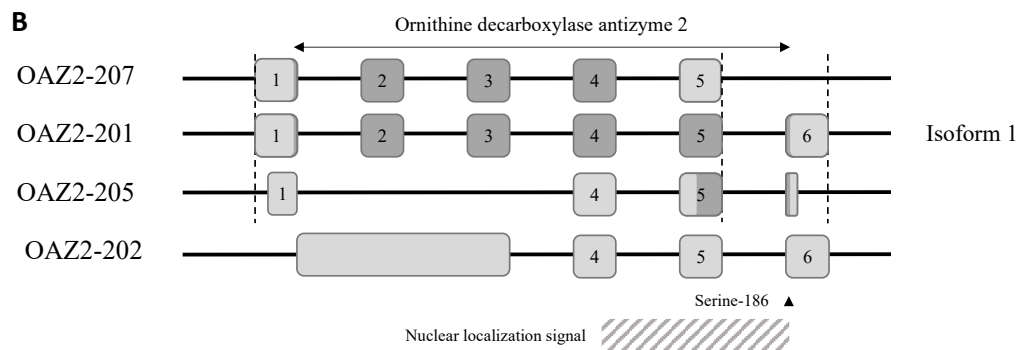
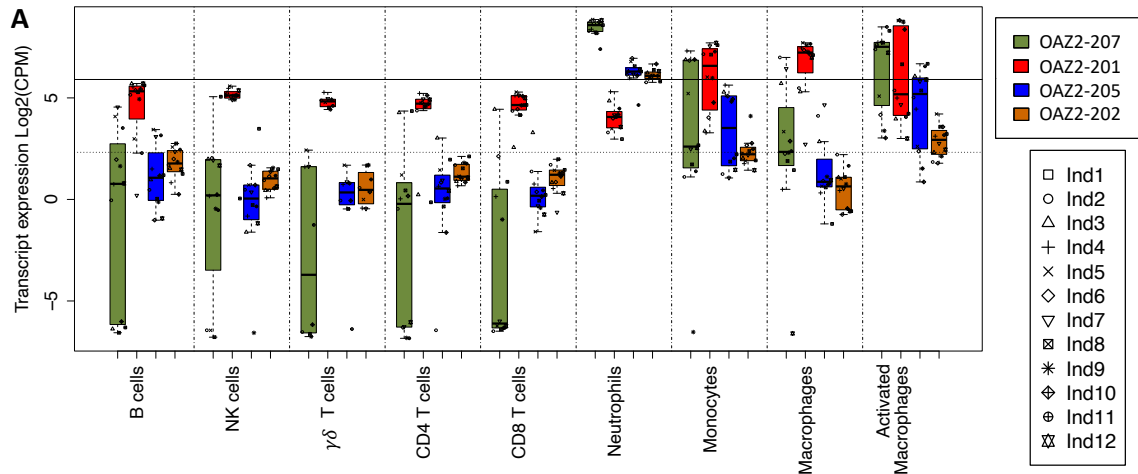
Seven different transcripts for Ribosomal Protein L32 (*RPL32*), differing in coding and non-coding sequences, were found to be expressed in our study. Transcripts *RPL32-201*, *RPL32-203*, and *RPL32-208* were strongly and consistently expressed across all lymphocytes. Although there was interindividual variation in the expression of *RPL32-201* in monocytes, it was essentially absent for the other myeloid populations. Transcripts *RPL32-201*, *RPL32-203*, and *RPL32-208* encode protein isoforms of 153, 135, 133 amino acids in length, respectively (**Table S2**). *RPL32-201* has an exon 4 more long than *RPL32-203* and *-208* and its code to a protein that has 18 amino acids at position 32 of *RPL32* protein. For transcript *RPL32-208*, it loses the last 2 amino acids of the canonical protein. While it is currently unknown what functional differences exist between these two protein isoforms, these may impact on its functions within the large 60S ribosomal subunit and/or potentially on immune signaling as it has previously been demonstrated for *RPL13A*, another component of the 60S subunit (435).



Data summary for RPL32 gene. **A)** Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different RPL32 transcripts, identified in our data, and their respective protein isoforms. Exons 3 to 5 encode for the RPL32 protein, a.k.a the 60S ribosomal protein L32. RPL32-206 codes for the canonical protein isoform.

OAZ2

Our analyses identified three protein-coding transcripts for the antizyme ornithine decarboxylase antizyme 2 (*OAZ2*) gene that were differentially expressed across immune cells. *OAZ2* is known to bind the monomer ornithine decarboxylase (ODC) to inhibit the assembling of functional ODC homodimers, with ODC known to be critical in the regulation of M1 macrophage activation in response to infection (436). This interaction accelerates the ODC degradation by nuclear proteasomes (437). *OAZ2* negatively regulates intracellular polyamine biosynthesis and extracellular polyamine uptake. Studies with cloned mouse *Oaz2* found that it is primarily localized in nucleus, with diffuse expression in both cytoplasm and nucleus, and that its C-terminal end is necessary for the nuclear distribution (437). In this same study it was reported that the *Oaz2* protein can be phosphorylate at its Ser-186 but the phosphorylated protein has the same function than non-phosphorylated protein (437). In our data, the full length *OAZ2-201* transcript is found in all cell types although most highly expressed in monocytes and macrophages, whereas the shorter *OAZ2-207* transcript is highly expressed in neutrophils and activated macrophages, although with modest expression in monocytes and unstimulated macrophages. *OAZ2-207* transcript differs from *OAZ2-201* in that it is lacking amino acids at position 130 to 189 and has distinct sequence at amino acids 119 to 129. Given this difference, the protein encoded by *OAZ2-207* is likely to primarily have non-nuclear localization.

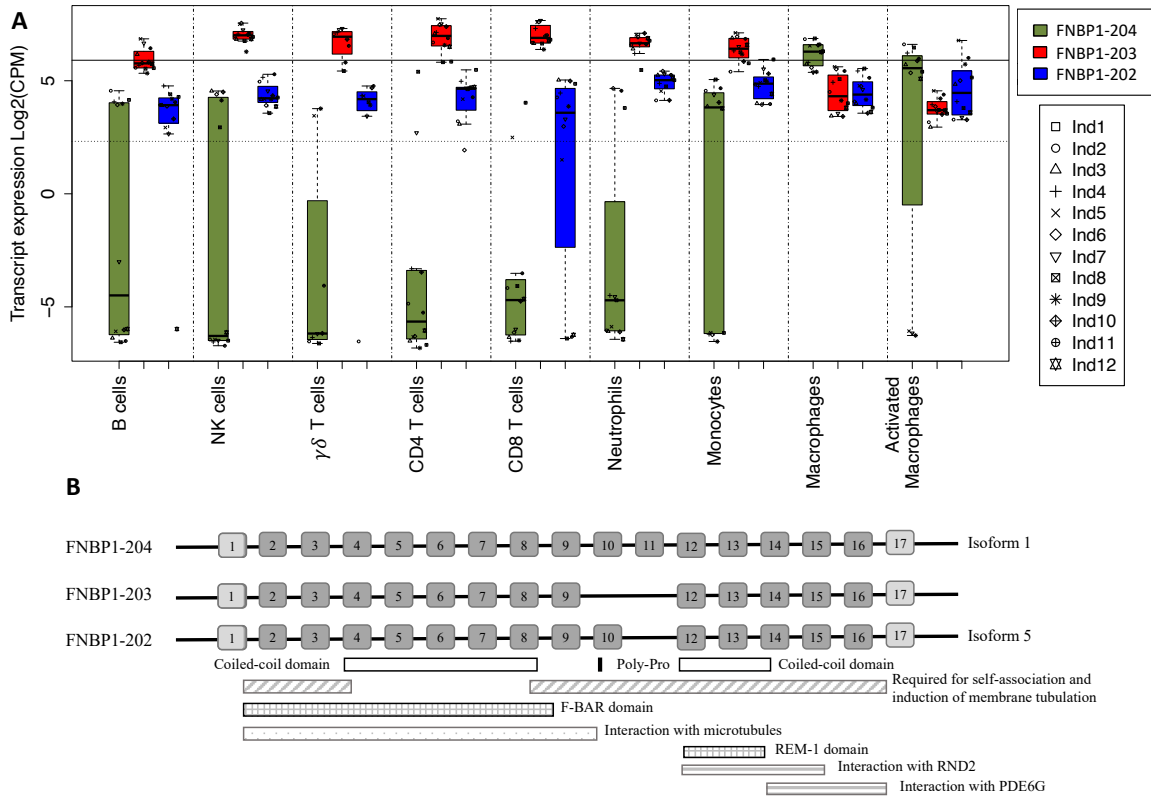


Data summary for OAZ2 gene. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different OAZ2 transcripts, identified in our data, and their respective protein isoforms. The OAZ2 protein as a Serine-186 and a nuclear localization signal (NLS) domain. The arrowhead represents the serine-186. The positions of Serine-186 and the NLS domain are obtained with the article (437). OAZ2-201 codes for the canonical protein isoform.

FNBP1

Formin binding protein 1 (*FNBP1*) alias *FBP17*, is a member of Toca family (TOCA1, FNBP1, and CIP4), a subgroup of the BAR domain superfamily that are known to be essential for N-WASP-mediated actin nucleation and membrane curvature (438). We detect three different transcripts in our data. *FNBP1-203* is the most abundant across the primary lymphoid and myeloid cells that we surveyed but decreases when monocytes are differentiated into macrophages. Conversely, *FNBP1-204*, was the predominant transcript in differentiated macrophages, which appears to be macrophage-specific; and *FNBP1-202* which appears to be myeloid-specific in our data. FNBP-202 was found in all cell types, albeit at modest levels. In terms of protein sequence, *FNBP1-203* encodes for the shortest protein, with -204 and -202 being longer due to additional amino acids in the region of the protein that has been shown to be required for self-assembly and for interaction with microtubules. This is also directly adjacent to the portion of the protein which is important for the interaction with numerous proteins including RND2 and has been proposed to be a link between signaling (e.g., by GTPases) and regulation of the actin cytoskeleton. In macrophages, FNBP1 recruits WASP, WIP and dynamin-2 proteins that form a complex involved in actin polymerization, podosome formation, phagocytic cups and this

complex is also implicated in both clathrin-dependent and -independent endocytosis (438-440). Indirectly, FNBP1 is responsible for chemotactic macrophage migration, adhesion structures and degradation of extracellular matrix (439, 441). It is highly likely that the shift in transcripts being expressed that we observed during differentiation and activation of the monocyte/macrophage lineage will impact on the efficiency of binding of the FNBP1-interaction proteins and these actin cytoskeleton functions.



Data summary for FNBP1 gene. A) Boxplots of the expression of alternative transcript isoforms in our 12 individuals in different immune cell populations (as described in **FYB1's figure**). **B)** Graphical representation (as in **FYB1's figure**) of the different *FNBP1* transcripts, identified in our data, and their respective protein isoforms. The FNBP1 protein has a Poly-Pro motif, two coiled-coil domain, two domain required for self-association and induction of membrane tubulation, a F-BAR domain, a domain that interacts with microtubule, a REM-1 domain, and two domains that interact with RDN2 and PDE6G, respectively. The domain that interacts with RND2 also binds PDE6G, TNKS, DNM1, DNM3, ARHGAP17, DAAM1, DIAPH1, and DIAPH2. *FNBP1-204* codes for the canonical protein isoform.

B.2. *Supplementary tables*

S3.1. Table **TableS1.xlsx file. Table S1:** Transcript expression and annotations in our RNA-Seq data.

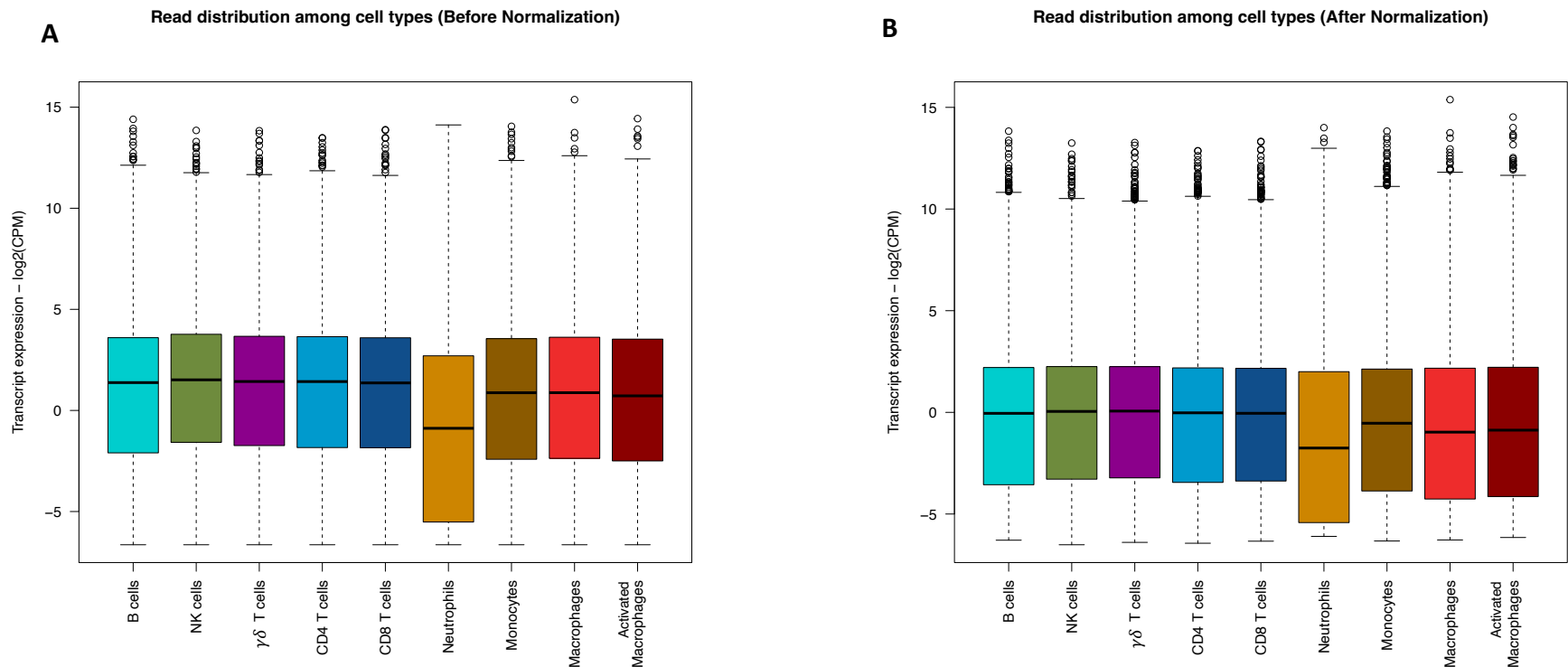
S3.2. Table **TableS2.xlsx file. Table S2:** Results of selected genes in lymphoid cell/monocyte analysis.

S3.3. Table Summaries of gene and transcript number for genes with at least two transcripts.

		Lymphoid cell types & monocytes	Monocytes, macrophages and activated macrophages
All genes	Gene Number	12,491	11,301
	Transcript Number	29,232	23,756
	Transcript encoding a protein isoform	21,555	19,497
Genes with at least two transcripts	Gene Number	6,985	5,773
	Transcript Number	23,736	18,228
	Transcript encoding a protein isoform	17,322	14,702

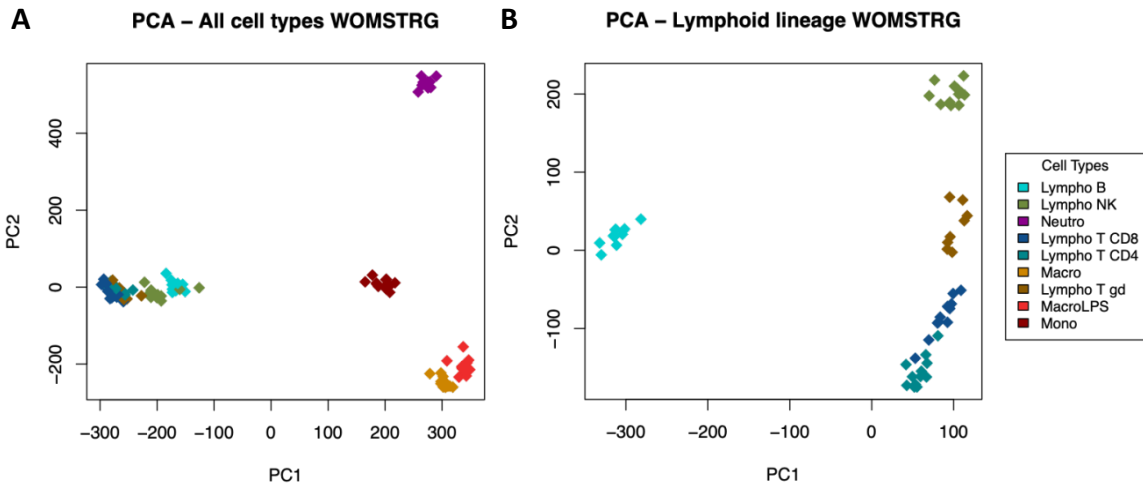
S3.4. Table **TableS4.xlsx file.** Results of selected genes in monocyte/macrophage lineage analysis.

B.3. Supplementary figures

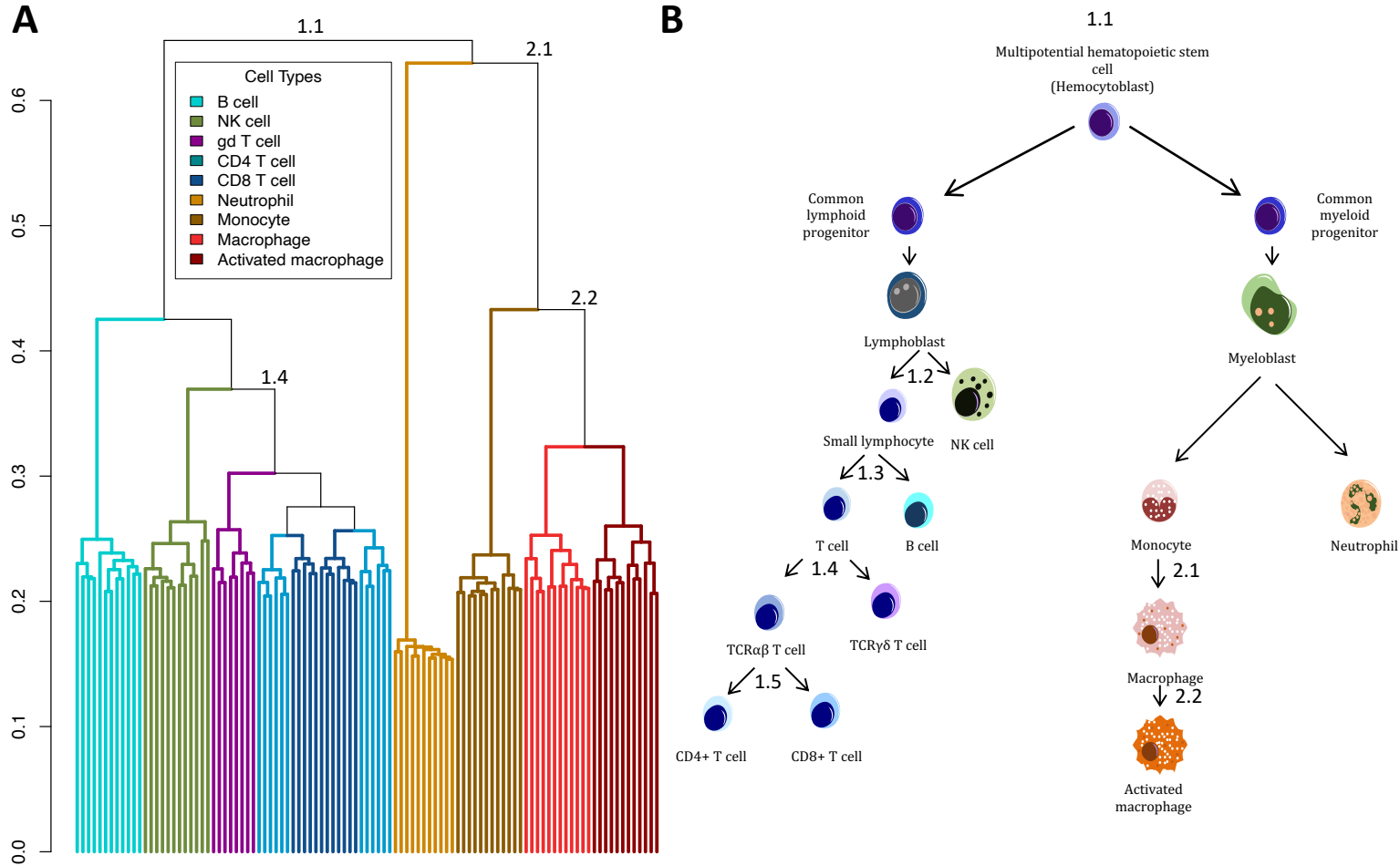


S3.1. Fig **Boxplots of transcript expression distribution before and after normalization.** Each boxplot represents the mean of transcript expression in log₂ for a cell type in our 12 individuals. **A)** Distribution of raw reads in log₂ of CPM. **B)** Distribution after normalization by library size and TMM normalization factor and variance stabilization with R function *voom*. The median

of neutrophil gene distribution is not aligned with other cell types because many more genes are not expressed in neutrophils than other primary immune cell types in our data (303).

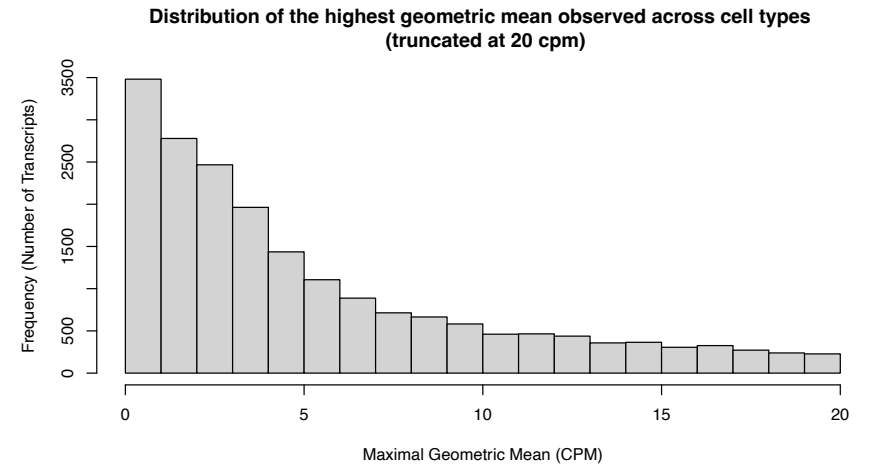
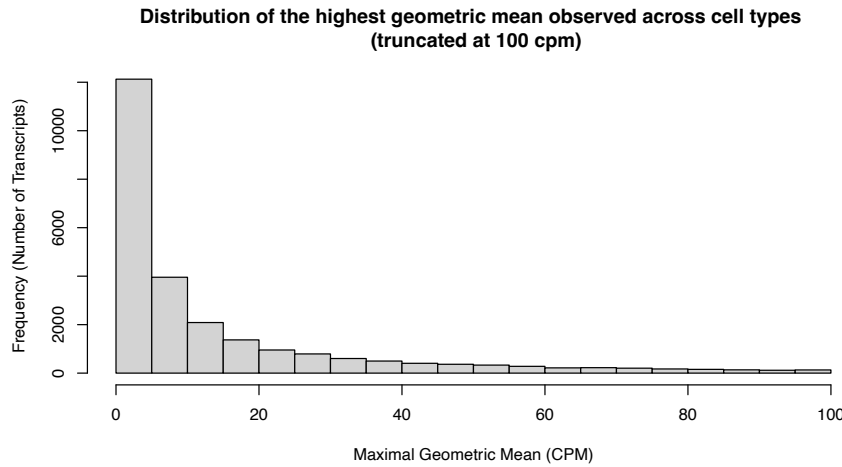
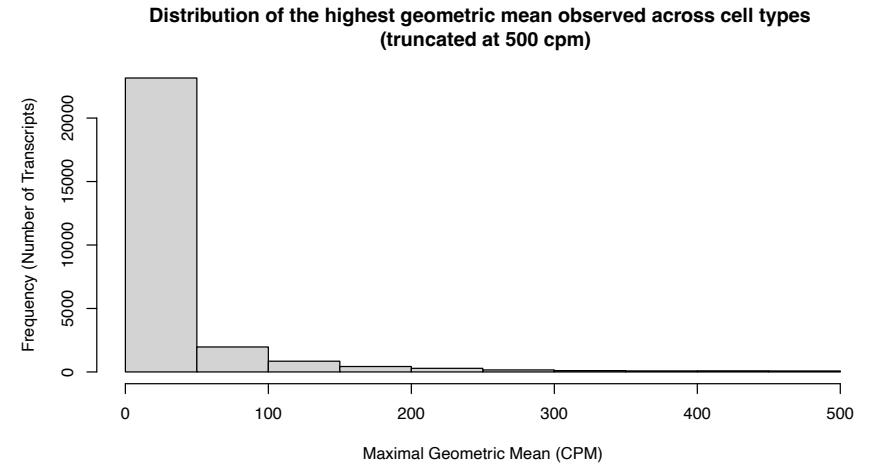
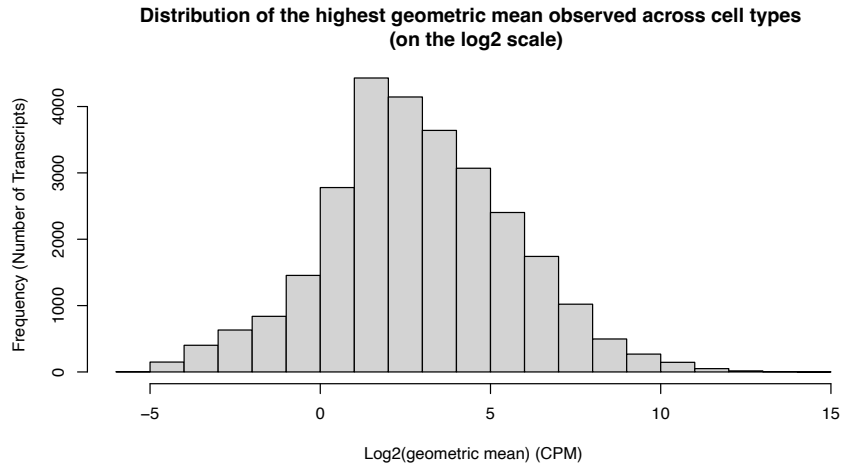


S3.2. Fig **PC analysis of transcriptomic data distinguishing immune cell types.** Representation of the two first principal components on the alternate transcript RNA-seq data from all individuals for all samples (**A**) and for lymphoid samples alone (**B**). The R function *prcomp* was used to perform principal component analysis, and each symbol represents an individual RNA-seq sample. The representation of the first principal components of transcript-level RNA-Seq data is similar to its representation of the first principal components of gene-level RNA-Seq data (54).

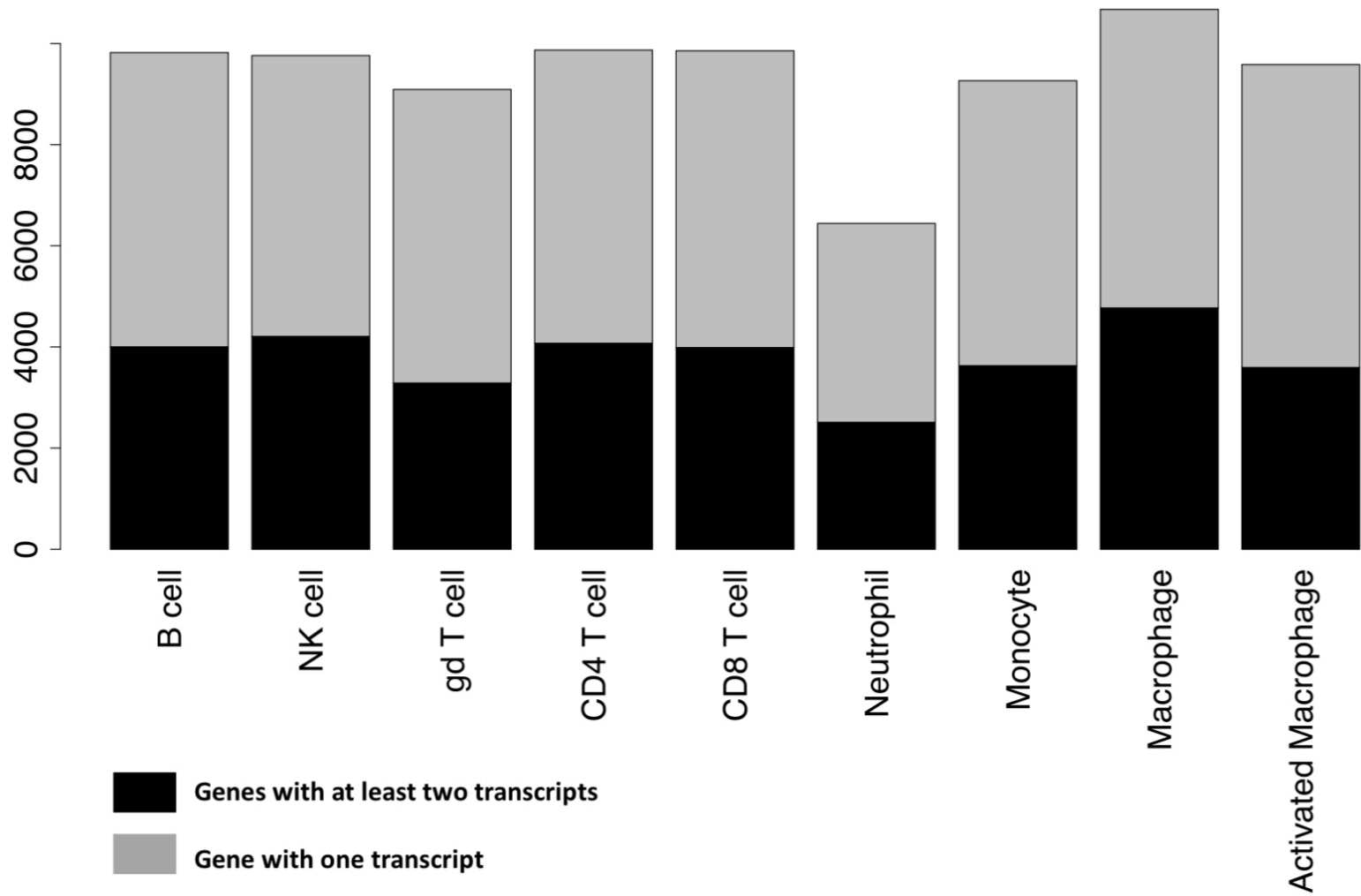


S3.3. Fig **Unsupervised hierarchical clustering of transcript RNA-seq data reflects hematopoietic differentiation.** **A)** A dendrogram representation of the unsupervised hierarchical clustering of the transcript-level RNA-seq data from across all individuals and cell types has a structure that closely resembles the known **B)** differentiation scheme of immune cells. Each dendrogram leaf represents a single RNA-Seq sample and Y-axis represents the distance based on Pearson correlation and the Arabic numerals represent some of the comparison branchpoint apply in our study. Black arrows represent a decision in the known differentiation scheme of immune cells and the Arabic numerals identify the comparison branchpoint apply in our

study. The dendrogram representation of the unsupervised hierarchical clustering of gene-level RNA-seq data is similar to this representation except for CD4 and CD8 T cells (54).



S3.4. Fig **Distribution of the highest geometric mean observed across cell type.** The maximum geometric mean (CPM) observed across the nine cell types is represented at different scales. Top-left histogram in the log2 scale to represent the full range of data, while the others are at the original scale and truncated to better illustrate the distribution at lower-moderate expression level.



S3.5. Fig **Transcript isoform distribution in our data by cell types.** Representation of two categories of genes: genes with one transcript in grey and genes with at least two transcripts in each cell types, in black.

B.4. *Supplementary references*

Uncategorized References

1. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics*. 2006;7:115. Epub 2006/05/18.
2. Gazit R, Garrison BS, Rao TN, Shay T, Costello J, Ericson J, et al. Transcriptome analysis identifies regulators of hematopoietic stem and progenitor cells. *Stem Cell Reports*. 2013;1(3):266-80. Epub 2013/12/10.
3. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015;348(6235):660-5. Epub 2015/05/09.
4. Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science*. 2015;350(6264):1092-6. Epub 2015/10/17.
5. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, et al. Identification and characterization of essential genes in the human genome. *Science*. 2015;350(6264):1096-101. Epub 2015/10/17.
6. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. 2015;163(6):1515-26. Epub 2015/12/03.
7. Boller S, Grosschedl R. The regulatory network of B-cell differentiation: a focused view of early B-cell factor 1 function. *Immunol Rev*. 2014;261(1):102-15. Epub 2014/08/16.
8. Lindner JM, Wong CS, Moller A, Nielsen PJ. A C-terminal acidic domain regulates degradation of the transcriptional coactivator Bob1. *Mol Cell Biol*. 2013;33(23):4628-40. Epub 2013/09/26.
9. Corcoran L, Emslie D, Kratina T, Shi W, Hirsch S, Taubenheim N, et al. Oct2 and Obf1 as Facilitators of B:T Cell Collaboration during a Humoral Immune Response. *Front Immunol*. 2014;5:108. Epub 2014/04/02.
10. Bordon A, Bosco N, Du Roure C, Bartholdy B, Kohler H, Matthias G, et al. Enforced expression of the transcriptional coactivator OBF1 impairs B cell differentiation at the earliest stage of development. *PLoS One*. 2008;3(12):e4007. Epub 2008/12/24.
11. Medvedovic J, Ebert A, Tagoh H, Busslinger M. Pax5: a master regulator of B cell development and leukemogenesis. *Adv Immunol*. 2011;111:179-206. Epub 2011/10/06.
12. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: the guardian of B cell identity and function. *Nat Immunol*. 2007;8(5):463-70. Epub 2007/04/19.
13. Mendez A, Mendoza L. A Network Model to Describe the Terminal Differentiation of B Cells. *PLoS Comput Biol*. 2016;12(1):e1004696. Epub 2016/01/12.
14. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, et al. MEF2B mutations lead to deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. *Nat Immunol*. 2013;14(10):1084-92. Epub 2013/08/27.
15. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*. 2006;7:29-59. Epub 2006/05/25.

16. Danino YM, Even D, Ideses D, Juven-Gershon T. The core promoter: At the heart of gene expression. *Biochim Biophys Acta*. 2015;1849(8):1116-31. Epub 2015/05/03.
17. Pang B, Snyder MP. Systematic identification of silencers in human cells. *Nat Genet*. 2020;52(3):254-63. Epub 2020/02/26.
18. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015;16(3):144-54. Epub 2015/02/05.
19. Pott S, Lieb JD. What are super-enhancers? *Nat Genet*. 2015;47(1):8-12. Epub 2014/12/31.
20. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307-19. Epub 2013/04/16.
21. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155(4):934-47. Epub 2013/10/15.
22. Onufriev AV, Schiessel H. The nucleosome: from structure to function through physics. *Curr Opin Struct Biol*. 2019;56:119-30. Epub 2019/02/03.
23. McGinty RK, Tan S. Nucleosome structure and function. *Chem Rev*. 2015;115(6):2255-73. Epub 2014/12/17.
24. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev*. 2011;25(10):1010-22. Epub 2011/05/18.
25. Vavouri T, Lehner B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol*. 2012;13(11):R110. Epub 2012/11/29.
26. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. *Cell*. 2018;172(4):650-65. Epub 2018/02/10.
27. Zhao Y, Garcia BA. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb Perspect Biol*. 2015;7(9):a025064. Epub 2015/09/04.
28. Stillman B. Histone Modifications: Insights into Their Influence on Gene Expression. *Cell*. 2018;175(1):6-9. Epub 2018/09/16.
29. Stryer L. *Biochemistry*. Freeman and Company^e éd. New York - Basingstoke: W. H. : Freeman and Company; 1995.
30. Eggeling R. Disentangling transcription factor binding site complexity. *Nucleic Acids Res*. 2018;46(20):e121. Epub 2018/08/08.
31. Lis M, Walther D. Erratum to: "The orientation of transcription factor binding site motifs in gene promoter regions: does it matter?". *BMC Genomics*. 2016;17:310. Epub 2016/04/30.
32. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012;13(9):613-26. Epub 2012/08/08.
33. Ramanouskaya TV, Grinev VV. The determinants of alternative RNA splicing in human cells. *Mol Genet Genomics*. 2017;292(6):1175-95. Epub 2017/07/15.
34. Mozaffari-Jovin S, Wandersleben T, Santos KF, Will CL, Luhrmann R, Wahl MC. Novel regulatory principles of the spliceosomal Brr2 RNA helicase and links to retinal disease in humans. *RNA Biol*. 2014;11(4):298-312. Epub 2014/03/20.
35. Dredge BK, Stefani G, Engelhard CC, Darnell RB. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J*. 2005;24(8):1608-20. Epub 2005/06/04.
36. Sanidas I, Polytarchou C, Hatzia Apostolou M, Ezell SA, Kottakis F, Hu L, et al. Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Mol Cell*. 2014;53(4):577-90. Epub 2014/01/28.

37. Rahhal R, Seto E. Emerging roles of histone modifications and HDACs in RNA splicing. *Nucleic Acids Res.* 2019;47(10):4911-26. Epub 2019/06/05.
38. Davis-Turak JC, Allison K, Shokhirev MN, Ponomarenko P, Tsimring LS, Glass CK, et al. Considering the kinetics of mRNA synthesis in the analysis of the genome and epigenome reveals determinants of co-transcriptional splicing. *Nucleic Acids Res.* 2015;43(2):699-707. Epub 2014/12/30.
39. Dever TE, Green R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol.* 2012;4(7):a013706. Epub 2012/07/04.
40. Merrick WC, Pavitt GD. Protein Synthesis Initiation in Eukaryotic Cells. *Cold Spring Harb Perspect Biol.* 2018;10(12). Epub 2018/05/08.
41. Mace K, Giudice E, Gillet R. [Protein synthesis by the ribosome: a pathway full of pitfalls]. *Med Sci (Paris).* 2015;31(3):282-90. Epub 2015/04/10. La synthèse des protéines par le ribosome - Un chemin semé d'embûches.
42. Yael Avissar (Cell Biology) RIC, Jung Choi (Genetics) GIoT, Jean DeSaix (Evolution) UoNCaCH, Vladimir Jurukovski (Animal Physiology) SCCC, Robert Wise (Plant Biology) UoW, Oshkosh, Connie Rye (General Content Lead) EMCC. *Biology: OpenStax*; 2016.
43. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, et al. Function of alternative splicing. *Gene.* 2013;514(1):1-30. Epub 2012/08/23.
44. Aguiar D, Cheng LF, Dumitrascu B, Mordelet F, Pai AA, Engelhardt BE. Bayesian nonparametric discovery of isoforms and individual specific quantification. *Nat Commun.* 2018;9(1):1681. Epub 2018/04/29.
45. Yang X, Coulombe-Huntington J, Kang S, Sheynkman GM, Hao T, Richardson A, et al. Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell.* 2016;164(4):805-17. Epub 2016/02/13.
46. Stastna M, Van Eyk JE. Analysis of protein isoforms: can we do it better? *Proteomics.* 2012;12(19-20):2937-48. Epub 2012/08/14.
47. Schwerk C, Schulze-Osthoff K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol Cell.* 2005;19(1):1-13. Epub 2005/07/02.
48. Tanuma N, Nakamura K, Shima H, Kikuchi K. Protein-tyrosine phosphatase PTPepsilon C inhibits Jak-STAT signaling and differentiation induced by interleukin-6 and leukemia inhibitory factor in M1 leukemia cells. *J Biol Chem.* 2000;275(36):28216-21. Epub 2000/06/22.
49. Lapointe F, Turcotte S, Roy J, Bissonnette E, Rola-Pleszczynski M, Stankova J. RPTepsilon promotes M2-polarized macrophage migration through ROCK2 signaling and podosome formation. *J Cell Sci.* 2020;133(5). Epub 2019/11/15.
50. Chaplin DD. Overview of the immune response. *J Allergy Clin Immunol.* 2010;125(2 Suppl 2):S3-23. Epub 2010/03/05.
51. Parkin J, Cohen B. An overview of the immune system. *Lancet.* 2001;357(9270):1777-89. Epub 2001/06/14.
52. Yatim KM, Lakkis FG. A brief journey through the immune system. *Clin J Am Soc Nephrol.* 2015;10(7):1274-81. Epub 2015/04/08.
53. Cruz-Migoni S, Caamano J. Fat-Associated Lymphoid Clusters in Inflammation and Immunity. *Front Immunol.* 2016;7:612. Epub 2017/01/10.
54. Mola S, Foisy S, Boucher G, Major F, Beauchamp C, Karaky M, et al. A transcriptome-based approach to identify functional modules within and across primary human immune cells. *PLoS One.* 2020;15(5):e0233543. Epub 2020/05/30.
55. Labrecque N, Cermakian N. Circadian Clocks in the Immune System. *J Biol Rhythms.* 2015;30(4):277-90. Epub 2015/04/23.

56. Dogra P, Rancan C, Ma W, Toth M, Senda T, Carpenter DJ, et al. Tissue Determinants of Human NK Cell Development, Function, and Residence. *Cell*. 2020;180(4):749-63 e13. Epub 2020/02/16.
57. Janeway C. *Immunobiology: The Immune System in Health and Disease*. 5th ed.^e éd. Garland Science: ew York : Garland Pub., c2001; 2001.
58. Paul S, Shilpi, Lal G. Role of gamma-delta (gammadelta) T cells in autoimmunity. *J Leukoc Biol*. 2015;97(2):259-71. Epub 2014/12/17.
59. Peters C, Kabelitz D, Wesch D. Regulatory functions of gammadelta T cells. *Cell Mol Life Sci*. 2018;75(12):2125-35. Epub 2018/03/10.
60. He Y, Wu K, Hu Y, Sheng L, Tie R, Wang B, et al. gammadelta T cell and other immune cells crosstalk in cellular immunity. *J Immunol Res*. 2014;2014:960252. Epub 2014/04/18.
61. Hou L, Wang T, Sun J. gammadelta T cells in infection and autoimmunity. *Int Immunopharmacol*. 2015;28(2):887-91. Epub 2015/04/14.
62. Stegelmeier AA, van Vloten JP, Mould RC, Klafuric EM, Minott JA, Wootton SK, et al. Myeloid Cells during Viral Infections and Inflammation. *Viruses*. 2019;11(2). Epub 2019/02/23.
63. Commins SP, Borish L, Steinke JW. Immunologic messenger molecules: cytokines, interferons, and chemokines. *J Allergy Clin Immunol*. 2010;125(2 Suppl 2):S53-72. Epub 2009/11/26.
64. Lunney JK. Cytokines orchestrating the immune response. *Rev Sci Tech*. 1998;17(1):84-94. Epub 1998/06/25.
65. Sokol CL, Luster AD. The chemokine system in innate immunity. *Cold Spring Harb Perspect Biol*. 2015;7(5). Epub 2015/01/31.
66. Arango Duque G, Descoteaux A. Macrophage cytokines: involvement in immunity and infectious diseases. *Front Immunol*. 2014;5:491. Epub 2014/10/24.
67. Maghazachi AA, Al-Aoukaty A, Schall TJ. CC chemokines induce the generation of killer cells from CD56+ cells. *Eur J Immunol*. 1996;26(2):315-9. Epub 1996/02/01.
68. Jones J, Stevens CE, Rubinstein P, Robertazzi RR, Kerr A, Cabbad MF. Obstetric predictors of placental/umbilical cord blood volume for transplantation. *Am J Obstet Gynecol*. 2003;188(2):503-9. Epub 2003/02/20.
69. Tecchio C, Cassatella MA. Neutrophil-derived chemokines on the road to immunity. *Semin Immunol*. 2016;28(2):119-28. Epub 2016/05/07.
70. Griffith JW, Sokol CL, Luster AD. Chemokines and chemokine receptors: positioning cells for host defense and immunity. *Annu Rev Immunol*. 2014;32:659-702. Epub 2014/03/25.
71. Vignola AM, Gjomarkaj M, Arnoux B, Bousquet J. Monocytes. *J Allergy Clin Immunol*. 1998;101(2 Pt 1):149-52. Epub 1998/03/21.
72. Gann AW, Amoroso JW, Einck VJ, Rice WP, Chambers JJ, Schnarr NA. A photoinduced, benzyne click reaction. *Org Lett*. 2014;16(7):2003-5. Epub 2014/03/15.
73. Cipriani B, Borsellino G, Poccia F, Placido R, Tramonti D, Bach S, et al. Activation of C-C beta-chemokines in human peripheral blood gammadelta T cells by isopentenyl pyrophosphate and regulation by cytokines. *Blood*. 2000;95(1):39-47. Epub 1999/12/23.
74. Borish LC, Steinke JW. 2. Cytokines and chemokines. *J Allergy Clin Immunol*. 2003;111(2 Suppl):S460-75. Epub 2003/02/20.
75. Xin P, Xu X, Deng C, Liu S, Wang Y, Zhou X, et al. The role of JAK/STAT signaling pathway and its inhibitors in diseases. *Int Immunopharmacol*. 2020;80:106210. Epub 2020/01/24.
76. Morris R, Kershaw NJ, Babon JJ. The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Sci*. 2018;27(12):1984-2009. Epub 2018/09/30.

77. Engel P, Boumsell L, Balderas R, Bensussan A, Gattei V, Horejsi V, et al. CD Nomenclature 2015: Human Leukocyte Differentiation Antigen Workshops as a Driving Force in Immunology. *J Immunol.* 2015;195(10):4555-63. Epub 2015/11/08.
78. Clark G, Stockinger H, Balderas R, van Zelm MC, Zola H, Hart D, et al. Nomenclature of CD molecules from the Tenth Human Leukocyte Differentiation Antigen Workshop. *Clin Transl Immunology.* 2016;5(1):e57. Epub 2016/02/24.
79. Gabius HJ, Kaltner H, Kopitz J, Andre S. The glycobiology of the CD system: a dictionary for translating marker designations into glycan/lectin structure and function. *Trends Biochem Sci.* 2015;40(7):360-76. Epub 2015/05/20.
80. Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project. *Nat Rev Immunol.* 2012;12(3):191-200. Epub 2012/02/22.
81. Orru V, Steri M, Sole G, Sidore C, Viridis F, Dei M, et al. Genetic variants regulating immune cell levels in health and disease. *Cell.* 2013;155(1):242-56. Epub 2013/10/01.
82. Lakschevitz FS, Hassanpour S, Rubin A, Fine N, Sun C, Glogauer M. Identification of neutrophil surface marker changes in health and inflammation using high-throughput screening flow cytometry. *Exp Cell Res.* 2016;342(2):200-9. Epub 2016/03/13.
83. Scapini P, Marini O, Tecchio C, Cassatella MA. Human neutrophils in the saga of cellular heterogeneity: insights and open questions. *Immunol Rev.* 2016;273(1):48-60. Epub 2016/08/26.
84. Depince-Berger AE, Aanei C, Iobagiu C, Jeraiby M, Lambert C. New tools in cytometry. *Morphologie.* 2016;100(331):199-209. Epub 2016/10/30.
85. McKinnon KM. Flow Cytometry: An Overview. *Curr Protoc Immunol.* 2018;120:5 1 -5 1 11. Epub 2018/03/08.
86. Choo SY. The HLA system: genetics, immunology, clinical testing, and clinical implications. *Yonsei Med J.* 2007;48(1):11-23. Epub 2007/02/28.
87. da Silva MB, da Cunha FF, Terra FF, Camara NO. Old game, new players: Linking classical theories to new trends in transplant immunology. *World J Transplant.* 2017;7(1):1-25. Epub 2017/03/11.
88. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* 2016;37(11):724-37. Epub 2016/09/12.
89. Leon Machado JA, Steimle V. The MHC Class II Transactivator CIITA: Not (Quite) the Odd-One-Out Anymore among NLR Proteins. *Int J Mol Sci.* 2021;22(3). Epub 2021/01/28.
90. Jongsma MLM, Guarda G, Spaapen RM. The regulatory network behind MHC class I expression. *Mol Immunol.* 2019;113:16-21. Epub 2017/12/12.
91. Reith W, LeibundGut-Landmann S, Waldburger JM. Regulation of MHC class II gene expression by the class II transactivator. *Nat Rev Immunol.* 2005;5(10):793-806. Epub 2005/10/04.
92. Krawczyk M, Reith W. Regulation of MHC class II expression, a unique regulatory system identified by the study of a primary immunodeficiency disease. *Tissue Antigens.* 2006;67(3):183-97. Epub 2006/04/01.
93. Bahram S, Bresnahan M, Geraghty DE, Spies T. A second lineage of mammalian major histocompatibility complex class I genes. *Proc Natl Acad Sci U S A.* 1994;91(14):6259-63. Epub 1994/07/05.
94. Wu J, Groh V, Spies T. T cell antigen receptor engagement and specificity in the recognition of stress-inducible MHC class I-related chains by human epithelial gamma delta T cells. *J Immunol.* 2002;169(3):1236-40. Epub 2002/07/23.

95. Cox ST, Madrigal JA, Saudemont A. Diversity and characterization of polymorphic 5' promoter haplotypes of MICA and MICB genes. *Tissue Antigens*. 2014;84(3):293-303. Epub 2014/06/26.
96. Cox ST, Hernandez D, Danby R, Turner TR, Madrigal JA. Diversity and characterisation of polymorphic 3' untranslated region haplotypes of MICA and MICB genes. *HLA*. 2018;92(6):392-402. Epub 2018/11/25.
97. Qian M, Geng J, Luo K, Huang Z, Zhang Q, Zhang JA, et al. BCL11B regulates MICA/B-mediated immune response by acting as a competitive endogenous RNA. *Oncogene*. 2020;39(7):1514-26. Epub 2019/11/02.
98. Van Rhijn I, Moody DB. CD1 and mycobacterial lipids activate human T cells. *Immunol Rev*. 2015;264(1):138-53. Epub 2015/02/24.
99. Chancellor A, Gadola SD, Mansour S. The versatility of the CD1 lipid antigen presentation pathway. *Immunology*. 2018;154(2):196-203. Epub 2018/02/21.
100. Kaczmarek R, Pasciak M, Szymczak-Kulus K, Czerwinski M. CD1: A Singed Cat of the Three Antigen Presentation Systems. *Arch Immunol Ther Exp (Warsz)*. 2017;65(3):201-14. Epub 2017/04/08.
101. Moody DB, Cotton RN. Four pathways of CD1 antigen presentation to T cells. *Curr Opin Immunol*. 2017;46:127-33. Epub 2017/08/02.
102. Eckhardt E, Bastian M. Animal models for human group 1 CD1 protein function. *Mol Immunol*. 2021;130:159-63. Epub 2021/01/02.
103. Hauptmann G, Bahram S. Genetics of the central MHC. *Curr Opin Immunol*. 2004;16(5):668-72. Epub 2004/09/03.
104. Colten HR. Expression of the MHC class III genes. *Philos Trans R Soc Lond B Biol Sci*. 1984;306(1129):355-66. Epub 1984/09/06.
105. Ling M, Murali M. Analysis of the Complement System in the Clinical Immunology Laboratory. *Clin Lab Med*. 2019;39(4):579-90. Epub 2019/11/02.
106. Belizario JE, Neyra JM, Setubal Destro Rodrigues MF. When and how NK cell-induced programmed cell death benefits immunological protection against intracellular pathogen infection. *Innate Immun*. 2018;24(8):452-65. Epub 2018/09/22.
107. Lanier LL. NKG2D Receptor and Its Ligands in Host Defense. *Cancer Immunol Res*. 2015;3(6):575-82. Epub 2015/06/05.
108. Jiang J, Natarajan K, Margulies DH. MHC Molecules, T cell Receptors, Natural Killer Cell Receptors, and Viral Immuno-evasins-Key Elements of Adaptive and Innate Immunity. *Adv Exp Med Biol*. 2019;1172:21-62. Epub 2019/10/20.
109. Dotiwala F, Lieberman J. Granulysin: killer lymphocyte safeguard against microbes. *Curr Opin Immunol*. 2019;60:19-29. Epub 2019/05/22.
110. de Saint Basile G, Menasche G, Fischer A. Molecular mechanisms of biogenesis and exocytosis of cytotoxic granules. *Nat Rev Immunol*. 2010;10(8):568-79. Epub 2010/07/17.
111. Osinska I, Popko K, Demkow U. Perforin: an important player in immune response. *Cent Eur J Immunol*. 2014;39(1):109-15. Epub 2014/01/01.
112. Tanaka S, Baba Y. B Cell Receptor Signaling. *Adv Exp Med Biol*. 2020;1254:23-36. Epub 2020/04/24.
113. Smulski CR, Eibel H. BAFF and BAFF-Receptor in B Cell Selection and Survival. *Front Immunol*. 2018;9:2285. Epub 2018/10/24.
114. Cowland JB, Borregaard N. Granulopoiesis and granules of human neutrophils. *Immunol Rev*. 2016;273(1):11-28. Epub 2016/08/26.

115. Liew PX, Kubes P. The Neutrophil's Role During Health and Disease. *Physiol Rev.* 2019;99(2):1223-48. Epub 2019/02/14.
116. Mortaz E, Alipoor SD, Adcock IM, Mumby S, Koenderman L. Update on Neutrophil Function in Severe Inflammation. *Front Immunol.* 2018;9:2171. Epub 2018/10/26.
117. Ravindran M, Khan MA, Palaniyar N. Neutrophil Extracellular Trap Formation: Physiology, Pathology, and Pharmacology. *Biomolecules.* 2019;9(8). Epub 2019/08/17.
118. Winterbourn CC, Kettle AJ, Hampton MB. Reactive Oxygen Species and Neutrophil Function. *Annu Rev Biochem.* 2016;85:765-92. Epub 2016/04/07.
119. Gao H, Dai W, Zhao L, Min J, Wang F. The Role of Zinc and Zinc Homeostasis in Macrophage Function. *J Immunol Res.* 2018;2018:6872621. Epub 2019/01/10.
120. Maares M, Haase H. Zinc and immunity: An essential interrelation. *Arch Biochem Biophys.* 2016;611:58-65. Epub 2016/03/30.
121. Freeman SA, Grinstein S. Phagocytosis: receptors, signal integration, and the cytoskeleton. *Immunol Rev.* 2014;262(1):193-215. Epub 2014/10/17.
122. Jain N, Moeller J, Vogel V. Mechanobiology of Macrophages: How Physical Factors Coregulate Macrophage Plasticity and Phagocytosis. *Annu Rev Biomed Eng.* 2019;21:267-97. Epub 2019/06/06.
123. Schumann J. It is all about fluidity: Fatty acids and macrophage phagocytosis. *Eur J Pharmacol.* 2016;785:18-23. Epub 2015/05/20.
124. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013;14(4):R36. Epub 2013/04/27.
125. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323. Epub 2011/08/06.
126. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* 2012;22(10):2008-17. Epub 2012/06/23.
127. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290-5. Epub 2015/02/19.
128. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 2016;11(9):1650-67. Epub 2016/08/26.
129. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017;14(4):417-9. Epub 2017/03/07.
130. Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotechnol.* 2014;32(5):462-4. Epub 2014/04/23.
131. Kim H, Bi Y, Pal S, Gupta R, Davuluri RV. IsoformEx: isoform level gene expression estimation using weighted non-negative least squares from mRNA-Seq data. *BMC Bioinformatics.* 2011;12:305. Epub 2011/07/29.
132. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559. Epub 2008/12/31.
133. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A.* 2014;111(51):E5593-601. Epub 2014/12/07.

134. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N. Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr Protoc Hum Genet.* 2015;87:11 6 1-6 4. Epub 2015/10/07.
135. Katz Y, Wang ET, Airoidi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods.* 2010;7(12):1009-15. Epub 2010/11/09.
136. Wu J, Akerman M, Sun S, McCombie WR, Krainer AR, Zhang MQ. SpliceTrap: a method to quantify alternative splicing under single cellular conditions. *Bioinformatics.* 2011;27(21):3010-6. Epub 2011/09/08.
137. Middleton R, Gao D, Thomas A, Singh B, Au A, Wong JJ, et al. IRFinder: assessing the impact of intron retention on mammalian gene expression. *Genome Biol.* 2017;18(1):51. Epub 2017/03/17.
138. Ergun A, Doran G, Costello JC, Paik HH, Collins JJ, Mathis D, et al. Differential splicing across immune system lineages. *Proc Natl Acad Sci U S A.* 2013;110(35):14324-9. Epub 2013/08/13.
139. Schaub A, Glasmacher E. Splicing in immune cells-mechanistic insights and emerging topics. *Int Immunol.* 2017;29(4):173-81. Epub 2017/05/13.
140. Byun S, Han S, Zheng Y, Planelles V, Lee Y. The landscape of alternative splicing in HIV-1 infected CD4 T-cells. *BMC Med Genomics.* 2020;13(Suppl 5):38. Epub 2020/04/04.
141. Ullrich S, Guigo R. Dynamic changes in intron retention are tightly associated with regulation of splicing factors and proliferative activity during B-cell development. *Nucleic Acids Res.* 2020;48(3):1327-40. Epub 2019/12/28.
142. West KO, Scott HM, Torres-Odio S, West AP, Patrick KL, Watson RO. The Splicing Factor hnRNP M Is a Critical Regulator of Innate Immune Gene Expression in Macrophages. *Cell Rep.* 2019;29(6):1594-609 e5. Epub 2019/11/07.
143. Rotival M, Quach H, Quintana-Murci L. Defining the genetic and evolutionary architecture of alternative splicing in response to infection. *Nat Commun.* 2019;10(1):1671. Epub 2019/04/13.
144. Liu H, Lorenzini PA, Zhang F, Xu S, Wong MSM, Zheng J, et al. Alternative splicing analysis in human monocytes and macrophages reveals MBNL1 as major regulator. *Nucleic Acids Res.* 2018;46(12):6069-86. Epub 2018/05/18.
145. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature.* 2013;498(7453):236-40. Epub 2013/05/21.
146. Kalam H, Fontana MF, Kumar D. Correction: Alternate splicing of transcripts shape macrophage response to Mycobacterium tuberculosis infection. *PLoS Pathog.* 2018;14(1):e1006833. Epub 2018/01/05.
147. Sinha S, van Nimwegen E, Siggia ED. A probabilistic method to detect regulatory modules. *Bioinformatics.* 2003;19 Suppl 1:i292-301. Epub 2003/07/12.
148. Liu H, Liu M, You H, Li X, Li X. Oncogenic Network and Hub Genes for Natural Killer/T-Cell Lymphoma Utilizing WGCNA. *Front Oncol.* 2020;10:223. Epub 2020/03/21.
149. Zhai X, Xue Q, Liu Q, Guo Y, Chen Z. Colon cancer recurrence-associated genes revealed by WGCNA coexpression network analysis. *Mol Med Rep.* 2017;16(5):6499-505. Epub 2017/09/14.
150. Chen S, Yang D, Lei C, Li Y, Sun X, Chen M, et al. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *PeerJ.* 2019;7:e7873. Epub 2019/10/15.

151. Barabási AL, Ravasza, E., Vicsek, T. Deterministic scale-free networks. Elsevier. 2001;299(3-4):559-64.
152. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. Nature. 2000;406(6794):378-82. Epub 2000/08/10.
153. Mason MJ, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. BMC Genomics. 2009;10:327. Epub 2009/07/22.
154. Langfelder P, Horvath S. Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol. 2007;1:54. Epub 2007/11/23.
155. Shay T, Jovic V, Zuk O, Rothamel K, Puyraimond-Zemmour D, Feng T, et al. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. Proc Natl Acad Sci U S A. 2013;110(8):2946-51. Epub 2013/02/06.
156. Heng TS, Painter MW, Immunological Genome Project C. The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol. 2008;9(10):1091-4. Epub 2008/09/19.
157. Immunological Genome P. ImmGen at 15. Nat Immunol. 2020;21(7):700-3. Epub 2020/06/25.
158. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816. Epub 2007/06/16.
159. Consortium EP. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011;9(4):e1001046. Epub 2011/04/29.
160. Consortium EP, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583(7818):699-710. Epub 2020/07/31.
161. Consortium EP, Snyder MP, Gingeras TR, Moore JE, Weng Z, Gerstein MB, et al. Perspectives on ENCODE. Nature. 2020;583(7818):693-8. Epub 2020/07/31.
162. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. Epub 2012/09/08.
163. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91-100. Epub 2012/09/08.
164. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, et al. A large-scale binding and functional map of human RNA-binding proteins. Nature. 2020;583(7818):711-9. Epub 2020/07/31.
165. Quinones-Valdez G, Tran SS, Jun HI, Bahn JH, Yang EW, Zhan L, et al. Regulation of RNA editing by RNA-binding proteins in human cells. Commun Biol. 2019;2:19. Epub 2019/01/18.
166. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009;10(10):669-80. Epub 2009/09/09.
167. Rye M, Saetrom P, Handstad T, Drablos F. Clustered ChIP-Seq-defined transcription factor binding sites and histone modifications map distinct classes of regulatory elements. BMC Biol. 2011;9:80. Epub 2011/11/26.
168. Wang J, Zhuang J, Iyer S, Lin XY, Greven MC, Kim BH, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res. 2013;41(Database issue):D171-6. Epub 2012/12/04.

169. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 2012;22(9):1798-812. Epub 2012/09/08.
170. Clark DN, Read RD, Mayhew V, Petersen SC, Argueta LB, Stutz LA, et al. Four Promoters of IRF5 Respond Distinctly to Stimuli and are Affected by Autoimmune-Risk Polymorphisms. *Front Immunol.* 2013;4:360. Epub 2013/11/14.
171. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference mapping of human transcription factor footprints. *Nature.* 2020;583(7818):729-36. Epub 2020/07/31.
172. Xiong T, Turner JE. Innate lymphoid cells in autoimmunity and chronic inflammatory diseases. *Semin Immunopathol.* 2018;40(4):393-406. Epub 2018/03/24.
173. Pockley AG, Henderson B. Extracellular cell stress (heat shock) proteins-immune responses and disease: an overview. *Philos Trans R Soc Lond B Biol Sci.* 2018;373(1738). Epub 2017/12/06.
174. Ribatti D. The concept of immune surveillance against tumors. The first theories. *Oncotarget.* 2017;8(4):7175-80. Epub 2016/10/21.
175. Conway EM, Pikor LA, Kung SH, Hamilton MJ, Lam S, Lam WL, et al. Macrophages, Inflammation, and Lung Cancer. *Am J Respir Crit Care Med.* 2016;193(2):116-30. Epub 2015/11/20.
176. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. *Nat Immunol.* 2017;18(7):725-32. Epub 2017/06/21.
177. Abbas AK, Lichtman AH, Pillai S. *Cellular and Molecular Immunology.* edition t, rédacteur2017 11th May 2017.
178. Murphy KM, Weaver C. *Janeway's Immunobiology.* edition t, rédacteur2016.
179. Underhill DM, Gordon S, Imhof BA, Nunez G, Bousso P. Elie Metchnikoff (1845-1916): celebrating 100 years of cellular immunology and beyond. *Nat Rev Immunol.* 2016;16(10):651-6. Epub 2016/08/02.
180. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science.* 1999;286(5439):531-7. Epub 1999/10/16.
181. Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med.* 2003;197(6):711-23. Epub 2003/03/19.
182. Lowes MA, Suarez-Farinas M, Krueger JG. Immunology of psoriasis. *Annu Rev Immunol.* 2014;32:227-55. Epub 2014/03/25.
183. Kim CC, Lanier LL. Beyond the transcriptome: completion of act one of the Immunological Genome Project. *Curr Opin Immunol.* 2013;25(5):593-7. Epub 2013/10/31.
184. Proserpio V, Mahata B. Single-cell technologies to study the immune system. *Immunology.* 2016;147(2):133-40. Epub 2015/11/10.
185. Shay T, Kang J. Immunological Genome Project and systems immunology. *Trends Immunol.* 2013;34(12):602-9. Epub 2013/05/02.
186. Bjorklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nat Immunol.* 2016;17(4):451-60. Epub 2016/02/16.
187. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science.* 2017;356(6335). Epub 2017/04/22.

188. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*. 2013;14:632. Epub 2013/09/24.
189. van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 2018;19(4):575-92. Epub 2017/01/13.
190. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20. Epub 2014/04/04.
191. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9. Epub 2009/06/10.
192. Delhomme N, Padioulet I, Furlong EE, Steinmetz LM. easyRNASeq: a bioconductor package for processing RNA-Seq data. *Bioinformatics*. 2012;28(19):2532-3. Epub 2012/08/01.
193. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139-40. Epub 2009/11/17.
194. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40(10):4288-97. Epub 2012/01/31.
195. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. Epub 2015/01/22.
196. Jolliffe IT, Morgan BJ. Principal component analysis and exploratory factor analysis. *Stat Methods Med Res*. 1992;1(1):69-95. Epub 1992/01/01.
197. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*. 2016;374(2065):20150202. Epub 2016/03/10.
198. Sainani KL. Introduction to principal components analysis. *PM R*. 2014;6(3):275-8. Epub 2014/02/26.
199. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17. Epub 2006/05/02.
200. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*. 2008;24(5):719-20. Epub 2007/11/21.
201. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57. Epub 2009/01/10.
202. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. The Universal Protein Resource (UniProt). *Nucleic Acids Res*. 2005;33(Database issue):D154-9. Epub 2004/12/21.
203. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res*. 2015;43(Database issue):D213-21. Epub 2014/11/28.
204. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015;43(Database issue):D1049-56. Epub 2014/11/28.
205. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457-62. Epub 2015/10/18.
206. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, et al. GeneCards Version 3: the human gene integrator. *Database (Oxford)*. 2010;2010:baq020. Epub 2010/08/07.

207. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011;2011:baq036. Epub 2011/01/20.
208. Anders ME, Evans DP. Comparison of PubMed and Google Scholar literature searches. *Respir Care*. 2010;55(5):578-83. Epub 2010/04/28.
209. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. 2006;34(Database issue):D108-10. Epub 2005/12/31.
210. Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2016;44(D1):D110-5. Epub 2015/11/05.
211. Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, et al. DNA-binding specificities of human transcription factors. *Cell*. 2013;152(1-2):327-39. Epub 2013/01/22.
212. Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*. 2015;527(7578):384-8. Epub 2015/11/10.
213. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. 2017;356(6337). Epub 2017/05/06.
214. Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res*. 2015;43(Database issue):D117-22. Epub 2014/11/08.
215. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*. 2014;158(6):1431-43. Epub 2014/09/13.
216. Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, et al. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol*. 2009;10(3):R29. Epub 2009/03/17.
217. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*. 2009;10(4):252-63. Epub 2009/03/11.
218. Wingender E, Schoeps T, Haubrock M, Donitz J. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res*. 2015;43(Database issue):D97-102. Epub 2014/11/02.
219. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxford)*. 2011;2011:bar030. Epub 2011/07/26.
220. Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, et al. Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol*. 2005;23(1):137-44. Epub 2005/01/08.
221. Veerla S, Ringner M, Hoglund M. Genome-wide transcription factor binding site/promoter databases for the analysis of gene sets and co-occurrence of transcription factor binding motifs. *BMC Genomics*. 2010;11:145. Epub 2010/03/03.
222. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*. 2013;14(2):144-61. Epub 2012/08/22.
223. Monaco G, Lee B, Xu W, Mustafah S, Hwang YY, Carre C, et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep*. 2019;26(6):1627-40 e7. Epub 2019/02/07.

224. Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, et al. Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A*. 2003;100(2):605-10. Epub 2003/01/17.
225. Dylke J, Lopes J, Dang-Lawson M, Machtaler S, Matsuuchi L. Role of the extracellular and transmembrane domain of Ig-alpha/beta in assembly of the B cell antigen receptor (BCR). *Immunol Lett*. 2007;112(1):47-57. Epub 2007/08/07.
226. Nutt SL, Kee BL. The transcriptional regulation of B cell lineage commitment. *Immunity*. 2007;26(6):715-25. Epub 2007/06/22.
227. Zandi S, Mansson R, Tsapogas P, Zetterblad J, Bryder D, Sigvardsson M. EBF1 is essential for B-lineage priming and establishment of a transcription factor network in common lymphoid progenitors. *J Immunol*. 2008;181(5):3364-72. Epub 2008/08/21.
228. Blackburn TE, Santiago T, Burrows PD. FCRLA-A Resident Endoplasmic Reticulum Protein that Associates with Multiple Immunoglobulin Isotypes in B Lineage Cells. *Curr Top Microbiol Immunol*. 2017;408:47-65. Epub 2017/09/08.
229. Muller J, Nitschke L. The role of CD22 and Siglec-G in B-cell tolerance and autoimmune disease. *Nat Rev Rheumatol*. 2014;10(7):422-8. Epub 2014/04/26.
230. Rinaldi A, Chiaravalli AM, Mian M, Zucca E, Tibiletti MG, Capella C, et al. Serotonin receptor 3A expression in normal and neoplastic B cells. *Pathobiology*. 2010;77(3):129-35. Epub 2010/06/03.
231. Li S, Liu J, Min Q, Ikawa T, Yasuda S, Yang Y, et al. Kelch-like protein 14 promotes B-1a but suppresses B-1b cell development. *Int Immunol*. 2018;30(7):311-8. Epub 2018/06/26.
232. Hagiwara S. Transgenic expression of VpreB-3 under the control of the immunoglobulin heavy chain enhancer and SV40 promoter. *Kobe J Med Sci*. 1996;42(1):43-59. Epub 1996/02/01.
233. Ebeling O, Duczmal A, Aigner S, Geiger C, Schollhammer S, Kemshead JT, et al. L1 adhesion molecule on human lymphocytes and monocytes: expression and involvement in binding to alpha v beta 3 integrin. *Eur J Immunol*. 1996;26(10):2508-16. Epub 1996/10/01.
234. Tabrizi SJ, Niuro H, Masui M, Yoshimoto G, Ino T, Kikushige Y, et al. T cell leukemia/lymphoma 1 and galectin-1 regulate survival/cell death pathways in human naive and IgM+ memory B cells through altering balances in Bcl-2 family proteins. *J Immunol*. 2009;182(3):1490-9. Epub 2009/01/22.
235. Imamura Y, Oda A, Katahira T, Bundo K, Pike KA, Ratcliffe MJ, et al. BLNK binds active H-Ras to promote B cell receptor-mediated capping and ERK activation. *J Biol Chem*. 2009;284(15):9804-13. Epub 2009/02/17.
236. Kabak S, Skaggs BJ, Gold MR, Affolter M, West KL, Foster MS, et al. The direct recruitment of BLNK to immunoglobulin alpha couples the B-cell antigen receptor to distal signaling pathways. *Mol Cell Biol*. 2002;22(8):2524-35. Epub 2002/03/23.
237. Klein J, Sato A. The HLA system. Second of two parts. *N Engl J Med*. 2000;343(11):782-6. Epub 2000/09/14.
238. Fortin JS, Cloutier M, Thibodeau J. Exposing the Specific Roles of the Invariant Chain Isoforms in Shaping the MHC Class II Peptidome. *Front Immunol*. 2013;4:443. Epub 2014/01/01.
239. Nguyen TA, Smith BRC, Tate MD, Belz GT, Barrios MH, Elgass KD, et al. SIDT2 Transports Extracellular dsRNA into the Cytoplasm for Innate Immune Recognition. *Immunity*. 2017;47(3):498-509 e6. Epub 2017/09/17.
240. Huber C, Martensson A, Bokoch GM, Nemazee D, Gavin AL. FGD2, a CDC42-specific exchange factor expressed by antigen-presenting cells, localizes to early endosomes and active membrane ruffles. *J Biol Chem*. 2008;283(49):34002-12. Epub 2008/10/08.

241. English WR, Puente XS, Freije JM, Knauper V, Amour A, Merryweather A, et al. Membrane type 4 matrix metalloproteinase (MMP17) has tumor necrosis factor-alpha convertase activity but does not activate pro-MMP2. *J Biol Chem.* 2000;275(19):14046-55. Epub 2000/05/09.
242. Medley QG, Kedersha N, O'Brien S, Tian Q, Schlossman SF, Streuli M, et al. Characterization of GMP-17, a granule membrane protein that moves to the plasma membrane of natural killer cells following target cell recognition. *Proc Natl Acad Sci U S A.* 1996;93(2):685-9. Epub 1996/01/23.
243. Chang GW, Hsiao CC, Peng YM, Vieira Braga FA, Kragten NA, Remmerswaal EB, et al. The Adhesion G Protein-Coupled Receptor GPR56/ADGRG1 Is an Inhibitory Receptor on Human NK Cells. *Cell Rep.* 2016;15(8):1757-70. Epub 2016/05/18.
244. Pazina T, Shemesh A, Brusilovsky M, Porgador A, Campbell KS. Regulation of the Functions of Natural Cytotoxicity Receptors by Interactions with Diverse Ligands and Alterations in Splice Variant Expression. *Front Immunol.* 2017;8:369. Epub 2017/04/21.
245. Perez-Quintero LA, Roncagalli R, Guo H, Latour S, Davidson D, Veillette A. EAT-2, a SAP-like adaptor, controls NK cell activation through phospholipase Cgamma, Ca⁺⁺, and Erk, leading to granule polarization. *J Exp Med.* 2014;211(4):727-42. Epub 2014/04/02.
246. Le Bouteiller P, Tabiasco J, Polgar B, Kozma N, Giustiniani J, Siewiera J, et al. CD160: a unique activating NK cell receptor. *Immunol Lett.* 2011;138(2):93-6. Epub 2011/02/18.
247. Cai G, Freeman GJ. The CD160, BTLA, LIGHT/HVEM pathway: a bidirectional switch regulating T-cell activation. *Immunol Rev.* 2009;229(1):244-58. Epub 2009/05/12.
248. Maher K, Konjar S, Watts C, Turk B, Kopitar-Jerala N. Cystatin F regulates proteinase activity in IL-2-activated natural killer cells. *Protein Pept Lett.* 2014;21(9):957-65. Epub 2014/04/08.
249. Stoeckle C, Gouttefangeas C, Hammer M, Weber E, Melms A, Tolosa E. Cathepsin W expressed exclusively in CD8⁺ T cells and NK cells, is secreted during target cell killing but is not essential for cytotoxicity in human CTLs. *Exp Hematol.* 2009;37(2):266-75. Epub 2008/12/23.
250. Yamazaki C, Miyamoto R, Hoshino K, Fukuda Y, Sasaki I, Saito M, et al. Conservation of a chemokine system, XCR1 and its ligand, XCL1, between human and mice. *Biochem Biophys Res Commun.* 2010;397(4):756-61. Epub 2010/06/15.
251. Lu D, Liu L, Ji X, Gao Y, Chen X, Liu Y, et al. The phosphatase DUSP2 controls the activity of the transcription activator STAT3 and regulates TH17 differentiation. *Nat Immunol.* 2015;16(12):1263-73. Epub 2015/10/20.
252. Zhu S, Phatarpekar PV, Denman CJ, Senyukov VV, Somanchi SS, Nguyen-Jackson HT, et al. Transcription of the activating receptor NKG2D in natural killer cells is regulated by STAT3 tyrosine phosphorylation. *Blood.* 2014;124(3):403-11. Epub 2014/06/04.
253. Lotem J, Levanon D, Negreanu V, Leshkowitz D, Friedlander G, Groner Y. Runx3-mediated transcriptional program in cytotoxic lymphocytes. *PLoS One.* 2013;8(11):e80467. Epub 2013/11/16.
254. Ohno S, Sato T, Kohu K, Takeda K, Okumura K, Satake M, et al. Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation. *Int Immunol.* 2008;20(1):71-9. Epub 2007/11/16.
255. Picozza M, Battistini L, Borsellino G. Mononuclear phagocytes and marker modulation: when CD16 disappears, CD38 takes the stage. *Blood.* 2013;122(3):456-7. Epub 2013/07/23.
256. Sierra-Filardi E, Nieto C, Dominguez-Soto A, Barroso R, Sanchez-Mateos P, Puig-Kroger A, et al. CCL2 shapes macrophage polarization by GM-CSF and M-CSF: identification of

- CCL2/CCR2-dependent gene expression profile. *J Immunol.* 2014;192(8):3858-67. Epub 2014/03/19.
257. Verreck FA, de Boer T, Langenberg DM, van der Zanden L, Ottenhoff TH. Phenotypic and functional profiling of human proinflammatory type-1 and anti-inflammatory type-2 macrophages in response to microbial antigens and IFN-gamma- and CD40L-mediated costimulation. *J Leukoc Biol.* 2006;79(2):285-93. Epub 2005/12/07.
258. Shinagawa S, Okazaki T, Ikeda M, Yudoh K, Kisanuki YY, Yanagisawa M, et al. T cells upon activation promote endothelin 1 production in monocytes via IFN-gamma and TNF-alpha. *Sci Rep.* 2017;7(1):14500. Epub 2017/11/05.
259. Subramanian Vignesh K, Deepe GS, Jr. Metallothioneins: Emerging Modulators in Immunity and Infection. *Int J Mol Sci.* 2017;18(10). Epub 2017/10/27.
260. Huang Y, Krein PM, Muruve DA, Winston BW. Complement factor B gene regulation: synergistic effects of TNF-alpha and IFN-gamma in macrophages. *J Immunol.* 2002;169(5):2627-35. Epub 2002/08/24.
261. Metz R, Duhadaway JB, Kamasani U, Laury-Kleintop L, Muller AJ, Prendergast GC. Novel tryptophan catabolic enzyme IDO2 is the preferred biochemical target of the antitumor indoleamine 2,3-dioxygenase inhibitory compound D-1-methyl-tryptophan. *Cancer Res.* 2007;67(15):7082-7. Epub 2007/08/03.
262. Kawata K, Kubota S, Eguchi T, Aoyama E, Moritani NH, Oka M, et al. A Tumor Suppressor Gene Product, Platelet-Derived Growth Factor Receptor-Like Protein Controls Chondrocyte Proliferation and Differentiation. *J Cell Biochem.* 2017;118(11):4033-44. Epub 2017/04/14.
263. Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A. Neutrophil function: from mechanisms to disease. *Annu Rev Immunol.* 2012;30:459-89. Epub 2012/01/10.
264. Jaillon S, Peri G, Delneste Y, Fremaux I, Doni A, Moalli F, et al. The humoral pattern recognition receptor PTX3 is stored in neutrophil granules and localizes in extracellular traps. *J Exp Med.* 2007;204(4):793-804. Epub 2007/03/29.
265. Vorobjeva NV, Pinegin BV. Neutrophil extracellular traps: mechanisms of formation and role in health and disease. *Biochemistry (Mosc).* 2014;79(12):1286-96. Epub 2015/02/27.
266. Yu Y, Wang J, Khaled W, Burke S, Li P, Chen X, et al. Bcl11a is essential for lymphoid development and negatively regulates p53. *J Exp Med.* 2012;209(13):2467-83. Epub 2012/12/12.
267. Liu P, Keller JR, Ortiz M, Tessarollo L, Rachel RA, Nakamura T, et al. Bcl11a is essential for normal lymphoid development. *Nat Immunol.* 2003;4(6):525-32. Epub 2003/04/30.
268. Lee BS, Lee BK, Iyer VR, Sleckman BP, Shaffer AL, 3rd, Ippolito GC, et al. Corrected and Republished from: BCL11A Is a Critical Component of a Transcriptional Network That Activates RAG Expression and V(D)J Recombination. *Mol Cell Biol.* 2018;38(1). Epub 2017/10/19.
269. Garvie CW, Stagno JR, Reid S, Singh A, Harrington E, Boss JM. Characterization of the RFX complex and the RFX5(L66A) mutant: implications for the regulation of MHC class II gene expression. *Biochemistry.* 2007;46(6):1597-611. Epub 2007/02/07.
270. Lochamy J, Rogers EM, Boss JM. CREB and phospho-CREB interact with RFX5 and CIITA to regulate MHC class II genes. *Mol Immunol.* 2007;44(5):837-47. Epub 2006/05/30.
271. Thakker S, Purushothaman P, Gupta N, Challa S, Cai Q, Verma SC. Kaposi's Sarcoma-Associated Herpesvirus Latency-Associated Nuclear Antigen Inhibits Major Histocompatibility Complex Class II Expression by Disrupting Enhanceosome Assembly through Binding with the Regulatory Factor X Complex. *J Virol.* 2015;89(10):5536-56. Epub 2015/03/06.

272. Boyd NH, Morgan JE, Greer SF. Polycomb recruitment at the Class II transactivator gene. *Mol Immunol*. 2015;67(2 Pt B):482-91. Epub 2015/08/19.
273. Mehta NT, Truax AD, Boyd NH, Greer SF. Early epigenetic events regulate the adaptive immune response gene CIITA. *Epigenetics*. 2011;6(4):516-25. Epub 2011/01/27.
274. Guo M, Price MJ, Patterson DG, Barwick BG, Haines RR, Kania AK, et al. EZH2 Represses the B Cell Transcriptional Program and Regulates Antibody-Secreting Cell Metabolism and Antibody Production. *J Immunol*. 2018;200(3):1039-52. Epub 2017/12/31.
275. Hashwah H, Schmid CA, Kasser S, Bertram K, Stelling A, Manz MG, et al. Inactivation of CREBBP expands the germinal center B cell compartment, down-regulates MHCII expression and promotes DLBCL growth. *Proc Natl Acad Sci U S A*. 2017;114(36):9701-6. Epub 2017/08/24.
276. Lin JH, Lin JY, Chou YC, Chen MR, Yeh TH, Lin CW, et al. Epstein-Barr virus LMP2A suppresses MHC class II expression by regulating the B-cell transcription factors E47 and PU.1. *Blood*. 2015;125(14):2228-38. Epub 2015/01/30.
277. Fujita N, Jaye DL, Geigerman C, Akyildiz A, Mooney MR, Boss JM, et al. MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation. *Cell*. 2004;119(1):75-86. Epub 2004/09/30.
278. Peng SL, Gerth AJ, Ranger AM, Glimcher LH. NFATc1 and NFATc2 together control both T and B cell activation and differentiation. *Immunity*. 2001;14(1):13-20. Epub 2001/02/13.
279. Tamura A, Hirai H, Yokota A, Kamio N, Sato A, Shoji T, et al. C/EBPbeta is required for survival of Ly6C(-) monocytes. *Blood*. 2017;130(16):1809-18. Epub 2017/08/16.
280. Murn J, Mlinaric-Rascan I, Vaigot P, Alibert O, Frouin V, Gidrol X. A Myc-regulated transcriptional network controls B-cell fate in response to BCR triggering. *BMC Genomics*. 2009;10:323. Epub 2009/07/18.
281. Habib T, Park H, Tsang M, de Alboran IM, Nicks A, Wilson L, et al. Myc stimulates B lymphocyte differentiation and amplifies calcium signaling. *J Cell Biol*. 2007;179(4):717-31. Epub 2007/11/14.
282. Rudolf R, Busch R, Patra AK, Muhammad K, Avots A, Andrau JC, et al. Architecture and expression of the nfatc1 gene in lymphocytes. *Front Immunol*. 2014;5:21. Epub 2014/02/20.
283. Muhammad K, Rudolf R, Pham DAT, Klein-Hessling S, Takata K, Matsushita N, et al. Induction of Short NFATc1/alphaA Isoform Interferes with Peripheral B Cell Differentiation. *Front Immunol*. 2018;9:32. Epub 2018/02/09.
284. Samitas K, Lotvall J, Bossios A. B cells: from early development to regulating allergic diseases. *Arch Immunol Ther Exp (Warsz)*. 2010;58(3):209-25. Epub 2010/05/12.
285. Miyazaki K, Miyazaki M, Murre C. The establishment of B versus T cell identity. *Trends Immunol*. 2014;35(5):205-10. Epub 2014/04/01.
286. Somasundaram R, Prasad MA, Ungerback J, Sigvardsson M. Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood*. 2015;126(2):144-52. Epub 2015/05/21.
287. Hu Y, Yoshida T, Georgopoulos K. Transcriptional circuits in B cell transformation. *Curr Opin Hematol*. 2017;24(4):345-52. Epub 2017/05/04.
288. Schmitt C, Tonnelle C, Dalloul A, Chabannon C, Debre P, Rebollo A. Aiolos and Ikaros: regulators of lymphocyte development, homeostasis and lymphoproliferation. *Apoptosis*. 2002;7(3):277-84. Epub 2002/05/09.
289. Itoh-Nakadai A, Matsumoto M, Kato H, Sasaki J, Uehara Y, Sato Y, et al. A Bach2-Cebp Gene Regulatory Network for the Commitment of Multipotent Hematopoietic Progenitors. *Cell Rep*. 2017;18(10):2401-14. Epub 2017/03/09.

290. Ji Y, Lee HJ, Goodman C, Uskokovic M, Liby K, Sporn M, et al. The synthetic triterpenoid CDDO-imidazole induces monocytic differentiation by activating the Smad and ERK signaling pathways in HL60 leukemia cells. *Mol Cancer Ther.* 2006;5(6):1452-8. Epub 2006/07/05.
291. Wang C, Petriello MC, Zhu B, Hennig B. PCB 126 induces monocyte/macrophage polarization and inflammation through AhR and NF-kappaB pathways. *Toxicol Appl Pharmacol.* 2019;367:71-81. Epub 2019/02/16.
292. Hume DA. The Many Alternative Faces of Macrophage Activation. *Front Immunol.* 2015;6:370. Epub 2015/08/11.
293. Pierson E, Consortium GT, Koller D, Battle A, Mostafavi S, Ardlie KG, et al. Sharing and Specificity of Co-expression Networks across 35 Human Tissues. *PLoS Comput Biol.* 2015;11(5):e1004220. Epub 2015/05/15.
294. Zhou L, Chong MM, Littman DR. Plasticity of CD4+ T cell lineage differentiation. *Immunity.* 2009;30(5):646-55. Epub 2009/05/26.
295. Kyaw T, Toh BH, Bobik A. Foxp3+CD4+ Regulatory T-Cell Subtypes and Atherosclerosis. *Circ Res.* 2016;119(11):1151-3. Epub 2017/01/05.
296. Kraj P, Ignatowicz L. The mechanisms shaping the repertoire of CD4(+) Foxp3(+) regulatory T cells. *Immunology.* 2018;153(3):290-6. Epub 2017/11/07.
297. de Klerk E, t Hoen PA. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet.* 2015;31(3):128-39. Epub 2015/02/05.
298. Zhao J, Hyman L, Moore C. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev.* 1999;63(2):405-45. Epub 1999/06/05.
299. Reyes A, Huber W. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Res.* 2018;46(2):582-92. Epub 2017/12/05.
300. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A.* 2003;100(1):189-92. Epub 2002/12/28.
301. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357-60. Epub 2015/03/10.
302. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004;5(10):R80. Epub 2004/10/06.
303. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014;15(2):R29. Epub 2014/02/04.
304. Akerman M, Fregoso OI, Das S, Ruse C, Jensen MA, Pappin DJ, et al. Differential connectivity of splicing activators and repressors to the human spliceosome. *Genome Biol.* 2015;16:119. Epub 2015/06/07.
305. Hermiston ML, Xu Z, Weiss A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu Rev Immunol.* 2003;21:107-37. Epub 2002/11/05.
306. Rheinlander A, Schraven B, Bommhardt U. CD45 in human physiology and clinical medicine. *Immunol Lett.* 2018;196:22-32. Epub 2018/01/26.
307. Sun J, Zheng MY, Li YW, Zhang SW. Structure and function of Septin 9 and its role in human malignant tumors. *World J Gastrointest Oncol.* 2020;12(6):619-31. Epub 2020/07/24.
308. Marcus J, Bejerano-Sagie M, Patterson N, Bagchi S, Verkhusha VV, Connolly D, et al. Septin 9 isoforms promote tumorigenesis in mammary epithelial cells by increasing migration

- and ECM degradation through metalloproteinase secretion at focal adhesions. *Oncogene*. 2019;38(30):5839-59. Epub 2019/07/10.
309. Bai X, Bowen JR, Knox TK, Zhou K, Pendziwiat M, Kuhlenbaumer G, et al. Novel septin 9 repeat motifs altered in neuralgic amyotrophy bind and bundle microtubules. *J Cell Biol*. 2013;203(6):895-905. Epub 2013/12/18.
310. Kuzmic M, Castro Linares G, Leischner Fialova J, Iv F, Salaun D, Llewellyn A, et al. Septin-microtubule association via a motif unique to isoform 1 of septin 9 tunes stress fibers. *J Cell Sci*. 2022;135(1). Epub 2021/12/03.
311. Hassn Mesrati M, Syafruddin SE, Mohtar MA, Syahir A. CD44: A Multifunctional Mediator of Cancer Progression. *Biomolecules*. 2021;11(12). Epub 2021/12/25.
312. Xu H, Manivannan A, Crane I, Dawson R, Liversidge J. Critical but divergent roles for CD62L and CD44 in directing blood monocyte trafficking in vivo during inflammation. *Blood*. 2008;112(4):1166-74. Epub 2008/04/09.
313. Qadri M, Almadani S, Jay GD, Elsaid KA. Role of CD44 in Regulating TLR2 Activation of Human Macrophages and Downstream Expression of Proinflammatory Cytokines. *J Immunol*. 2018;200(2):758-67. Epub 2017/12/03.
314. Johnson P, Ruffell B. CD44 and its role in inflammation and inflammatory diseases. *Inflamm Allergy Drug Targets*. 2009;8(3):208-20. Epub 2009/07/16.
315. Andreoli V, Gehrau RC, Bocco JL. Biology of Kruppel-like factor 6 transcriptional regulator in cell life and death. *IUBMB Life*. 2010;62(12):896-905. Epub 2010/12/15.
316. Syafruddin SE, Mohtar MA, Wan Mohamad Nazarie WF, Low TY. Two Sides of the Same Coin: The Roles of KLF6 in Physiology and Pathophysiology. *Biomolecules*. 2020;10(10). Epub 2020/10/02.
317. Date D, Das R, Narla G, Simon DI, Jain MK, Mahabeleshwar GH. Kruppel-like transcription factor 6 regulates inflammatory macrophage polarization. *J Biol Chem*. 2014;289(15):10318-29. Epub 2014/01/05.
318. Zhang Y, Lei CQ, Hu YH, Xia T, Li M, Zhong B, et al. Kruppel-like factor 6 is a co-activator of NF-kappaB that mediates p65-dependent transcription of selected downstream genes. *J Biol Chem*. 2014;289(18):12876-85. Epub 2014/03/19.
319. Kim GD, Das R, Goduni L, McClellan S, Hazlett LD, Mahabeleshwar GH. Kruppel-like Factor 6 Promotes Macrophage-mediated Inflammation by Suppressing B Cell Leukemia/Lymphoma 6 Expression. *J Biol Chem*. 2016;291(40):21271-82. Epub 2016/08/20.
320. Kim GD, Ng HP, Chan ER, Mahabeleshwar GH. Kruppel-like factor 6 promotes macrophage inflammatory and hypoxia response. *FASEB J*. 2020;34(2):3209-23. Epub 2020/01/08.
321. Vang T, Miletic AV, Arimura Y, Tautz L, Rickert RC, Mustelin T. Protein tyrosine phosphatases in autoimmunity. *Annu Rev Immunol*. 2008;26:29-55. Epub 2008/02/29.
322. Kopf A, Kiermaier E. Dynamic Microtubule Arrays in Leukocytes and Their Role in Cell Migration and Immune Synapse Formation. *Front Cell Dev Biol*. 2021;9:635511. Epub 2021/02/27.
323. Akira S, Misawa T, Satoh T, Saitoh T. Macrophages control innate inflammation. *Diabetes Obes Metab*. 2013;15 Suppl 3:10-8. Epub 2013/09/18.
324. Verdier-Pinard P, Salaun D, Bouguenina H, Shimada S, Pophillat M, Audebert S, et al. Septin 9_i2 is downregulated in tumors, impairs cancer cell migration and alters subnuclear actin filaments. *Sci Rep*. 2017;7:44976. Epub 2017/03/25.

325. Hanania R, Sun HS, Xu K, Pustyl'nik S, Jeganathan S, Harrison RE. Classically activated macrophages use stable microtubules for matrix metalloproteinase-9 (MMP-9) secretion. *J Biol Chem.* 2012;287(11):8468-83. Epub 2012/01/25.
326. Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell.* 2014;156(4):812-24. Epub 2014/02/18.
327. Koh WS, Porter JR, Batchelor E. Tuning of mRNA stability through altering 3'-UTR sequences generates distinct output expression in a synthetic circuit driven by p53 oscillations. *Sci Rep.* 2019;9(1):5976. Epub 2019/04/14.
328. Jia L, Mao Y, Ji Q, Dersh D, Yewdell JW, Qian SB. Decoding mRNA translatability and stability from the 5' UTR. *Nat Struct Mol Biol.* 2020;27(9):814-21. Epub 2020/07/29.
329. UniProt C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49(D1):D480-D9. Epub 2020/11/26.
330. Vallier P, Bata J, Colobert L. [Optimal conditions of alpha-amylase production by *Aspergillus oryzae* in liquid media]. *Ann Microbiol (Paris).* 1977;128(3):359-71. Epub 1977/10/01. Conditions optimales de production d'alpha-amylase en milieu liquide par *Aspergillus oryzae*.
331. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 2012;13(4):227-32. Epub 2012/03/14.
332. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci Rep.* 2015;5:10775. Epub 2015/06/09.
333. Le Bras S, Moon C, Foucault I, Breittmayer JP, Deckert M. Abl-SH3 binding protein 2, 3BP2, interacts with CIN85 and HIP-55. *FEBS Lett.* 2007;581(5):967-74. Epub 2007/02/20.
334. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet.* 2019;20(11):631-56. Epub 2019/07/26.
335. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. Epub 2012/10/30.
336. Singh ON, Burgess J. Characterization of albumin-alginate complex coacervation. *J Pharm Pharmacol.* 1989;41(10):670-3. Epub 1989/10/01.
337. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-9. Epub 2014/09/28.
338. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923-30. Epub 2013/11/15.
339. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11(3):R25. Epub 2010/03/04.
340. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. Epub 2014/12/18.
341. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525-7. Epub 2016/04/05.
342. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 2015;4:1521. Epub 2016/03/01.
343. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc.* 2012;7(3):562-78. Epub 2012/03/03.

344. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016;17:13. Epub 2016/01/28.
345. Turro E, Su SY, Goncalves A, Coin LJ, Richardson S, Lewin A. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol.* 2011;12(2):R13. Epub 2011/02/12.
346. Frazee AC, Perteza G, Jaffe AE, Langmead B, Salzberg SL, Leek JT. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol.* 2015;33(3):243-6. Epub 2015/03/10.
347. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics.* 2014;30(12):1660-6. Epub 2014/02/18.
348. Karamitros D, Stoilova B, Aboukhalil Z, Hamey F, Reinisch A, Samitsch M, et al. Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol.* 2018;19(1):85-97. Epub 2017/11/24.
349. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell.* 2017;65(4):631-43 e4. Epub 2017/02/19.
350. Shintaku H, Nishikii H, Marshall LA, Kotera H, Santiago JG. On-chip separation and analysis of RNA and DNA from single cells. *Anal Chem.* 2014;86(4):1953-7. Epub 2014/02/07.
351. Lovering RC, Gaudet P, Acencio ML, Ignatchenko A, Jolma A, Fornes O, et al. A GO catalogue of human DNA-binding transcription factors. *Biochim Biophys Acta Gene Regul Mech.* 2021;1864(11-12):194765. Epub 2021/10/22.
352. Chawla K, Tripathi S, Thommesen L, Laegreid A, Kuiper M. TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors. *Bioinformatics.* 2013;29(19):2519-20. Epub 2013/08/13.
353. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* 2019;47(D1):D33-D8. Epub 2018/09/12.
354. Feng C, Song C, Liu Y, Qian F, Gao Y, Ning Z, et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res.* 2020;48(D1):D93-D100. Epub 2019/10/11.
355. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 2013;41(Database issue):D991-5. Epub 2012/11/30.
356. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science.* 2004;306(5696):636-40. Epub 2004/10/23.
357. Czipa E, Schiller M, Nagy T, Kontra L, Steiner L, Koller J, et al. ChIPSummitDB: a ChIP-seq-based database of human transcription factor binding sites and the topological arrangements of the proteins bound to them. *Database (Oxford).* 2020;2020. Epub 2020/01/17.
358. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C. The sequence read archive. *Nucleic Acids Res.* 2011;39(Database issue):D19-21. Epub 2010/11/11.
359. Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, et al. hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. *Genomics Proteomics Bioinformatics.* 2020;18(2):120-8. Epub 2020/08/29.

360. Cao J, Dong R, Jiang L, Gong Y, Yuan M, You J, et al. LncRNA-MM2P Identified as a Modulator of Macrophage M2 Polarization. *Cancer Immunol Res.* 2019;7(2):292-305. Epub 2018/11/22.
361. Zhang P, Cao L, Zhou R, Yang X, Wu M. The lncRNA Neat1 promotes activation of inflammasomes in macrophages. *Nat Commun.* 2019;10(1):1495. Epub 2019/04/04.
362. Yan B, Yao J, Liu JY, Li XM, Wang XQ, Li YJ, et al. lncRNA-MIAT regulates microvascular dysfunction by functioning as a competing endogenous RNA. *Circ Res.* 2015;116(7):1143-56. Epub 2015/01/15.
363. Iwakiri J, Hamada M, Asai K. Bioinformatics tools for lncRNA research. *Biochim Biophys Acta.* 2016;1859(1):23-30. Epub 2015/08/15.
364. Zhu J, Fu H, Wu Y, Zheng X. Function of lncRNAs and approaches to lncRNA-protein interactions. *Sci China Life Sci.* 2013;56(10):876-85. Epub 2013/10/05.
365. Allantaz F, Cheng DT, Bergauer T, Ravindran P, Rossier MF, Ebeling M, et al. Expression profiling of human immune cell subsets identifies miRNA-mRNA regulatory relationships correlated with cell type specific expression. *PLoS One.* 2012;7(1):e29979. Epub 2012/01/26.
366. Salehi S, Taheri MN, Azarpira N, Zare A, Behzad-Behbahani A. State of the art technologies to explore long non-coding RNAs in cancer. *J Cell Mol Med.* 2017;21(12):3120-40. Epub 2017/06/21.
367. Dailey UG. Current Medical Thought. *J Natl Med Assoc.* 1916;8(3):166-7. Epub 1916/07/01.
368. Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. *J Cell Sci.* 2016;129(3):461-7. Epub 2016/01/21.
369. Hug N, Longman D, Caceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 2016;44(4):1483-95. Epub 2016/01/17.
370. Nguyen LS, Wilkinson MF, Gecz J. Nonsense-mediated mRNA decay: inter-individual variability and human disease. *Neurosci Biobehav Rev.* 2014;46 Pt 2:175-86. Epub 2013/11/19.
371. Bicknell AA, Ricci EP. When mRNA translation meets decay. *Biochem Soc Trans.* 2017;45(2):339-51. Epub 2017/04/15.
372. Smith JE, Baker KE. Nonsense-mediated RNA decay--a switch and dial for regulating gene expression. *Bioessays.* 2015;37(6):612-23. Epub 2015/03/31.
373. Silva AL, Ribeiro P, Inacio A, Liebhaber SA, Romao L. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. *RNA.* 2008;14(3):563-76. Epub 2008/01/31.
374. Silva AL, Pereira FJ, Morgado A, Kong J, Martins R, Faustino P, et al. The canonical UPF1-dependent nonsense-mediated mRNA decay is inhibited in transcripts carrying a short open reading frame independent of sequence context. *RNA.* 2006;12(12):2160-70. Epub 2006/11/02.
375. Inacio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, et al. Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. *J Biol Chem.* 2004;279(31):32170-80. Epub 2004/05/27.
376. Zhao LY, Song J, Liu Y, Song CX, Yi C. Mapping the epigenetic modifications of DNA and RNA. *Protein Cell.* 2020;11(11):792-808. Epub 2020/05/23.
377. Gjaltema RAF, Rots MG. Advances of epigenetic editing. *Curr Opin Chem Biol.* 2020;57:75-81. Epub 2020/07/04.

378. Santoso CS, Li Z, Lal S, Yuan S, Gan KA, Agosto LM, et al. Comprehensive mapping of the human cytokine gene regulatory network. *Nucleic Acids Res.* 2020;48(21):12055-73. Epub 2020/11/13.
379. Zhou L, Ma X, Sun F. The effects of protein interactions, gene essentiality and regulatory regions on expression variation. *BMC Syst Biol.* 2008;2:54. Epub 2008/06/28.
380. Bernal-Quiros M, Wu YY, Alarcon-Riquelme ME, Castillejo-Lopez C. BANK1 and BLK act through phospholipase C gamma 2 in B-cell signaling. *PLoS One.* 2013;8(3):e59842. Epub 2013/04/05.
381. Draber P, Vonkova I, Stepanek O, Hrdinka M, Kucova M, Skopcova T, et al. SCIMP, a transmembrane adaptor protein involved in major histocompatibility complex class II signaling. *Mol Cell Biol.* 2011;31(22):4550-62. Epub 2011/09/21.
382. Kralova J, Fabisik M, Pokorna J, Skopcova T, Malissen B, Brdicka T. The Transmembrane Adaptor Protein SCIMP Facilitates Sustained Dectin-1 Signaling in Dendritic Cells. *J Biol Chem.* 2016;291(32):16530-40. Epub 2016/06/12.
383. Weekes MP, Antrobus R, Talbot S, Hor S, Simecek N, Smith DL, et al. Proteomic plasma membrane profiling reveals an essential role for gp96 in the cell surface expression of LDLR family members, including the LDL receptor and LRP6. *J Proteome Res.* 2012;11(3):1475-84. Epub 2012/02/02.
384. Dominguez-Soto A, Aragonese-Fenoll L, Martin-Gayo E, Martinez-Prats L, Colmenares M, Naranjo-Gomez M, et al. The DC-SIGN-related lectin LSECtin mediates antigen capture and pathogen binding by human myeloid cells. *Blood.* 2007;109(12):5337-45. Epub 2007/03/07.
385. Weiss MJ, Henthorn PS, Lafferty MA, Slaughter C, Raducha M, Harris H. Isolation and characterization of a cDNA encoding a human liver/bone/kidney-type alkaline phosphatase. *Proc Natl Acad Sci U S A.* 1986;83(19):7182-6. Epub 1986/10/01.
386. Bergin DA, Reeves EP, Meleady P, Henry M, McElvaney OJ, Carroll TP, et al. alpha-1 Antitrypsin regulates human neutrophil chemotaxis induced by soluble immune complexes and IL-8. *J Clin Invest.* 2010;120(12):4236-50. Epub 2010/11/10.
387. Kamohara H, Matsuyama W, Shimozato O, Abe K, Galligan C, Hashimoto S, et al. Regulation of tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) and TRAIL receptor expression in human neutrophils. *Immunology.* 2004;111(2):186-94. Epub 2004/03/19.
388. Elghetany MT. Surface antigen changes during normal neutrophilic development: a critical review. *Blood Cells Mol Dis.* 2002;28(2):260-74. Epub 2002/06/18.
389. Starr AE, Bellac CL, Dufour A, Goebeler V, Overall CM. Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *J Biol Chem.* 2012;287(16):13382-95. Epub 2012/03/01.
390. Montell C. Mg²⁺ homeostasis: the Mg²⁺-sensitive TRPM channels. *Curr Biol.* 2003;13(20):R799-801. Epub 2003/10/17.
391. Nadolni W, Zierler S. The Channel-Kinase TRPM7 as Novel Regulator of Immune System Homeostasis. *Cells.* 2018;7(8). Epub 2018/08/22.
392. Coakley RJ, Taggart C, Greene C, McElvaney NG, O'Neill SJ. Ambient pCO₂ modulates intracellular pH, intracellular oxidant generation, and interleukin-8 secretion in human neutrophils. *J Leukoc Biol.* 2002;71(4):603-10. Epub 2002/04/03.
393. Barker H, Aaltonen M, Pan P, Vahatupa M, Kaipainen P, May U, et al. Role of carbonic anhydrases in skin wound healing. *Exp Mol Med.* 2017;49(5):e334. Epub 2017/05/20.
394. Zhou W, Zhao Q, Sutton R, Cumming H, Wang X, Cerruti L, et al. The role of p22 NF-E4 in human globin gene switching. *J Biol Chem.* 2004;279(25):26227-32. Epub 2004/04/16.

395. Gilbert C, Levasseur S, Desaulniers P, Dusseault AA, Thibault N, Bourgoin SG, et al. Chemotactic factor-induced recruitment and activation of Tec family kinases in human neutrophils. II. Effects of LFM-A13, a specific Btk inhibitor. *J Immunol.* 2003;170(10):5235-43. Epub 2003/05/08.
396. Xu K, Cooney KA, Shin EY, Wang L, Deppen JN, Ginn SC, et al. Adenosine from a biologic source regulates neutrophil extracellular traps (NETs). *J Leukoc Biol.* 2019. Epub 2019/03/26.
397. Liu YW, Yang T, Zhao L, Ni Z, Yang N, He F, et al. Activation of Adenosine 2A receptor inhibits neutrophil apoptosis in an autophagy-dependent manner in mice with systemic inflammatory response syndrome. *Sci Rep.* 2016;6:33614. Epub 2016/09/21.
398. Canela L, Lujan R, Lluís C, Burgueno J, Mallol J, Canela EI, et al. The neuronal Ca(2+) - binding protein 2 (NECAB2) interacts with the adenosine A(2A) receptor and modulates the cell surface expression and function of the receptor. *Mol Cell Neurosci.* 2007;36(1):1-12. Epub 2007/08/11.
399. Jiang S, Zhang M, Sun J, Yang X. Casein kinase 1alpha: biological mechanisms and theranostic potential. *Cell Commun Signal.* 2018;16(1):23. Epub 2018/05/26.
400. Pitard V, Roumanes D, Lafarge X, Couzi L, Garrigue I, Lafon ME, et al. Long-term expansion of effector/memory Vdelta2-gammadelta T cells is a specific blood signature of CMV infection. *Blood.* 2008;112(4):1317-24. Epub 2008/06/10.
401. Farber DL, Yudanin NA, Restifo NP. Human memory T cells: generation, compartmentalization and homeostasis. *Nat Rev Immunol.* 2014;14(1):24-35. Epub 2013/12/18.
402. Angelo LS, Banerjee PP, Monaco-Shawver L, Rosen JB, Makedonas G, Forbes LR, et al. Practical NK cell phenotyping and variability in healthy adults. *Immunol Res.* 2015;62(3):341-56. Epub 2015/05/28.
403. Boyette LB, Macedo C, Hadi K, Elinoff BD, Walters JT, Ramaswami B, et al. Phenotype, function, and differentiation potential of human monocyte subsets. *PLoS One.* 2017;12(4):e0176460. Epub 2017/04/27.
404. Rey-Giraud F, Hafner M, Ries CH. In vitro generation of monocyte-derived macrophages under serum-free conditions improves their tumor promoting functions. *PLoS One.* 2012;7(8):e42656. Epub 2012/08/11.
405. Pang DJ, Neves JF, Sumaria N, Pennington DJ. Understanding the complexity of gammadelta T-cell subsets in mouse and human. *Immunology.* 2012;136(3):283-90. Epub 2012/03/06.
406. Farag SS, Caligiuri MA. Human natural killer cell development and biology. *Blood Rev.* 2006;20(3):123-37. Epub 2005/12/21.
407. Veale M, Raab M, Li Z, da Silva AJ, Kraeft SK, Weremowicz S, et al. Novel isoform of lymphoid adaptor FYN-T-binding protein (FYB-130) interacts with SLP-76 and up-regulates interleukin 2 production. *J Biol Chem.* 1999;274(40):28427-35. Epub 1999/09/25.
408. Engelmann S, Togni M, Kliche S, Reinhold D, Schraven B, Reinhold A. The adhesion- and degranulation-promoting adaptor protein and its role in the modulation of experimental autoimmune encephalomyelitis. *Crit Rev Immunol.* 2015;35(1):1-14. Epub 2015/03/10.
409. Yuan M, Mogemark L, Fallman M. Fyn binding protein, Fyb, interacts with mammalian actin binding protein, mAbp1. *FEBS Lett.* 2005;579(11):2339-47. Epub 2005/04/26.
410. Samara P, Ioannou K, Tsitsilonis OE. Prothymosin Alpha and Immune Responses: Are We Close to Potential Clinical Applications? *Vitam Horm.* 2016;102:179-207. Epub 2016/07/28.

411. Wong LE, Bhatt A, Erdmann PS, Hou Z, Maier J, Pirkuliyeva S, et al. Tripartite phase separation of two signal effectors with vesicles priming B cell responsiveness. *Nat Commun.* 2020;11(1):848. Epub 2020/02/14.
412. Keller B, Shoukier M, Schulz K, Bhatt A, Heine I, Strohmeier V, et al. Germline deletion of CIN85 in humans with X chromosome-linked antibody deficiency. *J Exp Med.* 2018;215(5):1327-36. Epub 2018/04/11.
413. Yakymovych I, Yakymovych M, Zang G, Mu Y, Bergh A, Landstrom M, et al. CIN85 modulates TGFbeta signaling by promoting the presentation of TGFbeta receptors on the cell surface. *J Cell Biol.* 2015;210(2):319-32. Epub 2015/07/15.
414. Tibaldi EV, Reinherz EL. CD2BP3, CIN85 and the structurally related adaptor protein CMS bind to the same CD2 cytoplasmic segment, but elicit divergent functional activities. *Int Immunol.* 2003;15(3):313-29. Epub 2003/03/06.
415. Sun AQ, Yuksel KU, Jacobson TM, Gracy RW. Isolation and characterization of human glucose-6-phosphate isomerase isoforms containing two different size subunits. *Arch Biochem Biophys.* 1990;283(1):120-9. Epub 1990/11/15.
416. Gurney ME, Apatoff BR, Spear GT, Baumel MJ, Antel JP, Bania MB, et al. Neuroleukin: a lymphokine product of lectin-stimulated T cells. *Science.* 1986;234(4776):574-81. Epub 1986/10/31.
417. Read J, Pearce J, Li X, Muirhead H, Chirgwin J, Davies C. The crystal structure of human phosphoglucose isomerase at 1.6 Å resolution: implications for catalytic mechanism, cytokine activity and haemolytic anaemia. *J Mol Biol.* 2001;309(2):447-63. Epub 2001/05/24.
418. Repiso A, Oliva B, Vives-Corróns JL, Beutler E, Carreras J, Climent F. Red cell glucose phosphate isomerase (GPI): a molecular study of three novel mutations associated with hereditary nonspherocytic hemolytic anemia. *Hum Mutat.* 2006;27(11):1159. Epub 2006/10/17.
419. Nakajima K, Raz A. Autocrine motility factor and its receptor expression in musculoskeletal tumors. *J Bone Oncol.* 2020;24:100318. Epub 2020/10/27.
420. Shehata HM, Murphy AJ, Lee MKS, Gardiner CM, Crowe SM, Sanjabi S, et al. Sugar or Fat?-Metabolic Requirements for Immunity to Viral Infections. *Front Immunol.* 2017;8:1311. Epub 2017/11/01.
421. Bolcato-Bellemin AL, Mattei MG, Fenton M, Amar S. Molecular cloning and characterization of mouse LITAF cDNA: role in the regulation of tumor necrosis factor-alpha (TNF-alpha) gene expression. *J Endotoxin Res.* 2004;10(1):15-23. Epub 2004/03/18.
422. Tang X, Marciano DL, Leeman SE, Amar S. LPS induces the interaction of a transcription factor, LPS-induced TNF-alpha factor, and STAT6(B) with effects on multiple cytokines. *Proc Natl Acad Sci U S A.* 2005;102(14):5132-7. Epub 2005/03/29.
423. Guimaraes-Costa R, Iancu Ferfoggia R, Leonard-Louis S, Ziegler F, Magy L, Fournier E, et al. Phenotypic spectrum of Charcot-Marie-Tooth disease due to LITAF/SIMPLE mutations: a study of 18 patients. *Eur J Neurol.* 2017;24(3):530-8. Epub 2017/02/18.
424. Lacerda AF, Hartjes E, Brunetti CR. LITAF mutations associated with Charcot-Marie-Tooth disease 1C show mislocalization from the late endosome/lysosome to the mitochondria. *PLoS One.* 2014;9(7):e103454. Epub 2014/07/25.
425. Lee SM, Olzmann JA, Chin LS, Li L. Mutations associated with Charcot-Marie-Tooth disease cause SIMPLE protein mislocalization and degradation by the proteasome and aggresome-autophagy pathways. *J Cell Sci.* 2011;124(Pt 19):3319-31. Epub 2011/09/08.
426. Cantoni C, Bottino C, Augugliaro R, Morelli L, Marcenaro E, Castriconi R, et al. Molecular and functional characterization of IRp60, a member of the immunoglobulin

- superfamily that functions as an inhibitory receptor in human NK cells. *Eur J Immunol*. 1999;29(10):3148-59. Epub 1999/10/30.
427. Borrego F. The CD300 molecules: an emerging family of regulators of the immune system. *Blood*. 2013;121(11):1951-60. Epub 2013/01/08.
428. Simhadri VR, Andersen JF, Calvo E, Choi SC, Coligan JE, Borrego F. Human CD300a binds to phosphatidylethanolamine and phosphatidylserine, and modulates the phagocytosis of dead cells. *Blood*. 2012;119(12):2799-809. Epub 2012/02/04.
429. Carnec X, Meertens L, Dejarnac O, Perera-Lecoin M, Hafirassou ML, Kitaura J, et al. The Phosphatidylserine and Phosphatidylethanolamine Receptor CD300a Binds Dengue Virus and Enhances Infection. *J Virol*. 2016;90(1):92-102. Epub 2015/10/16.
430. Qureshi N, Perera PY, Shen J, Zhang G, Lenschat A, Splitter G, et al. The proteasome as a lipopolysaccharide-binding protein in macrophages: differential effects of proteasome inhibition on lipopolysaccharide-induced signaling events. *J Immunol*. 2003;171(3):1515-25. Epub 2003/07/23.
431. Pang B, Hu C, Wu G, Zhang Y, Lin G. Identification of Target Genes in Hypertension and Left Ventricular Remodeling. *Medicine (Baltimore)*. 2020;99(28):e21195. Epub 2020/07/16.
432. Capalbo L, Bassi ZI, Geymonat M, Todesca S, Copoiu L, Enright AJ, et al. The midbody interactome reveals unexpected roles for PP1 phosphatases in cytokinesis. *Nat Commun*. 2019;10(1):4513. Epub 2019/10/06.
433. Buschow SI, van Balkom BW, Aalberts M, Heck AJ, Wauben M, Stoorvogel W. MHC class II-associated proteins in B-cell exosomes and potential functional implications for exosome biogenesis. *Immunol Cell Biol*. 2010;88(8):851-6. Epub 2010/05/12.
434. Foulon M, Noel P. [Early post-traumatic epilepsy in childhood: significance and short term prognosis (author's transl)]. *Acta Neurol Belg*. 1977;77(5):276-84. Epub 1977/09/01. Epilepsie post-traumatique precoce dans l'enfance: signification et pronostic a court terme.
435. Zhang HX, Liu ZX, Sun YP, Zhu J, Lu SY, Liu XS, et al. Rig-I regulates NF-kappaB activity through binding to Nf-kappab1 3'-UTR mRNA. *Proc Natl Acad Sci U S A*. 2013;110(16):6459-64. Epub 2013/04/05.
436. Hardbower DM, Asim M, Luis PB, Singh K, Barry DP, Yang C, et al. Ornithine decarboxylase regulates M1 macrophage activation and mucosal inflammation via histone modifications. *Proc Natl Acad Sci U S A*. 2017;114(5):E751-E60. Epub 2017/01/18.
437. Murai N, Shimizu A, Murakami Y, Matsufuji S. Subcellular localization and phosphorylation of antizyme 2. *J Cell Biochem*. 2009;108(4):1012-21. Epub 2009/09/03.
438. Tsujita K, Kondo A, Kurisu S, Hasegawa J, Itoh T, Takenawa T. Antagonistic regulation of F-BAR protein assemblies controls actin polymerization during podosome formation. *J Cell Sci*. 2013;126(Pt 10):2267-78. Epub 2013/03/26.
439. Tsuboi S, Takada H, Hara T, Mochizuki N, Funyu T, Saitoh H, et al. FBP17 Mediates a Common Molecular Step in the Formation of Podosomes and Phagocytic Cups in Macrophages. *J Biol Chem*. 2009;284(13):8548-56. Epub 2009/01/22.
440. Shimada A, Niwa H, Tsujita K, Suetsugu S, Nitta K, Hanawa-Suetsugu K, et al. Curved EFC/F-BAR-domain dimers are joined end to end into a filament for membrane invagination in endocytosis. *Cell*. 2007;129(4):761-72. Epub 2007/05/22.
441. Meshik X, O'Neill PR, Gautam N. Physical Plasma Membrane Perturbation Using Subcellular Optogenetics Drives Integrin-Activated Cell Migration. *ACS Synth Biol*. 2019;8(3):498-510. Epub 2019/02/16.

