

Université de Montréal

**Génétique fonctionnelle et validation biologique d'un locus
quantitatif d'expression lié en *trans*- à un réseau de gènes
impliqués dans l'immunité innée**

Par Saloua Jeidane

Département Biologie Moléculaire
Faculté de médecine

Thèse présentée à la Faculté de médecine
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D)
En Biologie Moléculaire

Mars, 2017

© Saloua Jeidane, 2017

Résumé

Contrairement aux maladies génétiques mendéliennes, qui dépendent d'un seul gène causal, les traits quantitatifs complexes sont des caractéristiques mesurables d'organismes vivants, qui résultent de l'interaction entre plusieurs gènes et les facteurs environnementaux. La génomique fonctionnelle nous a permis d'identifier de nombreux locus génétiques liés aux caractères complexes, qui sont appelés «locus de traits quantitatifs» (QTL). Cependant, de telles études ne permettent pas une caractérisation précise de l'architecture génétique des traits complexes. Plus récemment, il est devenu possible d'identifier des locus génétiques associés aux niveaux d'expression de gènes, appelés «locus de traits quantitatifs d'expression» (eQTLs). Dans de tels cas, les variants génétiques peuvent affecter l'expression soit des gènes qui se situent dans leur voisinage (*cis*-eQTLs), soit de ceux qui résident plus loin (*trans*-eQTLs). Dans des cas particuliers, un même locus peut affecter l'expression de plusieurs gènes situés dans différents chromosomes, formant ce qu'on appelle des '*trans*-eQTLs hotspots'. Ceux-ci peuvent avoir d'importants intérêts biologiques, car ils sont généralement enrichis en gènes fonctionnellement apparentés qui peuvent influencer le même trait phénotypique. Dans cette thèse, en analysant l'expression des gènes dans des échantillons de cœurs obtenus à partir d'un panel de souches consanguines recombinantes de souris AxB / BxA, nous avons détecté un QTL lié en *trans*- à l'expression de 190 transcrits, dont la majorité est connue pour être sensible aux interférons de type I. Le même locus correspondait également à celui d'un *cis*-eQTL pour le gène *Ypel5*, ce qui suggère que ce dernier peut être un régulateur commun des gènes *trans*-eQTL. Donc, le but principale de cette thèse fut de valider biologiquement le rôle du gène *cis*-eQTL dans la régulation du '*trans*-eQTLs hotspots'.

Les travaux présentés dans cette thèse ont montrés que la réduction de l'expression de *Ypel5* dans des macrophages de souris a stimulée l'expression de plusieurs gènes qui appartiennent au '*trans*-eQTL hotspot', et ce d'une manière dépendante d'IFN β 1. Le knockdown de YPEL5 a également augmenté l'induction d'IFN β 1 dans les cellules humaines HEK293T. Lorsque ces dernières ont été soumis à des stimuli qui activent les kinases TBK1 / IKK ϵ , nous avons détecté des interactions fonctionnelles de YPEL5 avec l'activité de ces kinases, ainsi que des interactions physiques avec IKK ϵ . Nos résultats préliminaires (présentés

dans le chapitre3) suggèrent aussi l'implication de YPEL5 dans la régulation du cycle cellulaire et /ou de la sénescence.

En conclusion, nous sommes parmi les premiers groupes à fournir des preuves biologiques montrant le rôle d'un gène *cis*-eQTL en tant que régulateur commun de gènes appartenant à un 'hotspot de *trans*-eQTL'. La validation biologique des analyses génomiques a ainsi permis de découvrir Ypel5 comme un nouveau régulateur négatif de la réponse antivirale innée qui agit (au moins en partie) au niveau des kinases TBK1 / IKBKE.

Mots-clés: '*trans*-eQTL hotspot', YPEL5, IFNB1, ISGs, TBK1, IKBKE

Abstract

Unlike Mendelian genetic diseases, which depend on a single causal gene, complex quantitative traits are measurable characteristics of living organisms, which result from the interaction between several genes and environmental factors. Functional genomics has allowed us to identify many genetic loci linked to complex traits, which are called "quantitative trait loci" (QTL). However, such studies do not allow an accurate characterization of the genetic architecture of complex traits. More recently, it has become possible to identify genetic loci associated with gene expression levels, called "expression quantitative trait locus" (eQTLs). In such cases, the genetic variants can affect the expression of genes that are either located in their vicinity (*cis*-eQTLs) or that reside further away (*trans*-eQTLs). In particular cases, the same locus can affect the expression of several genes located on different chromosomes, forming so-called '*trans*-eQTL hotspots'. These may have important biological interests, as they are generally enriched in functionally related genes, which may influence the same phenotypic trait. In this thesis, by analyzing the expression of genes in hearts from a panel of AxB / BxA mouse recombinant inbred strains, we detected a QTL linked in *trans*- to the expression of 190 transcripts, the majority of which are known to be sensitive to type I interferon. The same locus also corresponded to that of a *cis*-eQTL for the *Ypel5* gene, suggesting that it could be a common regulator of the *trans*-eQTL genes. Therefore, the main purpose of this thesis was to biologically validate the role of the *cis*-eQTL gene in the regulation of the '*trans*-eQTL hotspot'.

The work presented in this thesis showed that the silencing of *Ypel5* expression in mouse macrophages stimulated the expression of several genes that belong to the '*trans*-eQTL hotspot' in an IFNB1-dependent manner. *YPEL5* knockdown also increased IFNB1 induction in human HEK293T cells. When the latter were subjected to stimuli that activate the TBK1/IKBKE kinases, we detected functional interactions of YPEL5 with the activity of these kinases and physical interactions with IKBKE. Our preliminary results (presented in Chapter 3) suggest also the involvement of YPEL5 in the regulation of cell cycle progression and / or senescence.

In conclusion, we are among the first groups to provide biological evidence showing the role of a *cis*-eQTL gene as a common regulator of genes belonging to a '*trans*-eQTL hotspot'.

The biological validation of genomic analysis thus revealed Ypel5 as a new negative regulator of the innate antiviral response that acts (at least in part) at the level of the TBK1 / IKKε kinases.

Keywords: ‘*trans*-eQTL hotspot’, YPEL5, IFNB1, ISGs, TBK1, IKKε

Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	v
Liste des tableaux.....	ix
Liste des figures.....	x
Remerciements.....	xvi
CHAPITRE1 : INTRODUCTION.....	1
1.1 Introduction à la génétique.....	1
1.1.1 L'information génétique :.....	1
1.1.2 De la génétique classique à la génétique quantitative.....	2
1.1.3 La variation génétique.....	3
1.2 Les traits quantitatifs complexes.....	4
1.2.1 Les études des loci de trait quantitatif (QTL).....	6
1.2.2 Les méthodes statistiques pour la détection des QTL.....	8
1.2.2.1 Méthode de l'analyse de la variance.....	8
1.2.2.2 Méthode de cartographie par intervalles.....	8
1.2.3 Les approches génétiques pour la cartographie des QTLs.....	10
1.2.3.1 Les analyses de liaisons.....	10
1.2.3.2 Les études d'association génétique.....	11
1.3 Les modèles murins utilisés pour les études des QTL.....	12
1.4 Limitation des études des QTLs.....	17
1.5 Les QTL d'expression (eQTLs).....	18
1.5.1 Génotypage et estimation des niveaux d'expression des gènes.....	20
1.5.2 La régulation de l'expression génique.....	21
1.5.3 Les QTLs d'expression locale, ou 'Cis-eQTLs'.....	21
1.5.4 Les QTL d'expression distale ou 'Trans-eQTLs'.....	23
1.5.5 Les limitations des études des eQTLs.....	24

1.5.6 La nature moléculaire des eQTLs	25
1.6 Les ‘ <i>trans</i> -eQTL hotspots’	26
1.6.1 L’importance des ‘ <i>trans</i> -eQTL Hotspots’	27
1.6.2 La régulation des ‘ <i>trans</i> -eQTL hotspots’	28
1.6.3 Les ‘ <i>trans</i> -eQTL hotspots’ détectés jusqu’à présent	30
1.7 But du travail de la thèse	34
1.7.1 L’utilisation du iBMQ ‘ <i>R/Bioconductor package for integrated Bayesian modeling of eQTL data</i> ’ pour la cartographie des ‘hotspots de <i>trans</i> -eQTLs’	34
1.7.2 Le ‘ <i>trans</i> -eQTL hotspot’ dans le chromosome 17	35
1.7.2.1 L’IFN de type I et les ISGs	35
1.7.2.2 Les ‘ <i>trans</i> -eQTL hotspot’ enrichis en gènes immunitaires et leur association avec des maladies	43
1.7 Objectifs et hypothèses	44
CHAPITRE 2: ASSOCIATION OF A NETWORK OF INTERFERON-STIMULATED GENES WITH A LOCUS ENCODING A NEGATIVE REGULATOR OF NON-CONVENTIONAL IKK KINASES AND IFNB1	46
2.1 Summary	47
2.2 Introduction	48
2.3 Results	50
2.3.1 Characteristics of <i>cis</i> -eQTL and <i>trans</i> -eQTL genes	50
2.3.2 Effects of <i>Ypel5</i> knockdown on <i>trans</i> -eQTL genes in mouse macrophage cell lines	52
2.3.3 Effects of YPEL5 on IFNB1 induction in human cells	53
2.4 Discussion	56
2.5 Experimental Procedures	59
2.5.1 Animals and ethics statement	59
2.5.2 Cells lines and culture	59
2.5.3 Expression vectors	60
2.5.4 Structural protein homology analysis	60
2.5.5 Gene expression analyses	60

2.5.6 Luciferase assays	61
2.5.7 Western immunoblot analyses	61
2.5.8 Co-immunoprecipitation	61
2.5.9 Biotin-RNA pull down assay	62
2.5.10 Ablation of the <i>YPEL5</i> gene in HEK293T cells by RNA-guided nuclease technology	62
2.5.11 Statistics	63
CHAPITRE 3 : LE ROLE DE <i>YPEL5</i> DANS LA REGULATION DU CYCLE CELLULAIRE ET/OU DE LA SENESENCE.....	91
3.1 Introduction.....	91
3.2 Procédures expérimentales.....	93
3.3 Résultats.....	94
3.3.1 <i>YPEL3</i> et <i>YPEL5</i> sont les deux gènes <i>YPEL</i> les plus abondamment exprimés dans les cellules HEK293T et MCF7.	94
3.3.2 H ₂ O ₂ induit l'expression de <i>YPEL5</i> , <i>p21</i> , et <i>TNFα</i> , mais pas celle de <i>YPEL3</i> dans les cellules MCF7.....	94
3.3.3 Le KD de <i>YPEL5</i> augmente l'expression de <i>YPEL3</i> , <i>p21</i> , et <i>TNFα</i> dans les cellules MCF7.	95
3.3.4 Le KD de <i>YPEL5</i> amplifie l'induction de <i>p21</i> par H ₂ O ₂	95
3.3.5 Le KD de <i>YPEL5</i> augmente l'expression de gènes associés à la sénescence et/ou l'apoptose dans les cellules RAW264.7.....	96
3.4 Discussion	96
CHAPITRE 4 : DISCUSSION ET CONCLUSION.....	106
4.1 Discussion	106
4.1.1 La validation biologiquement du 'trans-eQTL hotspot' au ch17	107
4.1.2 Le choix du gène régulateur du 'trans-eQTL hotspot'	107
4.1.3 La fonction ou l'intérêt biologique du 'trans-eQTL hotspot' au chr 17.....	108
4.1.4 La découverte d'un nouveau rôle pour <i>YPEL5</i>	109
4.1.4.1 L'interaction de <i>YPEL5</i> avec TBK1 / IKKBE	109
4.1.4.3 L'interaction entre <i>YPEL3</i> et <i>YPEL5</i>	117

4.2 Conclusion	118
Bibliographie.....	i

Liste des tableaux

Tableau 1.I : Tableau récapitulatif des ‘<i>trans</i>-eQTL hotspots’ détectés jusqu’à présent.	31
Table S1a (related to Figure 2.1) Enrichment of Gene Ontology (GO) terms within <i>trans</i> -eQTL genes.....	71
Table S1b (related to Figure 2.1) Enrichment of Gene Ontology (GO) terms within genes whose expression is affected by Ypel5 silencing in RAW264.7 cells.....	71
Table S2: (related to Figure 2.1) List of <i>trans</i> -eQTL genes reported in the “interferome database” as ISGs.....	72
Table S3 (related to Figure 2.1) Comparisons of 50 <i>trans</i> -eQTL genes with largest fold differences in mouse hearts to genes affected by <i>Ypel5</i> silencing in RAW264.6 cells	75
Table S4 (related to the “Experimental Procedures” section) Primers used for RT-qPCR validation studies	77
Tableau 3.I : SiYpel5 dans les cellules RAW164.7 régule de nombreux gènes associés à la voie de sénescence cellulaire et ou aux voies de la mort cellulaire programmée (Illumina screen)	105

Liste des figures

Figure 1.1: Au cœur de chaque noyau de cellule eucaryote se trouve de l'ADN.....	2
Figure 1.2: La taille est un exemple de trait quantitatif complexe classique.	6
Figure 1.3: Définition d'un locus de trait quantitatif (QTL).....	7
Figure 1.4: Représentation schématique de la génération de souches recombinantes consanguines.	17
Figure 1.5: Résultats potentiels de la combinaison de plusieurs types de données moléculaires, y compris les données génotypiques, l'expression génique, et les données phénotypiques.	20
Figure 1.6 : Représentation schématique de 'cis-eQTL', 'trans-eQTL', et 'trans-eQTL hotspot'	24
Figure 1.7: Représentation graphique de l'analyse des eQTLs	27
Figure 1.8: La voie de signalisation d'IFN de type I.....	39
Figure 2.1: Chromosomal distribution of eQTL genes linked either in <i>cis</i> - or <i>trans</i> - to rs33200140 on chr17.....	63
Figure 2.2: Ypel5 negatively regulates trans-eQTL genes in RAW264.7 cells in an <i>Ifnb1</i> -dependent manner	64
Figure 2.3: Ypel5 downregulation amplifies the effects of either intracellular or extracellular poly (I:C) on <i>Ifnb1</i> induction.....	65
Figure 2.4: Effects of YPEL5 down-regulation or overexpression in HEK293T cells with or without SeV infection	66
Figure 2.5: YPEL5 siRNA potentiates the induction of IFNB1-dependent luciferase activity by signaling events that activate the TBK1/IKBKE kinases	68
Figure 2.6: YPEL5 silencing increased the effects of overexpressed IKBKE and TBK1 on their respective auto- or cross-phosphorylation	69
Figure 2.7: YPEL5 interacts physically with IKBKE in transfected cells.....	70
Figure S1 (related to Figure 2.1): Heatmap of expression levels of <i>trans</i> -eQTL genes across mouse cell types and tissues	78
Figure S2 (related to Figure 2.1): Characteristics of the YPEL5 protein	79
Figure S4 (related to Figure 2.4): Effects of SeV infection on HEK293T cells transfected with either scrambled or <i>YPEL5</i> -specific siRNA	83

Figure S5 (related to Figure 2.4): <i>In vitro</i> interaction of biotinylated HCV RNA and poly (I:C) with cell lysates of transfected cells.....	84
Figure S6 (related to Figure 2.5): <i>YPEL5</i> siRNA potentiates the effects of <i>TBK1/IKBKE</i> overexpression on endogenous <i>IFNB1</i> and <i>TNF</i> expression in HEK293T cells	85
Figure S7 (related to Figure 2.7): Reverse co-IP	86
Figure S8 (related to Figure 2.5): Interactions of <i>YPEL5</i> siRNA with the effects of IRF3 (5D) and rcIFNB1	87
Figure S9 (related to Figure 2.5): Two additional <i>YPEL5</i> -specific siRNAs have effects that are similar to that of si- <i>YPEL5</i> -1	88
Figure S10 (related to Figure 2.5): Ablation of the <i>YPEL5</i> gene by RNA-guided nuclease in HEK293T cells.....	89
Figure 3.1 : L'expression relative de <i>YPEL1-5</i>	100
Figure 3.2 : (A) H ₂ O ₂ induit <i>YPEL5</i> , <i>YPEL3</i> , <i>p21</i> , et <i>TNFA</i> dans des cellules MCF7	101
Figure 3.3 : Les si- <i>YPEL5</i> et si- <i>YPEL3</i> amplifient l'induction d' <i>IFNB1</i> (A) et de <i>TNFα</i> (B) dans des cellules HEK 293T, avec une transfection des deux si-ARN ayant un effet additif	102
Figure 3.4 : le Knockdown de <i>YPEL5</i> amplifie l'induction de <i>p21</i> par le traitement de H ₂ O ₂	104

Liste des sigles

ABIN1	<i>A20 binding inhibitor of NF-κB 1</i>
ADN	acide désoxyribonucléique
ARN	acide ribonucléique
ARNdb	ARN bicaténaire
ARNm	ARN messenger
ARNsc	ARN monocaténaire
ANOVA	analyse de la variance
CC	<i>collaborative cross</i>
CDCV	maladie commune / variante commune
cDNA	<i>complementary DNA</i>
CDK11	<i>cyclin-dependent kinase 11</i>
ChIP-seq	immunoprécipitation de la chromatine suivie d'un séquençage à haut débit
Chr	chromosome
CIM	<i>composite interval mapping</i>
cGAS	<i>GMP-AMP cyclique synthase</i>
CYLD	<i>deubiquitinating enzyme cylindromatosis</i>
DDR	<i>DNA damage response</i>
eQTL	<i>expression QTL</i>
ER	réticulum endoplasmique
GAS	<i>Gamma interferon activation site</i>
GO	<i>Gene Ontology</i>
GWAS	<i>genome wide association studies</i>
HIPK2	<i>homeodomain-interacting protein kinase 2</i>
H ₂ O ₂	peroxyde d'hydrogène
iBMQ	<i>integrated Bayesian modeling of eQTL data</i>
IDIN	<i>IRF7-driven inflammatory network</i>
IFNB1	interferon- β
IFIT1	<i>Interferon-induced protein with tetratricopeptide repeats 1</i>
IFN γ	interféron γ

IGFBP-3 *insulin-like growth factor binding protein-3*
IKBKE *inducible I κ B kinase*
IM *interval mapping* IFN *interférons de type I*
IRF-3 *IFN-regulatory factor 3*
IRCM *Institut de Recherches Cliniques de Montréal*
ISGs *interferon-stimulated genes*
ISGF3 *IFN-stimulated gene factor 3*
ISG56 *IFN-stimulated gene 56*
ISRE *interferon-sensitive response elements*
JAK1 *Janus kinase 1*
LD *déséquilibre de liaison*
LGP2 *laboratory of genetics and physiology 2*
LOD *Logarithm of the odds*
LPS *lipopolysaccharides*
MAF *fréquence des allèles mineurs*
MAVS *mitochondrial antiviral signaling*
MDA5 *melanoma differentiation-associated gene 5*
MIB1 *E3 ubiquitin ligases mind bomb 1*
MIM *multiple interval mapping*
NF- κ B *nuclear factor-kappa B*
NIK *NF- κ B-inducing kinase*
NK *natural killer*
OASL1 *2'-5'-oligoadenylate synthase-like protein 1*
OMIM *Online Mendelian Inheritance in Man*
OS *outbred stocks*
PAMP *pathogen-associated molecular patterns*
PKR *protéine kinase R*
PLK1 *Polo-like kinase 1*
PP2C β *phosphatase Mg²⁺/Mn²⁺ dependent 1B*
PRRs *pattern-recognition receptors*
PTP-2C *protein-tyrosine phosphatase 2C*

QTL *quantitative trait locus*
RanBPM *Ran Binding Protein in the Microtubule organizing center*
RIG-I *retinoic acid-inducible gene-1*
RIS souches recombinantes consanguines
RLR RIG-I like receptors
ROS dérivées réactives de l'oxygène
RT-qPCR *real-time polymerase chain reaction*
SASP *senescent associated secretory phenotype*
SeV *Sendai virus*
siRNA *small interfering RNA*
SLE *lupus érythémateux disséminé*
SIKE *suppressor of IKBKE*
SNP *single nucleotide polymorphism*
SOCS *Suppressor of cytokine signaling 3*
STAT1 *the latent cytoplasmic transcription factors signal transducer and activator of transcription 1*
STING *stimulator of IFN genes*
TBK1 *(TANK)-binding kinase 1*
T2D diabète de type 2
TF facteurs de transcription
TLR récepteurs Toll-like
TNF α facteur de nécrose tumorale alpha
TRAF3 *TNF receptor-associated factor3*
TRIF *TLR/IL-1R domain-containing adaptor protein inducing IFN- β*
TYK2 *tyrosine kinase 2*
USP18 *ubiquitin specific peptidase 18*
WTCCC *Wellcome Trust Case Control Consortium*
Ypel5 *Yipee like 5*

À mes parents

Remerciements

Je suis extrêmement reconnaissante à toutes les personnes qui ont fait de mon doctorat une si grande expérience. Tout d'abord, je tiens à exprimer mes remerciements spéciaux à mon directeur de recherche, Dr. Christian Deschepper, pour ses directions importantes, ses conseils et son constant soutien. Son enthousiasme pour la recherche et son éthique de travail ont été une grande source d'inspiration, et j'ai beaucoup appris de lui. Je remercie également tous les membres du laboratoire de Dr. Christian Deschepper, avec qui j'ai partagé plein de moments agréables. Un grand merci à Sylvie Picard, Marie Pier Scott-Boyer, Sophie Cardin, et Samantha Praktijnjo. Je voudrais particulièrement remercier Sylvie Picard pour son inestimable soutien, ses conseils et sa patience.

Mon projet de doctorat repose en partie sur une collaboration très productive avec le laboratoire de Dr. Daniel Lamarre au centre de recherche du centre hospitalier de l'université de Montréal (CRCHUM). Donc, je tiens à exprimer mes sincères remerciements à Dr. Daniel Lamarre, ainsi que tous les membres de son laboratoire, et particulièrement Martin Baril, Nicolas Tremblay, Salwa Es-Saad, et Alex Parc, qui ont été de grandes sources d'inspiration et d'aide. Je voudrais remercier Martin Baril pour les nombreuses discussions fructueuses sur ce projet. J'aimerais remercier également les responsables du département de biologie moléculaire à l'université de Montréal: Dr. Jean François Coté, Dr Richard Bertrand, et Madame Liu Yan. Un grand merci à la communauté de l'institut de recherche clinique de Montréal (IRCM) et particulièrement Virginie Leduc. Son aide dans toutes les affaires administratives a été inestimable. Pendant ces cinq années, j'ai eu la chance de faire de très chers (es) amis (es) à l'IRCM et au CRCHUM, ce qui a fait de ces deux endroits une place incroyable pour faire mon doctorat. Bien qu'il y en ait trop pour être listés ici, quelques noms méritent une mention spéciale: merci beaucoup à Salma Chehboun, Omar El Rifai, Yacine Kherdjemil, et Nehme el Hachem pour être de si grands amis. Enfin, et surtout, je suis extrêmement reconnaissant à tous les membres de ma famille: mes parents, mes sœurs et mon frère qui ont toujours cru en moi et qui m'ont soutenu. Sans leurs efforts et leurs sacrifices, cette thèse aurait été impossible. Je voudrais exprimer également ma gratitude à mon cher mari Saadi Hermas pour son soutien et sa patience.

CHAPITRE1 : INTRODUCTION

1.1 Introduction à la génétique

1.1.1 L'information génétique :

Chez tous les êtres vivants, l'information génétique est contenue dans une macromolécule biologique appelée «ADN» (acide désoxyribonucléique). La première preuve expérimentale qui montre que l'ADN constitue le matériel génétique fut introduite par Avery, MacLeod et McCarty, qui ont montré que les caractéristiques génétiques chez les bactéries pourraient être modifiées d'un type à l'autre par un traitement avec de l'ADN purifié ¹. Par la suite, la structure tridimensionnelle de l'ADN, proposée en 1953 par Watson et Crick, a donné plusieurs indices sur la manière dont l'ADN fonctionne ^{2,3}. Une molécule d'ADN est formée principalement d'un double hélice, où chaque brin est composé d'une chaîne nucléotidique formée essentiellement de quatre bases nucléiques ou de bases azotées : l'adénine (A), la cytosine (C), la guanine (G), et la thymine (T). L'ordre ou la séquence de ces bases détermine l'information disponible pour le fonctionnement d'un organisme. Dans le noyau des cellules eucaryotes, la molécule d'ADN est emballée dans des structures filamenteuses appelées chromosomes. Tous les organismes diploïdes portent deux sets de chromosomes homologues (Figure 1.1).

Un gène est un fragment d'ADN qui contient l'information génétique à partir de laquelle d'autres macromolécules (telle que les protéines, constituées d'une chaîne d'acides aminés) sont synthétisées. Selon la théorie fondamentale de la biologie moléculaire, décrite par Francis Crick en 1958, le transfert de l'information génétique de l'ADN vers la protéine se fait comme suit : tout d'abord, l'ADN est transcrit en une chaîne d'acides ribonucléiques appelée «ARN», qui est ensuite traduite en protéine. Plus récemment, il fut établi que le génome contient beaucoup d'autres gènes, appelés 'non-codants' (car ils ne codent pas pour des protéines). L'ARN qui code pour les protéines est appelé ARN messager «ARNm» ⁴. Un gène est l'unité physique et fonctionnelle de base de l'hérédité. Les allèles sont des éléments génétiques codant pour le même gène mais contenant de petites différences dans leur séquence de bases d'ADN (variation génétique). Ces petites différences contribuent aux caractéristiques physiques uniques de chaque individu ⁵.

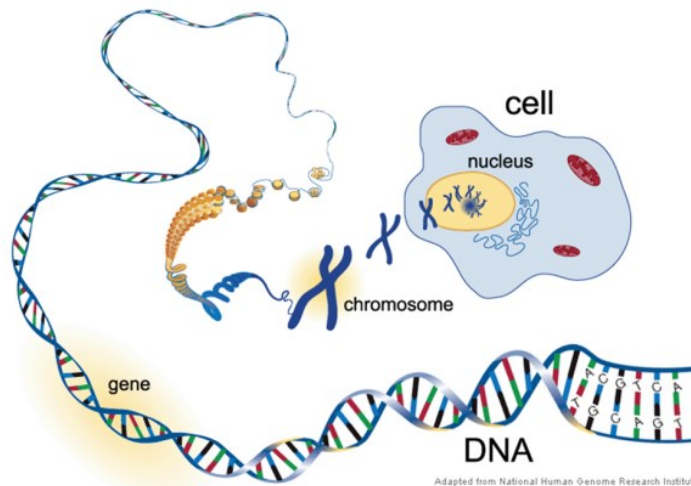


Figure 1.1: Au cœur de chaque noyau de cellule eucaryote se trouve de l’ADN. La double hélice de l’ADN contient l’information génétique sous forme d’une séquence de nucléotides. L’ADN est organisé en chromosomes. Source: <https://ratedrtech.wordpress.com/tag/dna/>

1.1.2 De la génétique classique à la génétique quantitative

Dans l’ouvrage “On the Origin of Species” (Darwin, 1859) ⁶, Charles Darwin a proposé la théorie de l’évolution qui repose sur les principes selon lesquels la variation naturelle entre les individus offre des avantages reproductifs différentiels et que cette variation est héréditaire. Bien que sa théorie puisse expliquer magnifiquement l’adaptation d’une population à son environnement et le développement de nouvelles espèces, les mécanismes par lesquels une telle variation pourrait se produire et la façon dont elle pourrait être transmise de génération en génération n’était pas claire à l’époque.

En 1853, Gregor Mendel avait commencé à travailler exactement sur ce problème. Ses recherches ont conduit Mendel à proposer ses fameuses ‘lois de l’hérédité’: la loi de ségrégation (qui décrit comment chaque individu contient une paire d’allèles pour un trait donné, dont l’un est transmis au hasard à sa progéniture) et la ‘loi de l’assortiment indépendant’ (qui décrit comment différents traits sont hérités indépendamment l’un de l’autre).

Ces découvertes ont suscité une nouvelle ère de la biologie, menant à des tentatives de trouver à la fois les mécanismes moléculaires qui pourraient entraîner des variations génétiques ainsi que les mécanismes qui pourraient conduire à leur transmission. À l’aide de ses étudiants (Hermann Muller, Alfred Sturtevant et Calvin Bridges), Morgan a pu élaborer l’effet des

mutations chez la *Drosophile*. Ceci a permis de découvrir plusieurs principes importants de la génétique, qui sont encore pertinents aujourd'hui. Parmi ceux-ci était l'observation que la progéniture de mouches femelles portant deux mutations situées sur deux copies distinctes du chromosome X (chr X), portaient parfois les deux mutations sur la même copie du chr X. Ceci l'amène à proposer le concept de « l'entrecroisement » ou 'crossing over', l'échange de matériel génétique entre chromosomes homologues pendant la méiose ⁷. Morgan a également raisonné que le degré de liaison entre deux régions (gènes) serait relatif à leur distance linéaire sur le chromosome. Lui et ses élèves ont utilisé ce phénomène pour développer la technique de la cartographie des gènes, en utilisant le taux de recombinaison entre les différents traits pour estimer les distances relatives des gènes les uns des autres. La première carte génétique, qui décrit l'arrangement de gènes sur le chr X, a été publiée en 1913 (Sturtevant 1913) ⁸.

Sept ans plus tard, Altenburg (un autre étudiant de Morgan) et Muller ont continué à utiliser la liaison génétique pour montrer qu'une mutation conduisant à des ailes tronquées était effectivement causée par des facteurs multiples sur différents chromosomes, tous hérités selon la loi mendélienne (Altenburg et Muller, 1920) ¹⁰. Cette découverte est un exemple du concept de traits quantitatifs, qui a été formalisé par le statisticien Ronald A. Fisher en 1918. En s'appuyant sur les travaux de Galton et Pearson, Fisher a décrit comment l'hérédité mendélienne pouvait entraîner la variation continue d'un trait (Fisher, 1918) ¹¹. Ce travail a introduit non seulement les concepts de variance et d'analyse de variance (ANOVA), mais a également jeté les bases du concept de traits quantitatifs et de génétique quantitative.

1.1.3 La variation génétique

Lorsque Darwin a proposé sa théorie de l'évolution, il n'était pas clair quels processus étaient responsables de la variation des populations. Aujourd'hui, nous savons bien sûr que le principal facteur de variation biologique est la mutation de l'ADN. Les mutations peuvent survenir spontanément, mais peuvent être aussi induites par des facteurs environnementaux, tels que l'exposition à différentes plages de température, ainsi que les rayons X et le radium (Muller, 1927) ¹².

Les mutations ou les variations génétiques sont appelées aussi polymorphismes. Le type de polymorphisme le plus courant est le polymorphisme d'un seul nucléotide [ou « *single*

nucleotide polymorphism » (SNP)], dans lequel une simple base d'ADN diffère entre les individus d'une population ¹³. Selon leur position dans le génome, les SNPs peuvent être non codants ou codants. Les SNPs codants ou non-synonymes (nsSNPs) sont ceux qui conduisent à un changement d'un acide aminé. Les SNPs synonymes sont ceux qui ne modifient pas la séquence d'acides aminés de la protéine codée. Beaucoup de SNPs n'ont aucun effet sur la fonction cellulaire. Par contre, les polymorphismes qui affectent les niveaux d'expression de divers gènes et/ou la fonction de protéines peuvent affecter ainsi divers traits de l'individu. Un caractère ou «un trait», pour un organisme vivant, peut résulter d'une variation génétique donnant naissance à une manifestation observable. Ces manifestations phénotypiques peuvent avoir parfois un caractère pathologique et/ou représenter une maladie. La relation entre la variation génétique et la manifestation phénotypique peut être soit simple, soit compliquée (voir section prochaine). La génétique est la science qui nous permet d'étudier cette relation, et de trouver la source du changement de caractères génétiques observables des organismes vivants ainsi que leur mécanisme de transmission héréditaire.

1.2 Les traits quantitatifs complexes

Une tâche commune en génétique est d'identifier les facteurs génétiques intervenant dans le déterminisme d'un caractère particulier. Historiquement, cette tâche a été résolue par la cartographie génétique, telle qu'introduite par Sturtevant en 1913. En utilisant des variations naturelles de l'ADN comme marqueurs génétiques ¹⁴, il a été possible d'effectuer des analyses de cartographie génétique chez les vertébrés. De telles études ont facilitées la découverte de facteurs génétiques qui sous-tendent les maladies à transmission mendélienne telles que la maladie de Huntington ¹⁵, ou la fibrose kystique ¹⁶.

Ces maladies se caractérisent par la présence d'une variation génétique unique ayant un effet délétère et suffisamment important pour provoquer des phénotypes remarquablement pénétrants et pathogènes. En général, les traits fortement influencés par la variation au sein d'un seul gène sont appelés traits monofactoriels ou monogénétiques. Les maladies dites 'mendéliennes' représentent des exemples de maladies monogéniques héréditaires, avec un mode de transmission qui suit les lois de Mendel ¹⁷. Comme les maladies monogéniques illustrent les effets biologiques résultant de la perturbation de gènes individuels, cette classe de maladie génétique a été inestimable pour aider à améliorer notre compréhension de l'architecture

génétiq ue de nombreuses maladies cliniquement importantes, actuellement documentés dans la base de données OMIM (*Online Mendelian Inheritance in Man*) ¹⁸.

Bien que les traits monogéniques forment la base de la génétique dite ‘classique’ et montrent des manifestations cliniques importantes, ils sont relativement rares et se manifestent dans un nombre restreint de familles ¹⁷. La plupart des traits de pertinence médicale ne suivent pas l'hérédité monogénique mendélienne simple. Ces «traits complexes» incluent les susceptibilités aux maladies cardiaques, à l'hypertension, au diabète, au cancer, et à l'infection.

Le terme «trait complexe» désigne tout phénotype qui ne présente pas d'hérédité mendélienne classique. En général, la complexité apparaît lorsque la correspondance simple entre le génotype et le phénotype se dissocie, soit parce que le même génotype peut aboutir à des phénotypes différents (d'où un gène peut montrer une pléiotropie en raison d'interaction avec différents conditions environnementaux), ou parce que différents génotypes peuvent contribuer conjointement au même phénotype (où un ou plusieurs gènes différents peuvent interagir) ¹⁹.

Les caractères polygéniques peuvent être classés soit comme des traits distincts, mesurés par une réaction spécifique (par exemple, le développement du diabète de type I ou l'infarctus du myocarde), soit comme des traits quantitatifs, mesurés par une variable continue [par exemple, la taille, la tension artérielle diastolique, ou les titres d'immunoglobuline E (IgE)]. La variance de ces caractères résulte en fait de l'action combinée de nombreux facteurs génétiques et non génétiques (Figure 1.2) ²⁰. Cependant, les variants génétiques contribuant aux maladies complexes ont des caractéristiques particulières qui ont un impact sur les stratégies de cartographie, car en général : 1) ils ne contribuent que de façon modeste à la variance du trait complexe; 2) comme leur effet sur la physiologie de l'organisme n'est pas très sévère, leur fréquence au sein des populations reste en général relativement élevée ²¹; et 3) leur impact est modulé par des interactions fréquentes avec des composantes environnementales ²².

La dissection génétique de traits complexes intéresse de nombreux chercheurs et concerne différents problèmes, tant dans le domaine de la génétique médicale, que d'autres disciplines, incluant les études fondamentales sur le développement des mammifères et le travail appliqué dans l'amélioration de l'agriculture.

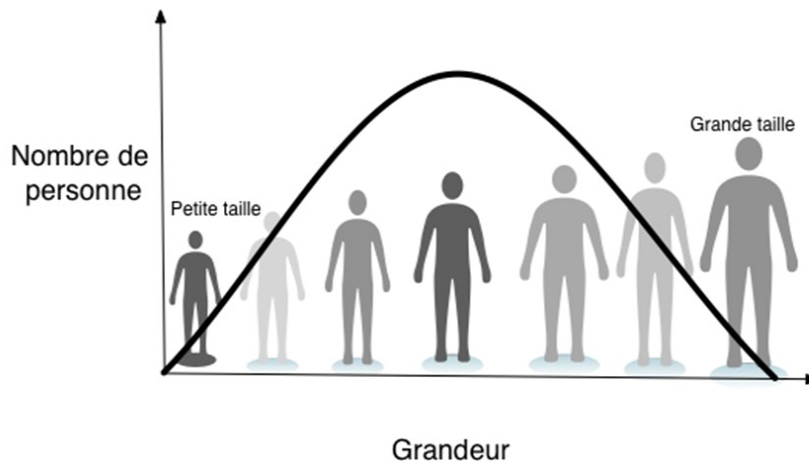


Figure 1.2: La taille est un exemple de trait quantitatif complexe classique. La distribution du phénotype suit une distribution normale, ce qui est caractéristique des traits complexes. Bien que plusieurs facteurs environnementaux (comme l'âge à la puberté, l'environnement périnatal et l'apport nutritionnel) aient affecté la taille adulte atteinte, jusqu'à 90% de sa variation est attribuée à des facteurs génétiques ou héréditaires ^{23,25}.

1.2.1 Les études des loci de trait quantitatif (QTL)

La génétique quantitative utilise les études de cartographie de 'locus de caractères quantitatifs', connue sous le terme QTL (*quantitative trait locus*), pour élucider la variation des caractères quantitatifs et leur base génétique. Un QTL est une région chromosomique qui est susceptible de contenir des facteurs génétiques contribuant (au moins en partie) à la variabilité des caractères phénotypique à l'étude (Figure 1.3). L'analyse des QTLs permet d'accumuler un grand nombre d'informations tel que : le nombre minimum de locus en cause, leur position dans le génome, leurs effets sur la variation du caractère, leurs caractéristiques génétiques (additivité, dominance, épistasie...), et leur pénétrance.

Dans l'approche de la génétique quantitative, un QTL individuel peut être disséqué en utilisant des marqueurs génétiques liés. Cette approche a été établie depuis longtemps ^{26,30}, mais jusqu'à récemment, elle a été considérée comme d'importance mineure en raison du manque de marqueurs génétiques suffisants. Grâce à la biologie moléculaire moderne, cette situation a été radicalement modifiée. La capacité de détecter la variation génétique directement au niveau de l'ADN a abouti à une offre essentiellement infinie de marqueurs pour tout organisme d'intérêt

¹⁴. Il n'est pas surprenant qu'il y ait eu une explosion dans l'utilisation de méthodes basées sur des marqueurs en génétique quantitative.

Les marqueurs génétiques sont des séquences d'ADN polymorphes dont la localisation sur le génome est connue. Les marqueurs les plus utilisés dans les cartes génétiques aujourd'hui sont les SNPs. Les SNPs sont hautement abondants et répartis dans tout le génome (ils représentent environ 90% de toutes les variations génétiques humaines, et se produisent toutes les 100 à 300 bases le long du génome humain) ³¹. Les SNPs sont plus stables du point de vue évolutionniste parce qu'ils ne changent pas beaucoup d'une génération à l'autre, ce qui facilite leur utilisation dans les études de population. En conséquence, les SNPs sont fréquemment utilisés dans les études d'association et dans les analyses de liaison.

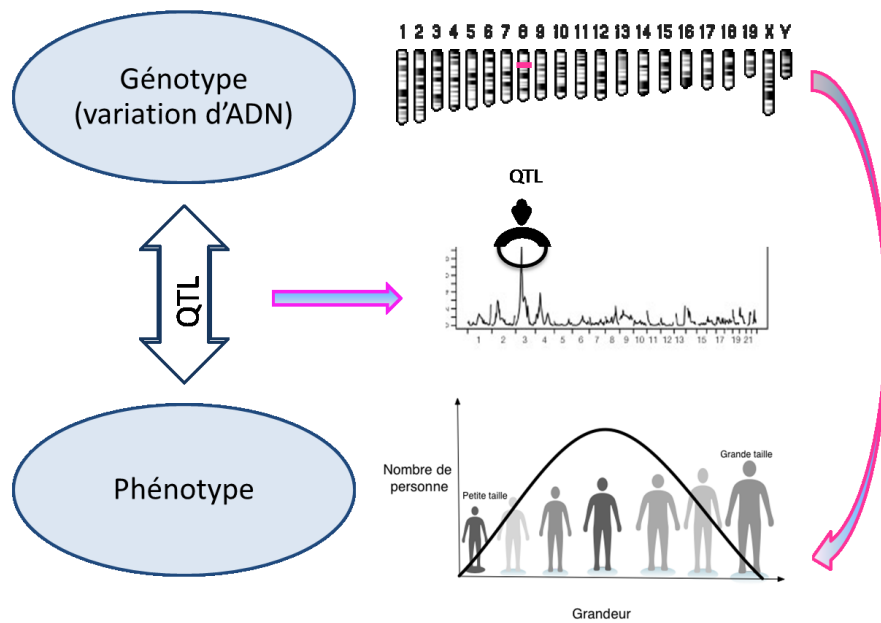


Figure 1.3: Définition d'un locus de trait quantitatif (QTL). Un QTL est une section d'ADN (un locus) qui est lié à, ou contient, les gènes qui contrôlent la variation d'un caractère quantitatif.

1.2.2 Les méthodes statistiques pour la détection des QTL

1.2.2.1 Méthode de l'analyse de la variance

La méthode la plus simple pour la cartographie de QTL est l'analyse de la variance (ANOVA, parfois appelée «régression par marqueur»), qui étudie l'effet des marqueurs sur les valeurs des phénotypes. Pour chaque marqueur génétique, on divise les individus d'une population en deux ou trois groupes, en fonction de leurs génotypes respectifs (du marqueur), et on compare les distributions phénotypiques des groupes. Lorsqu'un marqueur présente un génotype qui co-ségrègue de façon significative avec la distribution phénotypique dans la population, on dit que ce marqueur est «lié» à un QTL. Par exemple, si un groupe d'individus, ayant un génotype AA à un marqueur donné, ont des valeurs phénotypiques un peu plus élevées que ceux avec les génotypes AB et/ ou BB au niveau de ce marqueur, cela indique que le marqueur est lié à un QTL.

L'analyse de la variance possède plusieurs avantages. Par exemple, cette méthode ne nécessite pas une carte génétique pour les marqueurs; elle peut être facilement étendue pour tenir compte de multiples loci. Un autre avantage est l'inclusion facile de co-variables, tels que le sexe, un traitement ou un effet environnemental. L'approche ANOVA pour la cartographie de QTL a trois faiblesses importantes^{29,32}. Premièrement, elle ne donne pas d'estimations séparées pour l'emplacement du QTL et l'effet du QTL. Deuxièmement, il faut éliminer les individus dont les génotypes manquent au marqueur. Troisièmement, lorsque les marqueurs sont largement espacés, le QTL peut être assez éloigné des marqueurs utilisés, ce qui diminue la puissance de détection du QTL.

1.2.2.2 Méthode de cartographie par intervalles

Quand les marqueurs sont considérés individuellement dans la cartographie de QTL, les effets sont sous-estimés et la position ne peut pas être déterminée. Afin de remédier à ces inconvénients, Lander et Botstein (1989) ont introduit la cartographie par intervalles, connue sous l'acronyme IM (*interval mapping*)²⁹, comme une méthode systématique pour analyser le génome entier à la recherche des QTLs. Il a été démontré que la cartographie par intervalles a plus de puissance que les méthodes d'analyse directe des marqueurs^{26,29}.

La Méthode de IM utilise des marqueurs adjacents (flanquants) pour construire un intervalle, où l'existence d'un QTL va être testée. Chaque emplacement dans le génome est considéré comme un emplacement putatif de QTL. Étant donné que le génotype des marqueurs adjacents est connu, on peut calculer la probabilité qu'un individu ait un génotype à un QTL donné. La liaison entre le QTL et le phénotype est mesuré avec un score appelé '*Logarithm of odds*' (LOD), qui constitue le logarithme du rapport de la probabilité de la présence d'un QTL à la position donnée sur la probabilité de l'absence de QTL dans le génome.

L'approche d'IM considère un QTL à la fois dans le modèle de cartographie de QTL. Par conséquent, s'il y'a plus d'un QTL sur un chromosome, la statistique de test à la position testée sera affectée par ces autres QTLs, donc les positions estimées et les effets de QTL identifiés par cette méthode seront susceptibles d'être biaisés^{26,29,33,34}. L'une des raisons de ces lacunes est que le test utilisé dans la méthode de cartographie par intervalles n'est pas un test d'intervalle. Un test d'intervalle (qui teste l'effet du QTL dans un intervalle défini) devrait être indépendant des effets de QTL en dehors de la région. Sinon, même s'il n'y a pas de QTL dans un intervalle, le profil de vraisemblance sur l'intervalle peut encore dépasser le seuil de façon significative s'il existe un QTL dans une région voisine sur le même chromosome. Afin de surmonter la lacune de la méthode de cartographie d'intervalles, Zeng (1994) a proposé une méthode améliorée appelée cartographie d'intervalle composite, connue sous l'acronyme CIM (*composite interval mapping*), en combinant la cartographie d'intervalle avec l'analyse de régression multiple³⁴.

Bien que intuitives et largement utilisées, ces méthodes sont encore insuffisantes pour étudier l'architecture génétique des traits quantitatifs complexes qui sont affectés par de multiples QTLs. Lorsqu'un trait est affecté par plusieurs loci, il est plus efficace de rechercher statistiquement cet ensemble de QTLs. Dans ce contexte, une méthode étendue appelée cartographie par multiples intervalles, connue sous l'acronyme MIM (*multiple interval mapping*) a été proposée³⁵. Cette méthode s'adapte à tous les QTLs dans le modèle, et a la capacité d'analyser l'épistasie des QTLs.

1.2.3 Les approches génétiques pour la cartographie des QTLs

1.2.3.1 Les analyses de liaisons

Les analyses de liaison visent à identifier une relation explicite entre le phénotype et la transmission des régions chromosomiques à travers des familles contenant les individus touchés. L'objectif est d'identifier des marqueurs dont les modèles de transmission sont en corrélation avec le phénotype ³⁶.

Le point de départ d'une analyse de liaison est d'abord de réunir une population expérimentale génétiquement diverse constituée à partir de plusieurs pédigrées familiaux couvrant plusieurs générations. Au sein de telles populations, une « carte génétique » peut être construite en se basant sur la fréquence de recombinaison génétique de chaque marqueur, ce qui permet de déterminer leur position relative au sein du génome. Puisque l'emplacement de chaque marqueur est connu, il est possible d'identifier le locus d'une maladie en identifiant les marqueurs qui co-ségrèguent avec la manifestation phénotypique au sein de la population. Il faut noter que lorsque plusieurs SNPs sont proches l'un de l'autre, ils peuvent être liés en blocs, formant ce qu'on appelle des haplotypes, c'est-à-dire des régions de chromosome où la recombinaison est limitée, et où tous les SNPs sont transmis de façon conjointe.

Les premières études de cartographie génétique chez l'homme ont utilisé la cartographie de liaison ¹⁴. Chez les humains, les études de liaison ont réussi à montrer que des milliers de maladies mendéliennes résultent de la présence des variants génétiques à forte pénétrance (c'est-à-dire de variants qui, lorsque présents, s'accompagnent systématiquement d'une manifestation phénotypique détectable). Contrairement aux traits monogéniques, les approches de liaison se sont avérées moins appropriées pour élucider la génétique des traits complexes. Seulement les variants fréquents, associés à la variation de maladies courantes ("maladie commune / variante commune" (CDCV)), ont été identifiés dans ce genre d'études, par exemple le facteur VLeiden dans la thrombose veineuse profonde ³⁷, l'allèle APOEε-4 dans la maladie d'Alzheimer ¹⁵, et le PPARγ dans le diabète de type 2 ³⁸. Mais, dans le cas de traits complexes où des variations génétiques de faible effet présentent une pénétrance incomplète ou des effets pléiotropes, les études de liaison sont insuffisantes.

1.2.3.2 Les études d'association génétique

Une autre méthode pour l'étude génétique des maladies communes est représentée par les études d'association génétique, qui comparent dans une population la fréquence des variants génétiques entre individus affectés ou non affectés d'un trait particulier ³⁹. La distinction importante est qu'il n'y a pas de structure familiale présente, et les individus sont supposés être indépendants "non apparentés".

Une hypothèse importante à la base des études d'association est que dans des populations composées d'individus non apparentés, la majorité des variant génotypiques seront répartis de façon aléatoire par rapport au phénotype, même s'ils sont situés sur le même chromosome que la variation génétique causale. Les variants liés au génotype seront uniquement ceux qui sont très proche du locus causal. Lorsque deux loci sont associés de façon non-aléatoire au sein d'une population, on dit qu'ils sont en déséquilibre de liaison, ce qui est appelé en anglais « linkage disequilibrium » (LD).

Initialement, un des défis principaux des études d'association génétique était de pouvoir obtenir des cartes génétiques suffisamment denses ^{40,41}. Une autre difficulté était liée à la structure de la population, un facteur responsable de faux positifs ⁴². Ce n'est que lorsque des techniques telles que les microarrays d'ADN ont permis le génotypage à haut débit de nombreux variants qu'il est devenu possible de dépasser ces limites en étendant les études d'association génétique à l'ensemble du génome.

Les études d'association à l'échelle du génome, connue sous l'acronyme GWAS (*genome wide association studies*), ont aidé à comprendre la base génétique de plusieurs troubles complexes. Les cartes de variation génétique humaine bien documentées ^{43,44}, le grand nombre d'échantillons de patients avec les informations phénotypiques bien enregistrées, ainsi que les méthodes statistiques appropriées pour évaluer la signification, ont tous contribué au succès actuel des GWAS. La première GWAS chez l'homme, sur la dégénérescence maculaire liée à l'âge (AMD), a été publiée en 2005 ⁴⁵. Deux ans plus tard, le 'Wellcome Trust Case Control Consortium' (WTCCC) publiait son rapport historique sur les résultats de GWAS sur sept maladies différentes ⁴⁶. Depuis l'apparition de la phase II HapMap, (qui caractérise plus de 3,1 millions de SNPs humains) ⁴⁴, de nombreuses variations génétiques associées à un grand nombre de maladies complexes ont été rapportées et effectivement répliquées. Un catalogue actuel de

GWAS publié (<http://www.genome.gov/26525384>) comprend des SNPs non seulement associés aux principaux troubles courants [diabète de type 2 ^{47,48}, cancer de la prostate ⁴⁹, cancer du sein ⁵⁰, cancer du poumon ⁵¹ etc.], mais aussi à des traits quantitatifs [par ex. Indice de masse corporelle ⁵², ou la taille ⁵³].

L'étude d'association est la méthode la plus puissante à ce jour pour la détection de variants communs, à condition que les séquences causales soient en fort LD avec les SNPs génotypés ⁵⁴. De plus, avec un génotypage suffisamment dense, la cartographie d'association est plus susceptible de détecter des variants qui possèdent chacun un effet modéré sur le phénotype étudié (fréquence des allèles mineurs, MAF > 0,05) ³⁶. Bien que cette approche détecte rarement des variants fonctionnels, la résolution de la cartographie génétique est beaucoup plus fine par rapport aux études de liaison. Les études de liaison identifient généralement des régions génomiques de 5 à 10 Mb contenant des dizaines voire des centaines de gènes, alors que les études d'association sont en mesure de raffiner les loci génomiques à environ 10 à 100 kb.

1.3 Les modèles murins utilisés pour les études des QTLs

Étant donné les difficultés rencontrées pour étudier la physiopathologie ou la base génétique des maladies directement chez l'homme, les modèles animaux se sont avérés précieux pour les études de caractères complexes, y compris l'athérosclérose ^{55,56}, l'insuffisance cardiaque ^{57,58}, le diabète ^{59,60}, l'obésité ^{61,62}, l'ostéoporose ^{63,64}, le cancer et les troubles du comportement ⁶⁵. L'hypothèse de ces études est que même si les inférences ne se traduisent pas directement entre les modèles animaux et les humains, les voies qui contribuent aux pathologies seront partagées ⁶⁶. Effectivement, malgré les différences importantes, les modèles animaux ont généralement permis une meilleure compréhension à la fois de la biologie de base et de la pathologie. ^{67,70}.

Pour les besoins de mon projet, je vais restreindre la discussion à la description des modèles murins. Ces modèles ont de nombreux avantages : 1) ils sont très similaires à l'homme (ils partagent environ 98% de leur ADN codant avec les humains ⁷¹); 2) ils constituent une ressource accessible et relativement économique, ils sont de faible taille, et ont un cycle de vie court; ceci nous permet de faire des études au cours de leur cycle de vie, ou même à travers plusieurs générations ⁷²; 3) ils permettent d'éviter les problèmes éthiques que soulèverait

l'expérimentation chez les humains; 4) ils offrent des avantages de reproductibilité des méthodes expérimentales; 5) ils sont relativement faciles à gérer, car la plupart des facteurs environnementaux sont contrôlables ; et 6) ils sont accessibles, car des tissus et des organes peuvent être isolés pour des recherches expérimentales détaillées.

Les souris modifiées génétiquement (transgéniques, *knock-out* et *knock-in*) représentent des outils de choix qui ciblent spécifiquement l'expression de certains gènes ou transgènes. Grâce à ces modèles, il a été possible d'examiner les effets des perturbations de gènes uniques sur un fond génétique bien caractérisé, et d'identifier les mécanismes moléculaires (préalablement non soupçonnés) impliqués dans l'initiation et le développement de certaines maladies. Bien qu'ils aient l'avantage de permettre d'établir plus facilement des relations de causalité, ces modèles n'abordent pas la question des interactions entre multiples variants génétiques, un phénomène qui est à la base de traits complexes chez les mammifères. Par contre les populations de référence de souris permettent une analyse des interactions moléculaires dans un contexte qui est le plus pertinent pour le trait clinique, à savoir les perturbations génétiques multiples (comme dans une population naturelle) plutôt qu'une perturbation génétique individuelle (comme dans une souris transgénique) ⁷³.

Les souches parentales

Le croisement de souches consanguines phénotypiquement différentes permet de cartographier les loci de traits quantitatifs ou qualitatifs ⁷⁴. L'analyse d'un trait complexe implique généralement le choix de souches parentales phénotypiquement distinctes, mais ces souches doivent également être génotypiquement distinctes car la cartographie génétique dépend des différences polymorphes entre les parents. Dans le cas le plus simple, on peut choisir deux souches progénitrices consanguines (appelés ici les souches parentales A et B, qui diffèrent par le caractère d'intérêt), où tous les animaux de chaque souche sont homozygotes.

La génération F1

Les souches parentales sont croisées pour obtenir la première génération filiale (F1). Les individus F1 reçoivent une copie de chaque chromosome de chacune des deux souches parentales. Dans la génération F1, tous les gènes autosomaux (et ceux du chr X chez les femelles) sont hétérozygotes. Les souches parentales A et B ont été choisies pour avoir des

distributions phénotypiques nettement différentes. Bien que les individus au sein de chaque souche soient génétiquement identiques, il existe une certaine variation dans les phénotypes en raison des différences environnementales (non héréditaires) et de l'erreur de mesure. Lorsque les individus F1 peuvent avoir une valeur phénotypique identique à celle de A (ou B), on parle de dominance du caractère apporté par A (ou B) et de récessivité du caractère apporté par B (ou A). Le plus souvent, la distribution phénotypique pour la génération F1 est intermédiaire entre les deux souches parentales A et B : on parle alors de codominance du caractère. Dans des cas rares, on observe l'hétérosis (également connue sous le nom de vigueur hybride), où l'hybride F1 présente une aptitude supérieure à celle des souches parentales.

Les croisements expérimentaux

La génération F1 sert de base pour les croisements expérimentaux. En effet, la nature hétérozygote des individus F1 permet de créer un brassage allélique aléatoire grâce à la recombinaison méiotique dans les croisements subséquents. Ces individus F2 peuvent être utilisés dans deux types de croisement : 1) Le croisement en retour (ou rétrocroisement) entre la génération F1 et la souche parentale A (ou B), qui permet de générer une population N2, ou bien 2) L'inter-croisement des individus F1, qui permet de générer une population F2 (les individus de cette génération ne sont plus identiques en raison de la ségrégation des allèles des parents F1 hétérozygotes)⁷⁵. Pour créer « des souches recombinantes consanguines » (RIS), les frères et sœurs de la génération F2 sont croisés pour plus de 20 générations (Figure 1.4). Chacune des nouvelles RIS générées contient des proportions à peu près égales de variations génétiques des deux souches consanguines parentales initiales, mais possède une combinaison unique de variants alléliques (la moitié provenant de parent A et l'autre du parent B)^{76,77}. Les individus de chaque souche sont génétiquement presque identiques (isogéniques) et presque homozygotes à tous les loci du génome. L'ensemble des diverses souches recombinantes consanguines provenant d'un même croisement initial représente un « panel ».

Pour détecter des QTLs à forte pénétrance, l'utilisation de près d'une vingtaine de souches recombinantes consanguines RIS serait nécessaire. Plusieurs panels de RIS sont actuellement disponibles auprès du laboratoire The Jackson, et comprennent 1) l'ensemble des AXB (n= 16 souches), des BXA (n= 14 souches), 2) des BXD (n= 89 souches), 3) des BXH (n=13 souches), et 4) des CXB (n=13 souches)⁷⁵. Ces souches ont été utilisées pendant des

décennies pour diverses études génétiques ⁷⁸. Pour la réalisation du projet de cette thèse, nous avons utilisé le panel de souris RIS AXB/BXA dérivés de croisement réciproque des souches parentales consanguines C57BL/6J et A/J (les mâles étant B6 dans les croisements AXB, et A/J dans les croisements BXA). Ces dernières diffèrent dans leur susceptibilité à 27 maladies infectieuses, de même que pour des traits incluant l'athérosclérose, l'obésité, le diabète, l'hydrocéphalie, et le cancer. Ainsi, les souches AXB et BXA sont utiles dans l'analyse génétique de plusieurs maladies complexes ⁷⁹. Au total, le panel de souris RIS AXB et BXA comprend 30 souches de souris. Dans notre travail, le nombre de lignées utilisables a été réduit à 24 pour plusieurs raisons : 1) certaines souches sont très semblables génétiquement et sont donc «redondantes» d'un point de vue génétique ; 2) certaines souches ne se reproduisent que très difficilement, ce qui complique l'obtention d'animaux expérimentaux.

D'un point de vue expérimental, le panel de souris AXB/BXA offre les avantages suivants :

1) Le génotype de chaque lignée a été bien analysé, et est disponible dans des bases de données publiques ⁸⁰.

2) La mesure du phénotype peut être répétée sur plusieurs individus génétiquement identiques au sein d'une même lignée (cela réduit la variabilité due à l'environnement et/ou les méthodes utilisées pour mesurer le phénotype) ⁸¹.

3) Un phénotype peut être mesuré dans différentes conditions sur différents individus d'une même lignée (par exemple à différents âges, ou après exposition à des traitements précis) ⁸¹.

4) Avec les RIS, il est possible de comparer des groupes contrôles et expérimentaux dans une même lignée ⁸¹.

5) Ces souches peuvent être maintenues pendant de nombreuses générations, fournissant un stock théoriquement infini d'individus génétiquement identiques. Cela permet d'établir des croisements spécifiques qui peuvent être utilisés pour étudier des gènes individuels de manière ciblée.

6) Ces souches peuvent être interrogées de façon répétée. Ceci a permis de développer de grandes bases de données qui contiennent des données phénotypiques moléculaires, physiologiques et pathologiques détaillées dans une gamme de conditions environnementales ⁶⁶

Malgré tous ces avantages, les souches issues de deux souches parentales consanguines particulières ont une diversité génétique qui se limite à celle présente dans les deux souches

parentales. Par ailleurs, les chromosomes recombinants sont fixés dans l'état homozygote par inter-croisement en série, ce qui ne permet pas d'illustrer certains aspects des maladies complexes humaines ⁸². De plus, le nombre de générations utilisé limite la possibilité d'augmenter le nombre d'évènements de recombinaison allélique. Ceci mène à l'obtention de carte génétiques de densité moyenne, et donc à une résolution cartographique limitée ⁸³

Pour pallier à certains des inconvénients énumérés ci-dessus, des nouveaux panels de souris tels que le «*collaborative cross*» ou «CC», et le «*outbred stocks*» ou «OS», ont été développés au cours des dernières années. Le «CC» a été conçu pour améliorer l'efficacité de la cartographie des QTL grâce aux caractéristiques suivantes: 1) il se distingue des «panels» de RIS par le fait que huit souches, au lieu de deux, ont été utilisées comme souches progénitrices, ce qui augmente la diversité génétique ⁸⁴; 2) à travers les centaines de souches de «CC» produites, l'accumulation d'un grand nombre d'évènements de recombinaison diminue les blocs de déséquilibre de liaison, ce qui améliore par conséquent la résolution de la cartographie des QTL ⁷³

Les «OS» se caractérisent par un croisement aléatoire continu pour éviter la consanguinité et pour maintenir une diversité génétique et phénotypique maximale⁸⁵. Par opposition à des souris dérivées d'une souche consanguine, les souris au sein d'un «OS» ne sont pas génétiquement identiques. Parmi les «OS» les plus couramment utilisés sont le «*heterogeneous stock*» et le «*diversity outbred stock*» ^{84,86}. Les «*heterogeneous stocks*» ont été dérivées à partir de croisement de huit souches. Le stock a été maintenu pour plus de 60 générations, avec une distance moyenne de évènements recombinaison de <1 Mb ⁸⁶. Les «*diversity outbred stocks*» sont générés à partir d'un croisement aléatoire de 160 CC souches (dérivés d'un croisement de huit souches).

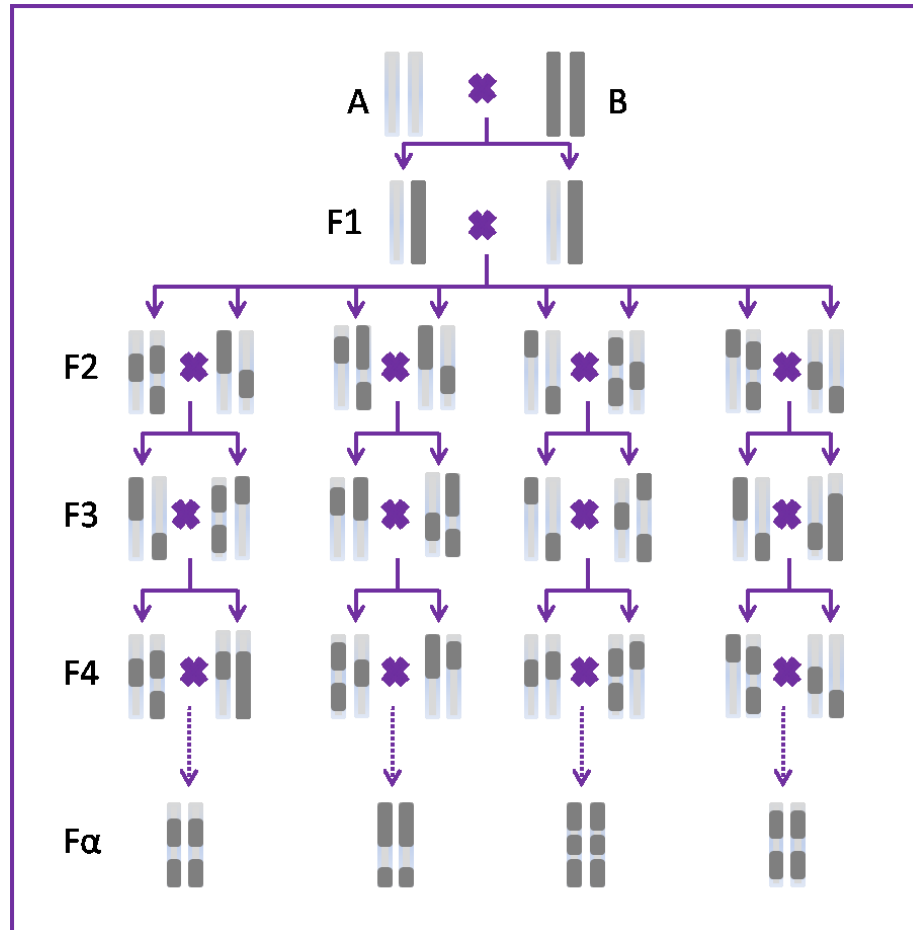


Figure 1.4: Représentation schématique de la génération de souches recombinantes consanguines. L'élevage des deux premières générations est identique à celle d'un croisement F2. Dans les générations suivantes, les frères et sœurs sont accouplés pour produire une progéniture qui est de moins en moins hétérozygote. Si l'on continue indéfiniment ces croisements, ce processus produira des individus qui sont complètement homozygotes à chaque locus, mais avec des chromosomes qui sont une mosaïque des chromosomes parentaux. Dans la pratique, entre 10 et 20 générations de croisements sont réalisés.

1.4 Limitation des études des QTLs

La cartographie de locus associés aux traits complexes peut être très compliquée, vu que le caractère considéré est sous le contrôle de plusieurs loci, ayant chacun un effet individuel trop modeste, ce qui rend l'identification de chaque locus difficile. De plus, les effets de tous les loci

associés à un trait complexe ne représentent qu'une faible proportion de la variabilité globale (ou de l'héritabilité estimée) du trait (due partiellement à des facteurs environnementaux et leur interaction avec divers gènes influençant les traits)²¹, ce qui rend impossible la caractérisation précise de l'architecture génétique des traits complexes.

Les études de cartographie des QTLs ont permis d'identifier des milliers de variants génétiques associés à des phénotypes particuliers^{70,87}. Toutefois, l'identification du SNP ou du gène causal était toujours un défi⁸⁸. Un facteur de complication est que les variants associés à la maladie pourraient ne pas être les véritables variants causaux, mais sont en LD avec le véritable variant pathogène, ce qui rend difficile d'impliquer le gène correct d'un locus dans la pathogenèse de la maladie.

Un autre problème plus important des études de QTL est la compréhension des effets biologiques des variants détectés. En fait, la fonctionnalité suggérée de ces variants génétiques est typiquement spéculative, basée sur l'annotation disponible des gènes à proximité des variants. Puisque, la plupart des variants alléliques associés à des maladies complexes se situent en dehors des régions codantes et sont des régulateurs putatifs de l'expression génique^{54,89}, il a été suggéré qu'en étudiant la régulation de l'expression génique, on pourrait mieux comprendre le mécanisme moléculaire de la maladie.

1.5 Les QTL d'expression (eQTLs)

Des études de l'expression génique dans différents organismes vivants ont montré que la variation naturelle des niveaux d'expression génique est héréditaire et répandue dans toutes les populations. L'héritabilité élevée de la variation de l'expression des gènes a suggéré que l'identification des déterminants génétiques de l'expression des gènes peut donner un aperçu sur la base moléculaire des traits quantitatifs complexes. «*Genetical genomics*»⁹⁰, ou «*expression genetics*»⁹¹ offre la promesse d'accélérer l'identification des gènes contribuant à la variation des phénotypes complexes, et permet la dissection de la régulation de l'expression génique⁹². Ces études, appelées aussi études des eQTLs (pour QTLs d'expression), ont été effectuées dans la levure^{93,94}, la souris⁹⁵, et plus récemment chez l'homme^{97,99}, et ont révélé certaines des caractéristiques fondamentales de l'architecture génétique de la variation d'expression génique (Figure 1.5).

Un avantage important de l'étude de la génétique de l'expression des gènes est qu'il s'est avéré que, dans de nombreux cas, l'abondance des transcrits peut agir comme un phénotype intermédiaire entre la variation génétique et les phénotypes complexes. En fait, l'importance de ces études a été illustrée par plusieurs études GWAS montrant que la plupart des variants alléliques associés aux maladies complexes correspondent à des séquences d'ADN non codantes^{89,89,99,100}. Ceci indique que dans le cas des traits complexes, les effets de variants génétiques impliqueraient plus souvent des phénomènes de régulation de gènes plutôt que des changements de séquences altérant les acides aminés au sein d'une protéine. Il a été démontré que les mutations dans les régions régulatrices affectent une variété de phénotypes à travers la régulation de l'expression génique. Par exemple, des variants régulateurs affectant l'initiation de la transcription, l'épissage, la stabilité de l'ARN et l'efficacité de traduction sont connus pour jouer un rôle dans des conditions incluant les maladies auto-immunes (CTLA4¹⁰¹), la malaria (DARC¹⁰²), et divers cancers (SMYD3¹⁰³), suggérant que les variants génétiques altérant l'expression des gènes (plutôt que la séquence protéique codée par ces gènes) peuvent constituer une des raisons premières de la variation naturelle des caractères complexes.

En plus de l'identification des régions du génome ayant des effets fonctionnels sur l'expression des gènes, la cartographie des eQTLs peut aider aussi à trouver les voies impliquées dans une variété de traits complexes¹⁰⁴ (Figure 1.5). Par exemple, en combinant la cartographie des eQTLs avec les résultats d'études traditionnelles de liaison ou d'association de maladie humaine, on peut prioriser et même prédire le rôle régulateur spécifique pour des sites polymorphes précédemment associés à des maladies^{36,105,106}.

Dans le projet de cette thèse, les niveaux d'expression des gènes furent traités comme des traits quantitatifs, et les données d'expression provenant des ventricules gauches ont été cartographiés en combinant des études de variation d'expression génique avec les résultats du génotypage des 24 souches de souris AXB/BXA.

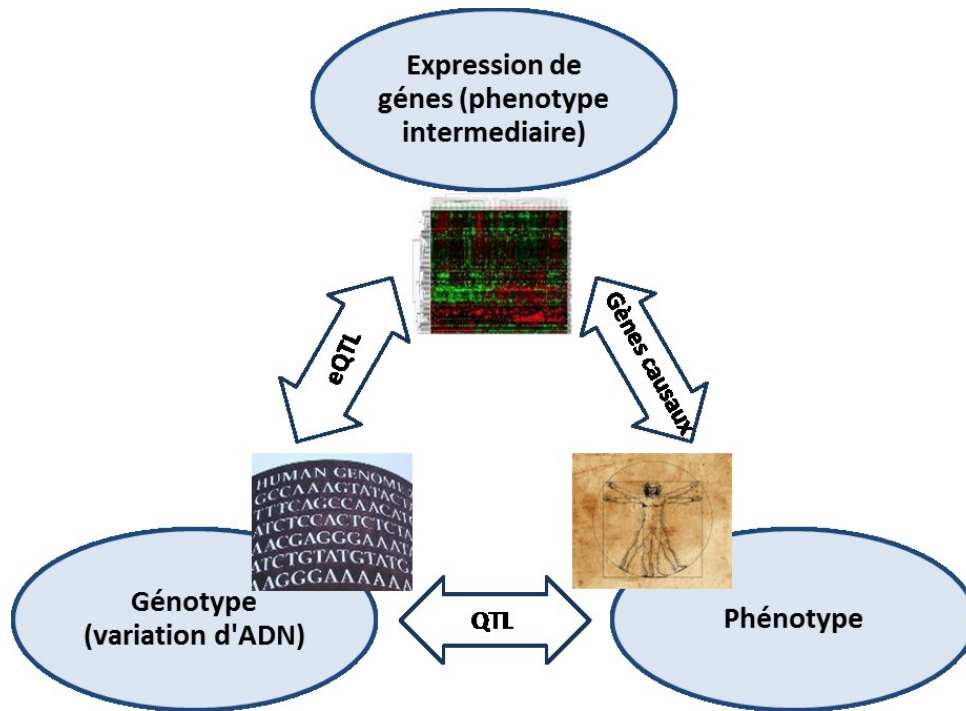


Figure 1.5: Résultats potentiels de la combinaison de plusieurs types de données moléculaires, y compris les données génotypiques, l'expression génique, et les données phénotypiques. L'analyse combinée des effets de la variation naturelle de l'ADN avec le profil d'expression des gènes (dans les populations de ségrégation) permet la cartographie des eQTL. La cartographie des eQTLs peut aider à identifier des gènes causaux des traits complexes.

1.5.1 Génotypage et estimation des niveaux d'expression des gènes

Les niveaux d'expression des transcrits ont été mesurés à l'aide de micropuces à ADN Illumina. Ces techniques sont très efficaces pour l'analyse à grande échelle. Cependant, étant donné qu'elles ne permettent que la mesure de transcrits d'ARN dont la séquence est connue, elles ne conviennent pas pour la détection de nouveaux transcrits ou de nouvelles isoformes d'épissage. De plus, elles sont souvent incapables de distinguer l'expression des transcrits ayant une faible abondance par rapport au bruit de fond ¹⁰⁷. La détection artéfactuelle d'un eQTL peut également se produire lorsque des polymorphismes surviennent entre les sondes utilisées dans les micropuces et les brins d'ARN eux-mêmes, affectant ainsi l'efficacité du processus d'hybridation ¹⁰⁸.

Les améliorations récentes du séquençage à haut débit permettent maintenant le séquençage et la quantification directs des bibliothèques d'ADNc ¹⁰⁹. Cette méthode, appelée ARN-Seq, permettra de détecter tous les variants génétiques (même ceux qui sont rares) ainsi que de découvrir de nouveaux gènes préalablement non identifiés ¹¹⁰.

1.5.2 La régulation de l'expression génique

La régulation de l'expression d'un gène qui code pour une protéine est un processus très complexe, qui implique l'interaction de plusieurs facteurs génétiques (la transcription, les modifications post-transcriptionnelle, et la traduction) ^{110,112,114}, épigénétiques ^{114,115}, et environnementaux ^{22,116,117}.

La composante génétique de la régulation génique est constituée de gènes, d'éléments *cis*- et de régulateurs agissant en *trans*-. Les régulateurs sont le plus souvent des protéines, appelées facteurs de transcription (TF), mais des petites molécules, comme par exemple les ARN et les métabolites, participent parfois aussi à la régulation globale. Les interactions et la liaison des régulateurs aux éléments dans la région *cis*- des gènes contrôlent le niveau d'expression de ces derniers pendant la transcription. Les régions *cis*- servent à agréger les signaux d'entrée, médiés par les régulateurs, et donc à effectuer un signal d'expression génique très spécifique.

L'état de la chromatine est aussi un des facteurs épigénétiques qui influe l'expression génique ¹¹⁸. En fait, la structure de la chromatine est dynamiquement altérée par la présence de variants d'histones (y compris l'acétylation, la méthylation et la phosphorylation de résidus spécifiques) ¹¹⁹, permettant ainsi soit la condensation de la chromatine dans un état transcriptionnellement inactif, soit l'ouverture des régions de la chromatine ¹²⁰. Ces zones de chromatine accessibles sont fonctionnellement liées à l'activité transcriptionnelle, étant donné que cet état remodelé est nécessaire pour la liaison de protéines régulatrices ou de complexes avec des éléments dans la séquence d'ADN.

1.5.3 Les QTLs d'expression locale, ou 'Cis-eQTLs'

Les différences dans l'expression de l'ARNm sont également générées par des interactions complexes et dynamiques des facteurs environnementaux et génétiques. Dans la

littérature de la cartographie des eQTLs, les polymorphismes régulateurs (eQTLs) sont souvent classées sur la base de 1) leur distance physique par rapport au gène régulé (dans ce cas-là, on parle de eQTL local et distal), et 2) leur mode d'action à travers lequel ils affectent l'expression de ces gènes (*cis*- et *trans*-) (Figure 1.6).

Lorsque l'expression d'un gène donné s'associe à un polymorphisme génétique qui se trouve proche du locus (ou dans la région régulatrice) de ce gène, le QTL d'expression correspondant (eQTL) est identifié comme eQTL proximal ou local (Figure 1.6A). Une telle proximité indique que la variation est susceptible d'agir sur l'expression du gène en *cis*- (c'est-à-dire uniquement dans la région chromosomique dans laquelle il réside), d'où la plupart des eQTLs locaux sont appelées *cis*-eQTLs. Un '*cis*-eQTL' peut être facilement détecté en mesurant seulement le niveau d'expression de chaque allèle^{122,124}. Bien sûr, il est possible aussi que certains eQTL locaux agissent en *trans*- par exemple, un gène pourrait montrer une rétroaction et pourrait réguler le gène sur le chromosome homologue. Il est donc plus précis d'utiliser le terme " local" ou " proximal" ¹²³.

L'enrichissement des eQTLs qui sont proches des gènes qu'ils régulent est compatible avec l'opinion selon laquelle les changements dans les éléments de régulation *cis*- sont moins susceptibles d'avoir des effets délétères, parce que les mutations dans ces éléments sont plus susceptibles de n'affecter la régulation que d'un seul gène¹⁰⁰. Un exemple d'un '*cis*-eQTL' correspond à la situation où une variation génétique dans la région régulatrice (un éléments de régulation *cis*- comme le promoteur, l'amplificateur, ou bien le site de liaison des siRNA) est responsable de la variation du niveau de transcription¹²⁴.

Dans des études à l'échelle du génome, des eQTLs ont été détectés en utilisant un large éventail de sources d'ARN, y compris des lignées de lymphoblastoïdes^{114,126,128}, le sang entier¹²⁸, les monocytes^{129,130}, les cellules B¹²⁹, les cellules hépatiques^{131,132} et des cellules tumorales de cancer du sein¹³³. Toutes ces études ont montré qu'une grande fraction de gènes humains (ou peut-être même tous les gènes) est régulée par des *cis*-eQTLs.

Dans le cas d'un «eQTL local», la distance entre le eQTL et le gène en question varie selon la population étudiée (entre 250 kilobases (kb) à 1 megabases (Mb) dans les populations naturelles et de 1 à 5 Mb dans des populations ségréguées)^{36,134}. Dans notre étude, nous avons défini les '*cis*-eQTLs' comme étant les eQTLs dont le pic était situé à moins de 1 Mb du site de début de transcription du gène correspondant. Cette région relativement étroite pourrait ne pas

inclure des régions régulatrices situées à de plus longues distances (comme les amplificateurs...etc) ^{135,136}. Fait à noter, pour différencier de façon plus rigoureuse les ‘*cis*-eQTLs’ des ‘*trans*-eQTLs’, on peut étudier l’effet de l’expression d’allèles hétérozygotes en utilisant RNA-Seq. Cependant, avec les souris homozygotes de notre population, cela n’est pas possible.

1.5.4 Les QTLs d'expression distale ou ‘*Trans*-eQTLs’

Dans d'autres cas, lorsque le niveau d'expression d'un gène s'associe à un locus (une variation génétique) clairement distinct de celui du gène lui-même, il est défini comme un eQTL distal, également appelé ‘*trans*-eQTL’ (Figure 1.6B). Contrairement aux loci agissant en *cis*-, on s'attendrait à ce que les loci agissant en *trans*- affectent de façon égale les deux allèles du gène cible. La distance précise requise pour qu'un eQTL soit considéré comme ‘distal’ est arbitraire et diffère d’une étude à l’autre. Par exemple, une telle distance peut aller de 10 kb dans la levure ⁹³ à 2 Mb chez l'homme ¹¹³. Certaines études exigent même que le eQTL soit situé sur un chromosome différent de celui du gène dont l’expression est affectée ¹³⁷. Un exemple possible de ‘*trans*-eQTL’ serait une situation où le *trans*-eQTL affecte le niveau d’expression d’un TF qui lui-même affecte l’expression d’autres gènes dont le locus est différent de celui de *trans*-eQTL.

Les études d'héritabilité suggèrent que plus de la moitié de la variance génétiquement liée à l'expression des gènes est due à des variants agissant en *trans*- ¹³⁸. Cependant, les ‘*trans*-eQTLs’ sont toujours plus difficiles à détecter que les *cis*-eQTL, parce que : 1) les effets *trans*- (les effets du second ordre agissant sur la transcription d'autres gènes), sont généralement indirects et plus faibles que les effets *cis*- (les effets du premier ordre qui affectent la transcription du gène lui-même) ⁹⁵; 2) un grand nombre de tests doit être effectué pour faire une recherche globale dans le génome pour les ‘*trans*-eQTLs’, ce qui implique l'utilisation de seuils de signification stricts ^{36,139,140}. Les ‘*trans*-eQTLs’ semblent aussi être plus spécifiques d’un tissu que les ‘*cis*-eQTLs’, ce qui complique davantage leur détection ^{141,143}. généralement, la proportion de ‘*trans*-eQTLs’ reproduite à travers plusieurs études s’est avérée plus faible que les ‘*cis*-eQTLs’ ^{105,128,130,131}.

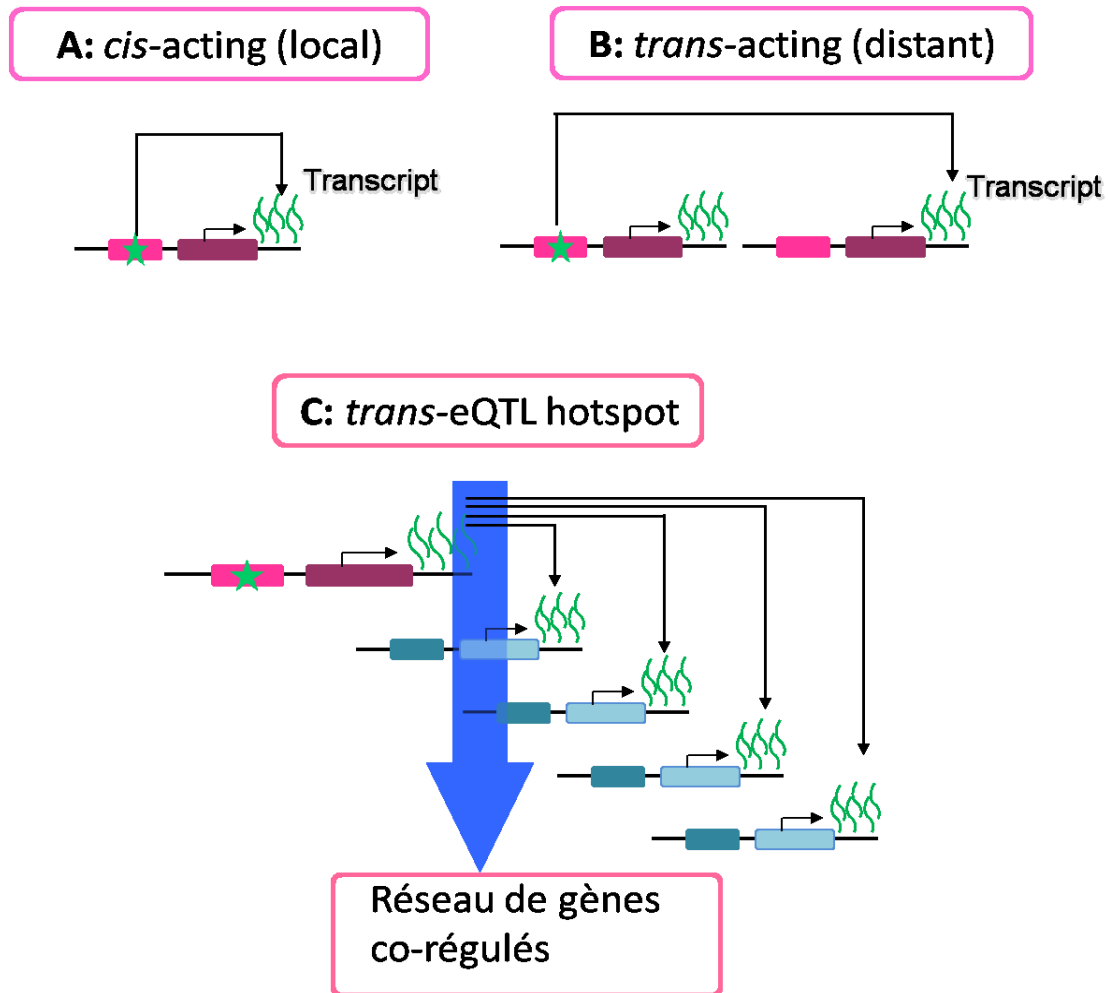


Figure 1.6 : Représentation schématique de ‘cis-eQTL’, ‘trans-eQTL’, et ‘trans-eQTL hotspot’. A) Un ‘cis-eQTL’ indique le cas où le locus de eQTL est le même que celui du gène. B) Dans le cas du ‘trans-eQTL’, le locus de eQTL est différent de celui du gène. C) un ‘trans-eQTL hotspot’ représente le cas où un seul locus est associé à l’expression de plusieurs gènes situés dans différents chromosomes.

1.5.5 Les limitations des études des eQTLs

Les études de la cartographie des eQTLs sont très prometteuses en tant que méthodes pour analyser la distribution spatiale des variants régulateurs dans le génome ¹²⁷, la spécificité temporelle de l’effet des séquences régulatrices sur l’expression génique (à savoir que certains éléments régulateurs n’influencent que l’expression des gènes dans certaines conditions), et les

changements d'expression à l'état d'équilibre associés à la variation des éléments régulateurs *cis-* ou *trans-*^{113,137}.

Malgré les nombreuses propriétés attrayantes des approches de cartographie des eQTL, la simple cartographie d'un locus associé à la variation d'expression génique ne fournit pas d'informations directes sur le mécanisme perturbé par la variation génétique associée, même si l'on suppose que le variant causal a effectivement été identifié. Pour comprendre quels mécanismes de régulation pourraient être affectés par les eQTLs, le processus de cartographie des eQTLs doit prendre en considération une grande variété de tests génomiques qui se rapportent à des aspects de la régulation des gènes, y compris l'état de la chromatine¹⁴³, les empreintes de transcription des TFs^{145,148}, les profils de différents marqueurs épigénétiques¹⁴⁸, et les modifications posttranscriptionnelles (y compris l'épissage alternatif et la dégradation de l'ARNm)^{150,154}. Il est maintenant évident que tous ces processus peuvent être affectés par des variants régulateurs. La logique est que si un eQTL agit en perturbant un mécanisme de régulation particulier, par exemple une modification d'histone, le eQTL devrait également être associé à des mesures du (ou des) mécanisme(s) de régulation pertinent.

1.5.6 La nature moléculaire des eQTLs

Les nouvelles technologies alimentées par les progrès dans le séquençage permettent un examen détaillé de la façon dont la variation de séquence influe les étapes individuelles de l'expression des gènes. Les eQTLs peuvent influencer la variation de niveaux d'ARNm mature par le biais de la régulation d'une variété de mécanismes transcriptionnels. Tout d'abord, les eQTLs peuvent affecter les modifications épigénétiques et l'initiation de la transcription. Ceux-ci incluent des processus régulateurs tels que la liaison des TFs, les modifications des histones, l'activité des amplificateurs (peut-être médiée par l'architecture et la conformation de la chromatine) et la méthylation de l'ADN. Les mécanismes de transcription, et en particulier la liaison des TFs, sont probablement les contributeurs les plus forts à la variation des niveaux d'ARNm à l'état d'équilibre. À l'appui de cela, des études ont montré que la majorité des eQTLs se produisent dans la chromatine ouverte (après détection de sites hyper-sensibles à la DNaseI)¹⁴⁴, et qu'ils sont particulièrement enrichis dans les sites de liaison des TFs¹⁵⁴, ce qui suggère que beaucoup ont un impact direct sur les interactions protéine-ADN (en changeant l'affinité de liaison de l'ADN pour les TFs actifs)^{143,145,154,155}. Une analyse conjointe de l'expression des

gènes, de l'affinité des TFs, de l'accessibilité à la chromatine, et de la méthylation, révèle que les variations associées à la modification de la liaison des TFs, peuvent également perturber la régulation des gènes de manière plus subtile - par exemple, en modifiant le modèle de méthylation de l'ADN ¹⁴⁸, et ou en perturbant le processus de remodelage de la chromatine ^{144,145,149,157,159}.

D'autres variations génétiques peuvent influencer le niveau d'expression des gènes en affectant le fonctionnement ou l'expression des facteurs régulant le processus transcriptionnel et cotranscriptionnel. Ceux-ci comprennent des mécanismes tels que l'épissage, et la dégradation de l'ARNm. Par exemple, plusieurs études ont cartographié la base génétique de la variation de l'épissage de l'ARNm. Elles ont constaté qu'une partie non négligeable de loci responsables de la variation de l'épissage, est enrichie près des sites d'initiation de la transcription, dans la région 5' UTR, et à l'intérieur des sites de liaison des TFs ¹⁵⁹. Cela suggère que les mécanismes d'épissage pourraient agir en même temps sur la régulation de l'expression des gènes grâce à des interactions avec des composants du mécanisme de la régulation transcriptionnelle. Un nombre restreint d'études ont également étudié le rôle des variations génétiques associées à la modulation de la liaison des microARN (miARNs) ^{161,163}. Ces variations présentent un enrichissement dans les motifs 3' UTR ou dans des éléments régulateurs impliqués dans la stabilité des transcrits, tels que des éléments riches en adénylate-uridylylate (AU) et des sites de fixation des miARN ^{153,161}. Fait intéressant, les variations perturbant la liaison des miARNs sont également souvent associées à la variation des niveaux d'expression de l'ARNm ^{161,162}. De plus, des études ont montré qu'à travers des mécanismes agissant en *trans*-, la variation génétique affectant l'expression des ARN non codants (comme lincRNA) pourrait à son tour influencer les niveaux d'ARNm de gènes codant les protéines ^{163,164}.

Les analyses combinées de données sur de multiples types de mécanismes de régulation nous permettent souvent d'obtenir une image plus complète de l'ordre des événements causaux qui mènent à des programmes de régulation génique précis.

1.6 Les '*trans*-eQTL hotspots'

Jusqu'ici, nous avons mis l'accent sur l'identification de variations génétiques qui influencent l'expression de gènes individuels. Cependant, certaines fonctions biologiques de l'organisme sont complexes et impliquent l'interaction de plusieurs gènes. Dans certains cas un

‘*trans*-eQTL’ est associé avec l'expression de dizaines ou centaines de gènes, formant ce qu'on appelle un ‘*trans*-eQTLs hotspot’^{165,166} (Figure 1.6C) et (Figure 1.7). Ceux-ci peuvent affecter la régulation de plusieurs traits (tel que observés dans plusieurs ensembles de données)^{168,171}

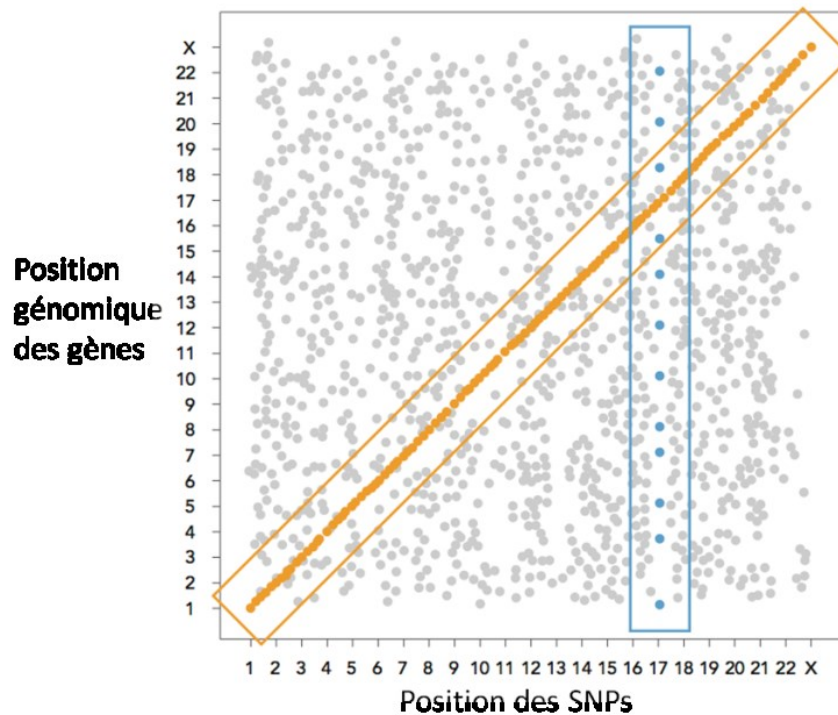


Figure 1.7 : Représentation graphique de l'analyse des eQTLs. L'axe des x représente la position génomique des SNPs, et l'axe des y montre la position génomique des gènes dont l'expression a été mesurée. Les points représentent les eQTLs. Chaque point dans le graphique représente une association entre le génotype à un locus génomique spécifique et l'expression d'un gène. La bande diagonale est appelée ‘bande *cis*-eQTL’ et représente une association entre le niveau d'expression d'un gène et le génotype au niveau du locus du gène. La bande verticale appelée ‘hotspot de *trans*-eQTL’ illustre les associations entre l'expression de nombreux gènes (situés dans différents chromosomes) et le génotype à un locus génomique unique.

1.6.1 L'importance des ‘*trans*-eQTL Hotspots’

Tel que discuté jusqu'ici, une étude de cartographie des eQTLs peut conduire à des aperçus plus convaincants sur les relations entre les variations de séquence d'ADN et les niveaux

d'expression génique. Pourtant, il est très peu probable que le niveau d'expression d'un seul gène (perturbé par une seule variation génétique) induira un phénotype cliniquement identifiable ou physiologiquement significatif, étant donné que la plupart des gènes opèrent dans des réseaux. Il est donc important de considérer ce fait lors de l'évaluation de l'impact de la variation de séquence d'ADN sur les niveaux d'expression de gènes multiples. Des études antérieures ont montré qu'un grand nombre de SNPs fortement associés aux niveaux d'expression de gènes particuliers, sont également associés aux niveaux d'expression de nombreux autres gènes ¹⁷¹. La raison pour laquelle le SNP pourrait être associé aux niveaux d'expression d'un ensemble de gènes est le fait que ces derniers appartiennent souvent aux mêmes réseaux et participent à des processus ou des voies biologiques communes.

Donc, les études des '*trans*-eQTL hotspots' peuvent avoir d'importants intérêts fonctionnels. Généralement, un hotspot représente le cas d'une ou plusieurs variations qui sont associées avec l'expression d'un ensemble de gènes fonctionnellement apparentés ^{173,175}. Donc, en faisant ce genre d'étude, on pourrait caractériser une ou plusieurs variations génétiques ^{94,175,176} qui influencent une fonction cellulaire spécifique ou des processus biologiques entiers. De plus, cette stratégie pourrait nous permettre d'identifier de nouveaux régulateurs, et même faire des prédictions fonctionnelles pour des gènes dont la fonction n'a pas encore été caractérisée mais qui peuvent être impliqués dans des voies de signalisations déjà connues ^{94,96,142}. Les '*trans*-eQTL hotspots' peuvent aussi chevaucher de manière significative avec des loci contenant des signaux d'association pour des maladies complexes ^{105,128,177}.

1.6.2 La régulation des '*trans*-eQTL hotspots'

Tel que déjà mentionné, un '*trans*-eQTL hotspot' représente un cas où l'expression de gènes multiples est associée à un locus génétique unique et commun. Cela soulève la possibilité qu'un ou plusieurs gènes influencés par la variation de séquence en question soit plus en amont dans le réseau, et induit une cascade d'événements physiologiques qui affectent plusieurs gènes du même réseau ^{94,96,142}. Donc, idéalement, une région '*trans*-eQTL' doit également contenir un ou plusieurs gènes '*cis*-eQTL', qui pourraient être considérés comme des régulateurs potentiels des gènes du hotspot ^{105,166,178}.

De nombreux mécanismes peuvent être cohérents avec ces résultats. Dans le scénario le plus simple, le gène ‘*cis*-eQTL’ pourrait être un régulateur transcriptionnel direct (un TF), qui cible la plupart des gènes du ‘*trans*-eQTL hotspot’. Des études testant l’enrichissement des sites de liaison des TFs au sein des promoteurs de gènes dans les ‘*trans*-eQTL hotspots’ ont conduit à l’identification de plusieurs TFs [y compris *KLF4*¹⁷⁹, *IRF2*¹⁷⁷, *IKZF1*¹⁰⁵ et *IRF7*¹⁸⁰] en tant que «hub genes» dans le réseau de gènes *trans*-eQTL. Par ailleurs, d’autres études ont également montré qu’un «*cis*-eQTL» lié à l’expression d’*IRF5* est également lié en *trans*- à l’expression de plusieurs autres gènes. Ainsi, un SNP associé au niveau de l’ARNm d’*IRF5* a également été associé aux niveaux d’ARNm de plusieurs gènes impliqués dans la voie IFN type I, ainsi qu’au lupus systémique chez les populations asiatiques¹⁸¹. Dans ce cas cependant, le nombre de gènes affectés est relativement plus faible que dans les études précédentes dont le but premier était la cartographie des ‘*trans*-eQTL hotspots’.

Un gène ‘*cis*-eQTL’ pourrait également être celui d’un régulateur transcriptionnel intermédiaire, qui à son tour module d’autres gènes du ‘*trans*-eQTL hotspot’. Donc, le gène ‘*cis*-eQTL’ pourrait affecter l’expression de gènes par des mécanismes différents de ceux utilisés par les TFs. Parmi les cas décrits dans la littérature, on peut relever les exemples suivants : 1) une enzyme de type « cytochrome P450 » (qui peut affecter les concentrations de métabolites influençant eux-mêmes l’expression d’autres gènes¹⁷⁷), 2) le gène *LYZ* (qui code pour l’enzyme lysozyme, dont l’activité bactériolytique peut affecter plusieurs fonctions des monocytes¹⁸²), 3) le gène *MKT1* (qui code pour une protéine qui se lie à l’ARN¹⁷⁶, et peut ainsi affecter la stabilité des ARNm de plusieurs gènes), ou 4) le gène *GPA1* (qui code pour une protéine de type G qui se lie au récepteurs de phéromones et influence l’expression de gènes régulés par ces derniers chez les levures⁹⁴). Quel que soit le mécanisme en cause, l’identification d’un *trans*-eQTL fournit un point de départ permettant d’identifier en son sein un gène affectant de façon plus large une voie biologique spécifique.

1.6.3 Les 'trans-eQTL hotspots' détectés jusqu'à présent

Références	organisme	taille de population	nombre des eQTLs locaux	nombre des eQTLs distants	Seuils de detection des eQTLs	Nombre des hotspots	Le nombre de hotspots validés
Brem et al., Science, 2002 ⁹³	Levure	40	185	385	$p < 5 \times 10^{-5}$	8	0
Schadt et al., Nature, 2003 ⁹⁵	souris	111	1,022	1,985	LOD>4.3	7	0
Yvert et al., Nat Genet, 2003 ⁹⁴	levure	86	578	1,716	$p < 3.4 \times 10^{-5}$	13	0
Morley et al., Nature, 2004 ⁹⁷	humain	14 CEPH families	29	118	$p < 4.3 \times 10^{-7}$	2	0
Mehrabian et al., Nat Genet, 2005 ¹⁶⁷	souris	111	20,107	20,107	LOD>2	1	0
Chesler et al., Nat Genet, 2005 ¹⁴¹	souris	35	83	5	FDR= 0.05	7	0
Hubner et al., Nat Genet, 2005 ⁵⁷	rat	259	622	1,211	$p < 0.05$	2	0
Bystrykh et al., Nat Genet, 2005 ¹⁸³	souris	30	478	136	$p < 0.005$	plusieurs	0
Li et al., PLoS Genet, 2006 ¹⁸⁴	C. elegans	80	414	308	FDR= 0.04	1	0
Lan et al., PLoS Genet, 2006 ¹⁸⁵	souris	60	723	5,293	LOD>3.4	15	0
DeCook et al., Genetics, 2006 ¹⁸⁶	Arabidopsis	30	3,525	3,525	FDR= 2.3%	5	0
Wang et al., PLoS Genet, 2006 ⁶¹	souris	312	2,118	4,556	$p < 5 \times 10^{-5}$	7	0
McClurg et al., Genetics, 2007 ¹⁸⁷	souris	32	N.A.	N.A.	N.A.	25	0
Keurentjes et al., PNAS, 2007 ¹³⁹	Arabidopsis	160	1,875	1,958	FDR= 0.05	29	0
West et al., Genetics, 2007 ¹⁷⁰	Arabidopsis thaliana	211	15,771	N.A	$P < 0.05$	17	0
Ghazalpour et al., PLoS Genet, 2008 ¹⁶⁶	Souris	110	471	701	FDR = 0.1	4	0
Grieve et al., Plos one 2008 ¹⁸⁸	rat	29	3,240	5,029	$p < 0.05$	81	0
Wu et al., PLoS Genet, 2008 ¹⁸⁹	souris	28	600	885,840	$p < 0.003$	1,659	1
Schadt et al., PLoS Biol, 2008 ¹³²	humain	427	3,210	242	$p < 1.6 \times 10^{-12}$	23	0
Heinig et al., Nature 2010 ¹⁹⁰	rat humain	30 1490	N.A	305	FDR< 0.1%	2	1Local (ce n'est pas un TF)

Rockman et al., science 2010 ¹⁶⁹	C. elegans	208	1,500	N.A	FDR=0.05	3	0
Fehrmann et al., Plos genetics 2011 ¹²⁸	humain	1,469	679	130	FDR<0.05	1	0
Small et al., nature genetics 2011 ¹⁷⁹	humain	776	N.A	N.A	p-value <0.05	1	1 Local (c'est un TF)
Orozco et al., Cell 2012 ¹⁶⁸	souris	92	4,747	9,344	FDR<0.05	147	1 local (ce n'est pas un TF)
Rotival et al., Plos genetics 2011 ¹⁸²	humain	1,490	N.A	44	FDR<10 ⁻³	3	0
Westra et al., 2013 Nat Genetic ¹⁰⁵	humain	5,311	6418	103	FDR< 0.05	3	1 local (c'est un TF)
Pierce et al., Plos Genetics 2014 ¹⁷⁸	humain	1,799	39	434	FDR <0.05	26	0
Fairfax et al., Science 2014 ¹⁷⁷	humain	432	multiple	1838	FDR <0.05	10	1 Local (c'est un TF)
Kang et al., Cell reports 2014 ¹⁷⁴	rat	200	67%	190	p =4 × 10 ⁻⁷	1	1 local et 1 trans (ce n'est pas des TF)
Wang et al., J Exp Bot 2014 ¹⁹¹	riz	210	5079	8568	P=0.05	138	0
Lee et al., Science 2014; ¹⁸⁰	humain	534	264	N.A	FDR< 0.05	1 (10 gènes)	1 Local (c'est un TF)
Tian et al., Genetics 2015 ¹⁹²	souris	500	2400	~2085	LOD ≥5	Plusieurs	0
Quach et al., Cell 2016 ¹⁹³	humain	200	N.A	42-945	FDR< 0.05	2	0
Hasin-Brumshtein et al eLife 2016 ¹⁹⁴	souris	282	2600	>800	FDR< 0.05	2	0

Tableau 1.I : Tableau récapitulatif des ‘*trans*-eQTL hotspots’ détectés jusqu’à présent. Ce tableau fournit une compilation des études de ‘*trans*-eQTLs hotspots’ effectuées à ce jour; il montre aussi que le nombre de hotspots validés biologiquement ne constitue qu’une minorité des hotspots détectés.

Jusqu’à présent, un nombre non négligeable de ‘*trans*-eQTLs hotspots’ a été détecté. Les approches utilisées pour la détection des hotspots, et le choix du gène(s) candidat(s) suggéré(s) d’être un régulateur commun des gènes du ‘*trans*-eQTLs hotspots’ sont différents d’une étude à l’autre. La plupart des études utilisent les variants génétiques qui sont connues pour être associés à des maladies et des phénotypes complexes comme point de départ, et évaluent si ces

variants sont également associés au niveau d'expression génique. Les gènes dont l'expression est modulée par ces variants génétiques sont alors considérés comme des gènes causaux qui lient la variabilité du génome à la maladie complexe. Ces études sont généralement basées sur l'hypothèse que les loci de risque identifiés par GWAS sont aussi des eQTLs qui affectent en *cis*- un ou plusieurs TFs qui, à leurs tours, influencent plusieurs cibles en aval (des '*trans*-eQTLs'). Par exemple, une étude a montré qu'un SNP lié au lupus érythémateux disséminé (SLE) est également associé en *cis*- à un TF (*IKZF1*), et en *trans*- à l'expression d'un groupe de gènes connus pour être impliqués dans le SLE ¹⁰⁵. Une autre étude a montré que le même ensemble de SNPs influence 1) en *cis*- l'expression d'un TF (*KLF14*), 2) en *trans*- l'expression d'un réseau de gènes, et 3) une variété de traits métaboliques incluant le diabète de type 2 (T2D) et le cholestérol HDL ¹⁷⁹.

Par ailleurs, plusieurs d'autres se sont concentrées sur l'identification de variants qui régulent un ensemble de gènes dans des contextes environnementaux bien spécifiques, en considérant que cela peut constituer une ressource utile pour évaluer le rôle de ces variants dans les mécanismes moléculaires et cellulaires qui sous-tendent des traits connus (comme la susceptibilité aux infections virales et aux maladies auto-immunes). Ces études ont pu détecter des '*trans*-eQTLs hotspots' dans des conditions de stimulations variées. Par exemple, un traitement avec les lipopolysaccharides (LPS) a permis la détection d'un eQTL avec un LOD score hautement significatif, qui est associé en *cis*- à l'expression d'un TF (*IRF2*) et en *trans*- à un réseau de gènes impliqués dans la réponse immunitaire ¹⁷⁷. D'autres études ont pu détecter un locus associé en *cis*- à un TF (*IRF7*), et en *trans*- à l'expression d'un groupe de gènes, après stimulation avec le virus de la grippe ¹⁸⁰. Un grand avantage de ce type d'études est que le traitement induit des effets régulateurs, et augmente l'expression des gènes à des niveaux quantifiables, facilitant ainsi la détection des '*trans*-eQTLs hotspots'. Cependant, la relation entre le '*cis*-eQTL' et le réseau de gènes régulés en *trans*- sont temporelles, et dépendent du type et de la durée de la stimulation.

De plus, conformément à l'hypothèse selon laquelle la régulation *trans*- est médiée par des effets locaux, les SNP *trans*- agissant dans les hotspots sont également associés à l'expression de plusieurs gènes locaux ('*cis*-eQTLs') qui ne sont pas eux-mêmes des TFs, (mais qui peuvent affecter l'activité de certains TFs. Parmi les exemples publiés, on relève : 1) le gène *Ebi2* (qui régule *Irf7*) ¹⁹⁰, 2) *LYZ* (qui interagit avec MYB, un activateur transcriptionnel), 3)

YEATS4 (un gène montrant un degré d'homologie élevée avec MLLT1 et MLLT3, des sous-unités de complexe d'élongation) ¹⁸², et 4) *TLRI* (un des récepteurs transmembranaires qui affecte l'expression d'un groupe de gènes pro-inflammatoires) ¹⁹³.

1.6.4 La validation biologique des ‘*trans*-eQTLs hotspots’

Bien que l'identification des ‘hotspots de *trans*-eQTL’ se soit améliorée, grâce au développement de techniques bioinformatiques, la validation biologique reste toujours un défi. Malgré les nombreuses difficultés qui existent, quelques études de validation biologique ont été effectuées. Dans le cas où le gène régulateur est un ‘*cis*-eQTL’ qui correspond à un TF, il est plus probable que le niveau d'activité de ce TF peut être reflété dans les niveaux d'expression des gènes cibles (régulés en *trans*-). Il est alors recommandé d'évaluer si un motif de liaison du TF est enrichi dans les gènes cibles. Cela a été effectué en utilisant des bases de données bioinformatiques ^{105,179}, ou en faisant des tests expérimentaux comme la ChIP-seq (l'immunoprécipitation de la chromatine suivie d'un séquençage à haut débit) ¹⁷⁷. Il existe des exemples de validation biologique pour certains cas de « *cis*-eQTL » correspondant à des gènes ayant des actions régulatrices autres que celles utilisées par les TFs. Dans les situations où il existe plusieurs régulateurs candidats, des expériences de knockdown ont été réalisées afin de prioriser davantage le (ou les) gène (s) candidat (s). Donc, l'effet du knockdown de tous les ‘*cis*-eQTLs’ est évalué en mesurant le niveau d'expression des gènes ‘*trans*-eQTL’ par une qPCR ^{168,174}. Dans des cas particuliers, un ‘*cis*-eQTLs’ peut réguler un ‘*trans*-eQTL hotspot’ à travers la régulation d'un ‘*trans*-eQTL’. Par exemple, une étude a montré qu'un locus régule en *cis*- *Ebi2*, et en *trans*- un groupe de gènes régulés par *IRF7* ‘*IRF7-driven inflammatory network*’ ou (IDIN)¹⁹⁰. *Ebi2* régule IDIN à travers la régulation d'expression d'*Irf7* (qui est lui-même un ‘*trans*-eQTL’). Le rôle d'*Irf7* dans la régulation des IDIN a été confirmé par des analyses d'enrichissement du site de liaison de TF. Une autre étude a montré qu'un ‘*cis*-eQTL’ appelé *Trem2* régule l'expression d'un ‘*trans*-eQTL’ appelé *kcnn4*, qui à son tour régule un réseau de gènes ‘*trans*-eQTL’, impliqués dans la multinucléation des macrophages. *Kcnn4* a été considéré comme gène régulateur car il est parmi les gènes les plus régulés dans le hotspot, et car son rôle dans la régulation du phénotype étudié a été déjà documenté ¹⁷⁴.

D'après le tableau récapitulatif, on constate que parmi le grand nombre de 'trans-eQTL hotspots' détecté (à peu près 34 hotspots), très peu (6 études seulement) ont été validés expérimentalement. Ce manque de validation peut être dû à une variété de raisons, telles que les interactions des gènes avec les facteurs environnementaux (GxE) ¹⁹¹. Les effets synergiques et compensatoires peuvent également expliquer le manque de validation, car il peut être nécessaire de cibler simultanément deux gènes ou plus pour observer un effet dans certains des gènes 'trans-eQTL'. Par exemple, une étude a montré que les effets combinés de deux gènes candidats dans un hotspot du chromosome 8 (2310061C15Rik et Gcsh) pourraient seulement expliquer la régulation d'environ 12% des gènes totaux qui ont été cartographiés à ce locus ¹⁶⁸. Il est important de noter que la variation génétique de l'expression génétique est importante et très complexe, car l'expression d'un gène peut être contrôlée par un ou plusieurs loci (avec des effets semblables ou des effets opposés), et que ces effets peuvent varier en fonction du fond génétique.

1.7 But du travail de la thèse

Dans cette étude, nous avons généré des données eQTL à partir d'échantillons de cœurs obtenus d'un panel de 24 souches de RIS (AxB / BxA) ⁷⁹. Nous avons précédemment utilisé ce panel de souris pour la détection des QTLs liés à la masse ventriculaire gauche cardiaque ¹⁹⁵. Ici, nous avons concentré nos efforts sur la caractérisation génomique et fonctionnelle d'un QTL lié à l'expression de nombreux gènes, c'est-à-dire d'un locus décrit dans la littérature comme un « hotspot de trans-eQTLs », ou région « *master eQTL* ».

1.7.1 L'utilisation du iBMQ '*R/Bioconductor package for integrated Bayesian modeling of eQTL data*' pour la cartographie des 'hotspots de trans-eQTLs'

La détection des 'trans-eQTL hotspot' a connu un succès limité à cause de l'absence de méthodes statistiques performantes : 1) les méthodes de calcul utilisées pour détecter les 'trans-eQTLs', y compris les approches bayésiennes, ont tendance à être exigeantes en termes de calcul et nécessitent des logiciels spécialisés ^{196,197} ; 2) la plupart des études des eQTLs utilisent des méthodes univariées traitant chaque gène indépendamment, ce qui conduit à des résultats sous-

optimaux ^{196,197}. Pour pallier au manque de puissance des techniques de calcul utilisé précédemment dans la détection de ‘*trans*-eQTL hotspots’, nous avons appliqué un nouvel outil d'analyse bayésienne des eQTLs, appelé iBMQ ¹⁹⁸, aux données générées précédemment ¹⁹⁹, pour détecter et cartographier les eQTLs. Les données d'expression comprennent 8725 gènes et 977 marqueurs (SNPs) dans le tissu cardiaque de 24 RIS (AxB / BxA), mesurées à l'aide du ‘*Illumina Mouse Ref-8 v2.0 BeadChips*’. Un seuil de détection correspondant à un "logarithm of odds" de 3,3, a été utilisé ²⁰⁰.

L'iBMQ est un logiciel qui met en œuvre un modèle bayésien hiérarchique commun, où tous les gènes et SNPs sont modélisés simultanément. Les méthodes bayésiennes fournissent des cadres de modélisation naturels pour analyser les eQTL, où l'information partagée entre les marqueurs et / ou les gènes peut augmenter le pouvoir de détecter les eQTL. Le logiciel iBMQ R / Bioconductor intègre des données génotypiques et des données d'expression génique dans un seul modèle, il fait face à la grande dimensionnalité des données eQTL (le grand nombre de gènes), et il contrôle le nombre de faux positifs à un niveau souhaitable.

Les principales fonctions implémentées par de IBMQ sont: 1) cartographier les eQTLs (ou les SNPs) significatifs, 2) classifier les eQTLs soit comme ‘*cis*-eQTLs’, soit comme ‘*trans*-eQTLs’, et 3) trouver les ‘hotspots de *trans*-eQTL’ (identifier des marqueurs uniques associés à plusieurs gènes). Parmi les ‘*trans*-eQTLs’, iBMQ a détecté trois groupes de ≥ 50 gènes formant des ‘hotspots de *trans*-eQTL’. IBMQ montre une plus grande sensibilité que les méthodes univariées traditionnelles pour détecter les ‘*trans*-eQTL hotspots’ contenant un grand nombre de gènes ¹⁹⁸.

1.7.2 Le ‘*trans*-eQTL hotspot’ dans le chromosome 17

En particulier, le plus grand ‘hotspot de *trans*-eQTLs’, détecté par le iBMQ, correspond à un locus sur le chromosome 17 (chr17), qui contient 190 transcrits régulés en *trans*-. Parmi ces transcrits, 177 correspondaient à des gènes identifiés par un nom de gènes officiel, et parmi ces derniers, 110 correspondaient à des ‘*Interferon stimulated genes*’ (ISGs), dont la plupart sont stimulés par l'IFN de type I.

1.7.2.1 L'IFN de type I et les ISGs

La régulation d'expression de l'IFN de type I

Les interférons de type I (IFN) sont des polypeptides sécrétés par les cellules infectées. Ils activent les programmes antimicrobiens intracellulaires et influencent le développement de réponses immunitaires innées et adaptées^{201,202}. Il existe deux types d'IFN de type I : IFN α et IFN β . La plupart des types de cellules produisent de l'IFN β , alors que les cellules hématopoïétiques, en particulier les cellules dendritiques plasmacytoïdes, sont les producteurs prédominants d'IFN α . L'IFN β est codé par un seul gène *IFNB*, alors qu'IFN α est codé par 14 gènes distincts. Les IFN de type I peuvent être produits par presque n'importe quel type cellulaire dans le corps en réponse à la stimulation d'un réseau de récepteurs transmembranaires et cytosoliques appelés PRRs (*pattern-recognition receptors*), ces derniers reconnaissent différentes composantes structurales du virus, y compris l'ADN viral, l'ARN bicaténaire (ARNdb), l'ARN monocaténaire (ARNsc) et les glycoprotéines de surface, qui sont reconnus comme étant des PAMPs (*pathogen-associated molecular patterns*)²⁰³.

Il existe un certain nombre de PRRs, chacun étant spécialisé dans la détection de PAMP distincts; ces PRRs recrutent également des adaptateurs différents pour activer TBK1 (*TRAF family member-associated NF- κ B activator* ou (*TANK*)-binding kinase 1)^{204,205}. La famille des hélicases de type RIG-I, connue sous l'acronyme RLR (RIG-I like receptors), reconnaissent l'ARN bicaténaire cytosolique et l'ARN 5' triphosphorylé^{206,207}. Cette famille est constituée de trois membres, RIG-I (*retinoic acid-inducible gene-1*) codée par le gène *DDX58*^{206,208}, MDA5 (*melanoma differentiation-associated gene 5*) codée par le gène *IFIH1*, et LGP2 (*laboratory of genetics and physiology 2*) codée par le gène *DHX58*, qui partagent une organisation modulaire commune, incluant deux domaines CARDs N- terminaux (absent chez LGP2), un domaine hélicase central (Hel), et un domaine C- terminal (CTD)²⁰⁹. La liaison de l'ARN, au niveau des domaines Hel et CTD, induit un changement de conformation et une exposition des domaines CARDs qui interagissent avec le domaine CARD de la protéine MAVS (*mitochondrial antiviral signaling*) Cette dernière (aussi appelée IPS-1, Cardif, ou VISA), interagit avec TRAF3 (*TNF receptor-associated factor*)^{210,211}, TRAF2/5/et 6, permettant ainsi l'activation des facteurs de transcription NF- κ B et IRFs, et l'initiation de la réponse antivirale²¹²⁻²¹⁴.

La '*GMP-AMP cyclique synthase*' (cGAS, cGAMP synthase) a récemment été identifiée comme récepteur de PAMP cytosolique^{215,216}. Par un mécanisme qui reste à élucider, la détection cytosolique d'un ADN PAMP active l'activité de nucléotidyl transférase de cGAS et conduit à la production de cGAMP. Cette molécule joue le rôle d'un second messager qui lie et

active STING (*stimulator of IFN genes*)²¹⁷. STING est une protéine transmembranaire exprimée dans le réticulum endoplasmique (ER), en particulier dans les régions de contact entre les mitochondries et le ER. Ce domaine particulier (connu sous le nom '*the mitochondrial-associated membrane*') joue un rôle clé dans l'activation immunitaire innée^{218,219}. Les radicaux libres (« Reactive Oxygen Species », ou ROS) peuvent agir en tant que second messagers pour induire l'oligomérisation de MAVS et la production d'IFN de type I indépendamment de l'infection virale²²⁰. De plus, l'augmentation des ROS cellulaires amplifie la signalisation RIG-I et la fonction MAVS^{221,222}.

Les récepteurs Toll-like (TLR) 1, 2 et 4 sont des protéines transmembranaires qui se localisent à la surface cellulaire, où elles détectent des glycoprotéines virales et une variété de PAMPs bactériennes (la flagelline, les LPS, les peptidoglycanes, les toxines bactériennes et les acides nucléiques bactériens). De même, TLR 3, 7 et 9 sont également des protéines transmembranaires, mais ce groupe de récepteurs se localise dans les endosomes, où ils se spécialisent dans la détection de virus et d'acides nucléiques bactériens, de façon semblable aux RLRs cytosoliques²²³. L'induction de la production d'IFN de type I par les TLR3 et TLR4, implique la molécule adaptatrice TRIF (*TLR/IL-1R domain-containing adaptor protein inducing IFN- β*), qui s'associe avec la molécule TRAF3^{224,225}.

En fait, le recrutement de différentes molécules adaptatrices (telles que STING, MAVS, et/ou TRIF) par divers PRRs conduit à l'activation et au recrutement de TBK1 et d'IKBKE (*inducible I κ B kinase*, codée par le gène *IKBKE*^{226,227}). La transmission de signaux d'activation aux IKBKE et TBK1 conduit à la phosphorylation des facteurs de transcription '*IFN-regulatory factor 3*' (IRF-3) et IRF-7 aux résidus C-terminaux^{210,228}. IRF-3 et IRF-7 sont des acteurs importants dans l'induction de l'IFN de type I suite à l'infection virale^{223,225}. Alors que l'IRF-7 peut être induit au niveau transcriptionnel, IRF-3 se trouve sous forme latente dans le cytoplasme. Lors de la phosphorylation de la région C-terminale par TBK1 ou IKBKE, IRF-3 et IRF-7 hétérodimérisent et translocalisent dans le noyau, où ils induisent plusieurs gènes pro-inflammatoires et antiviraux ainsi que l'IFN de type I^{226,227,232,233}. L'analyse des cellules et/ou de souris knock-outs simples pour TBK1 et IKBKE suggère un certain niveau de redondance dans l'activité de ces deux²³⁴; la perte de TBK1 seule semble avoir un impact plus profond sur la cascade de signalisation de l'interféron de type I que la perte de IKBKE seul²³²⁻²³⁶.

L'IFN de type I sécrété se lie au récepteur IFN α / β (IFNAR) et déclenche la production de nombreux gènes antiviraux via la voie JAK / STAT ²³⁷. La production des ISGs sensibilise les cellules à la détection des agents pathogènes. Cependant, la plupart des régulateurs positifs nécessitent, en plus de l'induction de leurs niveaux d'expression, une activation supplémentaire pour remplir leurs rôles de signalisation. Par exemple, les PRR sont inactifs jusqu'à ce que les PAMPs soient détectés; IRF3 et IRF7 nécessitent une phosphorylation pour la translocation nucléaire et l'activité de renforcement de la transcription. La signalisation canonique par les protéines STAT nécessite également une phosphorylation. Une exception notable dans cette classe d'ISGs est IRF1, qui est capable de *trans*-activer d'autres gènes sans aucune modification post-traductionnelle ou signaux supplémentaires ²³⁸.

La voie de signalisation de l'IFN de type I

IFN α et IFN β se lient à un récepteur transmembranaire hétérodimère appelé récepteur IFN α (IFNAR), qui est composé des sous-unités IFNAR1 et IFNAR2. Dans la voie canonique de signalisation induite par IFN de type I, il a été montré que l'engagement de l'IFNAR activait les protéines tyrosine kinases associées aux récepteurs JAK1 (*Janus kinase 1*) et TYK2 (*tyrosine kinase 2*), qui phosphorylent et activent les facteurs de transcriptions STAT1 et STAT2 (*the latent cytoplasmic transcription factors signal transducer and activator of transcription 1/2*) ^{239,240} (Figure 1.8). Les STAT1 et STAT2 phosphorylés par la tyrosine se dimérisent et se déplacent vers le noyau, où ils s'assemblent avec l'IRF9 pour former un complexe trimoléculaire appelé ISGF3 (*IFN-stimulated gene factor 3*). L'ISGF3 se lie à ses séquences d'ADN, appelé ISRE (*interferon-sensitive response elements*), activant ainsi la transcription des ISGs impliqués dans la réponse antivirale.

En plus de l'activation de l'ISGF3, l'IFNAR active aussi les homodimères STAT1 et STAT3, qui se lient à une séquence appelée GAS (*Gamma interferon activation site*) ^{238,240-242}. Contrairement à STAT1, STAT3 active des gènes qui suppriment les réponses inflammatoires ²⁴³, ce qui suggère que les IFNs de type I peuvent activer une voie de signalisation parallèle STAT3 qui équilibre et restreint les voies pro-inflammatoires induites par les IFN de type I. L'équilibre entre l'activation de différents STATs par IFNAR est partiellement déterminé par les niveaux relatifs d'expression des STATs.

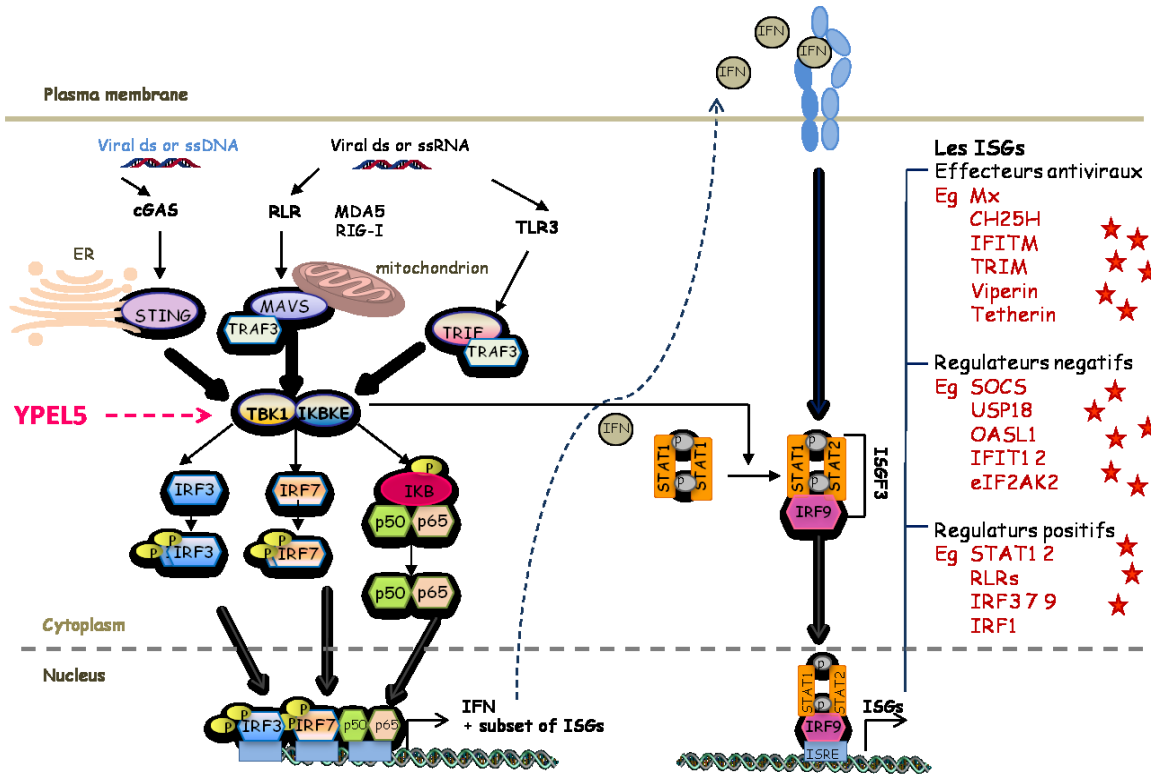


Figure 1.8 : La voie de signalisation d'IFN de type I. La reconnaissance de l'ARNdb ou de l'ADNdb par des récepteurs cytoplasmiques, tel que cGAS, RLR, et TLR3 déclenche la production d'IFN de type I via la signalisation par des molécules adaptatrices, y compris MAVS, STING et TRIF, qui s'associent à TBK1 et IKKε. Ces deux kinases forment un hétérodimère qui, lorsque activé : 1) stimule la phosphorylation de IRF3 et IRF7 qui induisent le gène IFNB en se liant à des éléments de réponse dans son promoteur ; et 2) phosphoryle IκBα, menant ainsi à la libération du complexe p50/p65 et à l'activation de la voie NFκB. Cette voie peut être également activée par la phosphorylation directe de la p65 initiée par IKKε. Par ailleurs, l'IFN de type I active la voie de signalisation de JAK-STAT. IKKε joue également un rôle à ce niveau en inhibant l'homodimérisation STAT1, ce qui favorise l'activation du complexe ISGF3. Le complexe ISGF3 induit de nombreux ISGs qui sont soit des effecteurs de la machine antivirale, soit des régulateurs positifs ou négatifs de cette dernière. L'expression d'IKKε est induite par diverses voies, y compris le facteur de nécrose tumorale (TNF) et les voies induites par les PRRs.

L'amplification de la signalisation basale de l'IFN de type I

Les composantes canoniques de la voie de signalisation de l'IFN de type I (i.e. IFNAR, JAK1, TYK2, STAT1, STAT2 et IRF9) sont exprimées de façon généralement ubiquitaire et, par conséquent, la plupart des types de cellules sont capables de manifester des réponses à l'IFN de type I. Cette capacité est maintenue dans des conditions homéostatiques par une boucle autocrine dans laquelle de faibles concentrations picomolaires d'IFN sont suffisantes pour augmenter l'expression de STAT1 et IRF9 (qui sont eux-mêmes des ISGs), conditionnant ainsi les cellules immunitaires pour une sensibilité accrue aux IFN de type I pendant les premières phases de l'infection, lorsque les quantités de cette cytokine sont limitées ²⁴⁴. L'induction préférentielle de l'expression de STAT1 est associée à une augmentation de la formation d'homodimères STAT1 et du complexes ISGF3, alors que la protéine IKBKE favorise la formation du complexes ISGF3 en inhibant l'homodimérisation de STAT1 ^{245,246}. La phosphorylation de STAT1 médiée par IKBKE supprime la formation d'homodimères STAT1, facilitant ainsi la formation d'ISGF3 ²⁴⁷.

En plus d'augmenter l'expression du gène STAT1, le conditionnement (« *priming* ») des cellules immunitaires induit aussi l'augmentation de la phosphorylation des résidus tyrosine et serine de la protéine STAT1 ²⁴⁰. La phosphorylation de la serine carboxy-terminale conservée dans STAT1 conduit à une activité transcriptionnelle accrue. La transcription génétique induite par les STATs et par le complexe ISGF3 implique aussi la coopération d'autres TFs, comme les membres de la famille IRF ²³¹. Les séquences d'ADN reconnues par la plupart des IRFs se chevauchent avec les ISRE, ainsi les IRFs et l'ISGF3 peuvent lier et activer les mêmes gènes cibles. De plus, l'expression de IRF1, IRF5, IRF7, IRF8 et IRF9 est induite par des IFNs à travers des mécanismes dépendants de STAT, et de nombreux ISGs contiennent à la fois des éléments GAS, ISRE et/ou IRF. Au total, il apparaît que IFN peut initier une cascade d'activation complexe où, dans un premier temps, STATs et l'ISGF3 activent directement certains ISGs, y compris IRF1, IRF5, IRF7, IRF8 et IRF9, et où ces IRFs activent dans un deuxième temps un second groupe d'ISGs ^{248,249} (Figure 1.8).

Dans l'ensemble, la signalisation IFNAR est amplifiée par divers mécanismes qui améliorent les réponses d'IFN de type I pendant les infections, en favorisant les fonctions antimicrobiennes et pro-inflammatoires des IFN de type I.

La régulation négative de la réponse IFN par les ISGs

Pour restreindre l'amplitude de la signalisation basale d'IFNAR, des voies de rétroaction négative (qui limitent la durée ou l'intensité de la réponse IFN de type I) sont induites par les ISGs, Il s'agit notamment de: 1) l'internalisation des IFNAR par des voies de signalisation activées par des cytokines pro-inflammatoires telles que l'IL-1, les TLRs, et le stress oxydatif ou métabolique ^{243,246}; 2) la désactivation des composants du signal IFNAR-JAK-STAT par l'induction de régulateurs négatifs, tels que SOCS (*suppressor of cytokine signaling*) (qui empêche la liaison des STAT au IFNAR, et qui supprime l'activité de JAK) ²⁵⁴, et USP18 (*ubiquitin specific peptidase 18*) (qui déplace JAK1 de IFNAR2) ^{254,255}; 3) la répression de la traduction des STATs et ou des IRFs; par exemple, la suppression de la traduction d'IRF7 par le répresseur de traduction général 4E-BP (qui est régulé par la signalisation IFN de type I) ²⁵⁶. La traduction d'IRF7 est supprimée aussi par la protéine OASL1 (*2'-5'-oligoadenylate synthase-like protein 1*) d'une manière inductible et très spécifique. Il existe encore d'autres ISGs qui sont aussi des régulateurs traductionnels, tels que IFIT1 et IFIT2 (*Interferon-induced protein with tetratricopeptide repeats 1/2*), qui se lient et inhibent la formation du complexe de préinitiation ribosomique médiée par eIF3 ^{257,258}; et la protéine kinase R (PKR, également connue sous le nom de eIF2AK2), qui phosphoryle et inactive le complexe eIF2, qui est important pour la fonction ribosomique ^{238,259}.

Le rôle de la réponse d'IFN de type I et des ISGs dans l'infection virale

Il existe plusieurs ISGs qui affectent l'infection virale par divers mécanismes antiviraux, et durant tous les stades d'infection. Parmi les effecteurs antiviraux, il y a : 1) les inhibiteurs de la sortie virale (tels que MX1, MX2, CH25H, les protéines IFITM, et les protéines TRIM; 2) les inhibiteurs de la traduction et de la réplication du virus, incluant la protéine antivirale zinc-finger (ZAP), les protéines IFIT1 et IFIT2, la voie OAS-RNaseL et la PKR ^{253,256}; et 3) les régulateurs

de la modification post-traductionnelle des protéines virales, incluant notamment la protéine de type ubiquitine, appelé ISG15, qui se lie de manière covalente à des protéines cibles dans un processus appelé ISGylation ²⁶⁴(parmi les protéines impliquées dans l'ISGylation et le deISGylation, il y a UBE2L6, HERC5, HERC6, UBE1LA, TRIM25 et USP18 qui sont également induits par l'IFN ²⁶⁵).

Les deux effecteurs antiviraux les plus étudiés et les plus fortement induits, sont :

1) la molécule appelée « tetherin » est codée par l'ISG *BST2*. En 2008, *BST2* a été découvert comme facteur responsable de la prévention de la libération du VIH-1 (avec le *Vpu* muté) des cellules hôtes ^{266,267}. Après ces découvertes, *BST2* a été renommée « tetherin » grâce à sa topologie unique qui permet de lier les virus enveloppés à la surface des cellules infectées ^{268,269}. Depuis, il a été montré que l'effet de liaison de *BST2* s'étendait à d'autres « virus enveloppés » ²⁶⁹⁻²⁷². En attachant les virus enveloppés à la surface des cellules infectées, la *BST2* limite non seulement la libération du virus, mais provoque également l'amplification de la réponse immunitaire innée ^{273,274}.

2) la viperine, codée par l'ISG *RSAD2* et connue pour son rôle antiviral puissant contre un certain nombre de virus, y compris le virus de l'immunodéficience humaine (HIV)²⁷⁵, le virus de l'hépatite C (VHC) ²⁷⁶, et plusieurs autres ²⁷⁷⁻²⁸⁰. Il a également été démontré que la viperine participe à la signalisation cellulaire du système immunitaire, en modulant la signalisation NFκB et AP-1 dans les cellules T ²⁸¹et les voies TLR-7 et TLR-9 dans les cellules dendritiques plasmacytoïdes ²⁸².

En plus de leurs effets antiviraux, les IFN de type I empêchent la réplication intracellulaire des bactéries ^{266,269}. D'autres effets antibactériens possibles de l'IFN de type I ont été postulés sur la base de son pouvoir à moduler la réponse innée et adaptative contre les bactéries, par exemple l'activation des cellules NK (*natural killer*), la production d'anticorps et l'activation des cellules dendritiques ²⁸⁷. Les IFN de type I peuvent moduler différents processus biologiques comme : l'autophagie, la différenciation, la migration cellulaire, la prolifération cellulaire, l'angiogenèse, et l'apoptose ^{288,289}.

1.7.2.2 Les ‘*trans*-eQTL hotspots’ enrichis en gènes immunitaires et leur association avec des maladies

De toute évidence, les IFN de type I induisent des réponses équilibrées dans lesquelles des signaux activateurs induisant des états antiviraux et favorisant les réponses immunitaires, sont contrebalancés par des signaux suppressifs limitant la toxicité pour l'hôte. Ces réponses équilibrées sont affinées par divers facteurs opérant à de multiples niveaux. Une signalisation aberrante, provoquée par la dérégulation des IFN de type I et des ISGs, peut conduire à plusieurs syndromes inflammatoires ou auto-immuns²⁹⁰. Les ISGs sont régulés de façon coordonnée, et sont souvent regroupés dans des réseaux de gènes.

Les mécanismes de régulation des réseaux de gènes immunitaires varient en fonction des individus et des types de cellules, reflétant la diversité génétique et épigénétique²⁹¹. Ils varient aussi selon les conditions environnementales (en présence ou en absence de stimulation). Un locus ou une variation génétique qui affecte l'expression d'un réseau de gènes immunitaires (ou un hotspot de gènes), peut aussi contribuer en partie à la susceptibilité à plusieurs maladies complexes associées à une activité immune inappropriée ou à une inflammation, y compris les maladies auto-immunes [comme les T1D, les auto-anticorps T1D,^{292,293} la maladie cœliaque²⁹⁴, l'hyperthyroïdie²⁹⁵, le vitiligo²⁹⁶, la polyarthrite rhumatoïde²⁹⁴], et d'autres traits complexes tels que la pression artérielle^{297,298}, la maladie rénale chronique²⁹⁹, la maladie d'Alzheimer³⁰⁰, l'athérosclérose³⁰¹, l'infarctus du myocarde³⁰², et le cancer.

Effectivement, des études antérieures ont montré plusieurs exemples de chevauchement entre les ‘hotspots de *trans*-eQTLs’ (enrichis en gènes immunitaires) et les loci contenant des signaux d'association à des maladies complexes. Par exemple, une étude de cartographie de ‘*trans*-eQTL hotspots’ a montré qu'un locus correspondant au gène 2310061C15Rik, qui régule l'expression de plusieurs gènes inflammatoires, est aussi susceptible de contribuer à l'athérosclérose¹⁶⁸. De plus, les études de cartographie de QTLs de type GWAS ont montré qu'une proportion significative des loci liés à la majorité des maladies complexes est susceptible d'affecter plusieurs *trans*-hotspots de gènes immunitaires ou antiviraux^{128,179,190,303}. Par exemple, une étude récente a montré que les variants génétiques associés au métabolisme du cholestérol et au diabète de type 1, régulent aussi des gènes impliqués dans la réponse immunitaire¹⁰⁵. Il a été montré que la diminution de la biosynthèse du cholestérol favorise

l'expression de gènes anti-viraux par un mécanisme qui implique une signalisation améliorée du complexe STING-TBK1-IRF3, induisant ainsi une production accrue d'interférons de type I ³⁰⁴⁻³⁰⁶.

En conclusion, on peut déduire que l'intégration de la cartographie des 'hotspots de *trans*-eQTL' avec les études de GWAS permet de mieux comprendre les effets en aval de nombreux variants génétiques associés aux traits complexes, et finalement d'élucider les causes de ces troubles. En outre, puisque les '*trans*-eQTL hotspots' sont souvent formés de gènes regroupés dans des voies immunologiques clés, il est probable qu'un '*trans*-eQTL hotspot' peut avoir un effet pléiotropique et affecter plusieurs traits qui ont une étiologie génétique commune (comme les maladies auto immunes) ³⁰⁷. Fait important, la cartographie des 'hotspots de *trans*-eQTL' peut être caractérisée dans des cellules exposées à différentes conditions de stimulation, permettant ainsi de détecter des effets très « contextuels ». Ceci met en évidence l'importance de déterminer la variation de la régulation génétique dans différentes conditions patho-physiologiques ^{105,168,177,180,193}.

1.7 Objectifs et hypothèses

Tel que mentionné précédemment, nous avons déjà détecté un 'hotspot de *trans*-eQTLs' sur le chr17. Comme le même locus est lié en *cis*- au gène '*Yippee like 5*' (*Ypel5*), nous avons voulu étudier si ce gène pouvait être un régulateur des gènes régulés en *trans*- et, si tel était le cas, de tester et valider expérimentalement le mécanisme de fonctionnement de *Ypel5*.

Le gène *YPEL5* fait partie d'une famille constituée d'un groupe de 5 gènes (*YPEL1-5*), présentant un haut degré d'homologie avec le gène *Yippee* de la Drosophile ³⁰⁸. *Yippee* a été découvert pour la première fois chez la drosophile en tant que protéine interagissant avec l'hémoline, une protéine de type immunoglobuline présente chez les lépidoptères et découverte initialement dans des larves de *Hyalophora cecropia* en condition de stress immunitaire ³⁰⁹. Des orthologues des gènes *YPEL* existent chez tous les eucaryotes (incluant les protozoaires et les plantes) et ces gènes montrent un haut degré de conservation dans tout l'arbre évolutif (de la moisissure aux humains). Les gènes *YPEL* se composent de 4 exons codant pour une protéine de 121 acides aminés (poids moléculaire prédit de 13,7 kDa). L'analyse de la séquence d'acides aminés a révélé la présence d'un domaine '*zinc binding Ring Finger*', un motif conservé

dans tous les homologues yippee. YPEL5 est le membre de la famille dont la séquence d'acide aminé est la plus homologue à celle du gène Yippee (71% d'identité). YPEL1-4 présentent un niveau d'homologie inférieur avec le gène Yippee (45% d'identité), mais un haut niveau d'identité les uns avec les autres.

La séquence de YPEL5 est identique (conservation à 100%) chez tous les vertébrés^{308,310}; cependant, sa fonction demeure inconnue. Dans mon projet de thèse, nous avons découvert un nouveau rôle pour *YPEL5* dans la régulation d'IFN β et des ISGs dans les macrophages de souris (RAW 264,7) et les cellules humaine (HEK 293T)³¹⁰. Par ailleurs, d'autres données expérimentales permettent de penser que le gène *YPEL5* est impliqué également dans le processus de sénescence cellulaire (voir chapitre III).

CHAPITRE 2: ASSOCIATION OF A NETWORK OF INTERFERON-STIMULATED GENES WITH A LOCUS ENCODING A NEGATIVE REGULATOR OF NON-CONVENTIONAL IKK KINASES AND IFNB1

Authors: Saloua Jeidane^{1,2}, Marie-Pier Scott-Boyer^{1,2,4}, Nicolas Tremblay^{2,3}, Sophie Cardin^{1,2}, Sylvie Picard¹, Martin Baril^{2,3}, Daniel Lamarre^{2,3}, Christian F. Deschepper^{1,2*}

Address:

¹ Cardiovascular Biology Research Unit, Institut de recherches cliniques de Montréal (IRCM), Montréal (QC), Canada H2W 1R7

² Département de Médecine, Faculté de Médecine, Université de Montréal, Montréal (QC), Canada

³ Centre de Recherche du Centre Hospitalier de l'Université de Montréal (CR-CHUM), Montréal (QC), Canada H2X 3J4

⁴current address: The Microsoft Research - University of Trento Centre for Computational and Systems Biology. Rovereto 38068. Italy

* Corresponding author: E-mail: deschec@ircm.qc.ca (CFD)

2.1 Summary

Functional genomic analysis of gene expression in mice allowed us to identify a quantitative trait locus (QTL) linked in *trans*- to the expression of 190 gene transcripts and in *cis*- to the expression of only two genes, one of which being *Ypel5*. Most of the *trans*-expression QTL genes were interferon-stimulated genes (ISGs), and their expression in mouse macrophage cell lines was stimulated in an *IFNB1*-dependent manner by *Ypel5* silencing. In human HEK293T, *YPEL5* silencing enhanced the induction of *IFNB1* by pattern recognition receptors and the phosphorylation of the TBK1/IKBKE kinases, while co-immunoprecipitation experiments revealed that *YPEL5* interacted physically with IKBKE. We thus found that the *Ypel5* gene (contained in a locus linked to a network of ISGs in mice) is a negative regulator of *IFNB1* production and innate immune responses that interacts functionally and physically with the TBK1/IKBKE kinases.

Highlights

- One genomic locus associates in *trans*- to expression of 190 genes in hearts from mouse strains
- Expression levels of the uncharacterized and conserved *Ypel5* gene associates in *cis*- to that locus
- *Ypel5* regulates interferon-responsive genes present in the hotspot of *trans*-regulated genes
- *YPEL5* interacts functionally and physically with the TBK1/IKBKE kinases

eTOC Blurb

After detecting a locus linked to the expression of many interferon-responsive genes in mice, Jeidane Saloua *et al* discovered that the locus contained the *Ypel5* gene, a negative regulator of *IFNB1* production and non-conventional IKK kinases.

2.2 Introduction

Functional genomics, which evaluate the functional consequences of genetic variations on intermediate molecular traits, have been proposed as a means to improve the power of detecting gene variants linked to complex traits³¹¹. This approach is supported by increasing evidence showing that gene regulatory changes often contribute to complex phenotypes [such as susceptibility to disease³¹²], and that a majority of genes have expression levels that associate with genetic variants¹⁷⁸. Quantitative trait loci (QTLs) linked to gene expression are called “expression-QTLs” (eQTLs). Genetic variants may affect the expression of genes that either map in their vicinity (*cis*-eQTLs) or reside further away (*trans*-eQTLs)¹⁷³. Heritability studies suggest that *trans*-acting variants are responsible for more than half of the genetically explained variance in gene expression¹³⁸.

In some particular cases, genetic variants may associate *in trans*- with expression levels of a great number of genes, with such clusters being commonly referred to as “*trans*-eQTL hotspots”¹⁶⁵. Their functional interest is highlighted by the facts that: 1) in some instances, they show enrichment for functionally related genes^{173,175}, which may give insights on how several genes in the network contribute to a biologic function; and 2) *trans*-eQTL hotspots overlap significantly with loci containing association signals for complex diseases^{105,128,177}.

The association of a single genetic variant to expression levels of a great number of genes *in trans*- suggests that genes in the cluster are downstream of a common regulator. Studies testing for enrichment of transcription factor (TF) binding sites within the promoters of genes in *trans*-eQTL hotspots have led to the identification of several TFs [including *KLF4*¹⁷⁹, *IRF2*¹⁷⁷, and *IRF7*¹⁹⁰] as “hub genes” within the *trans*-eQTL gene network. However, the loci of the latter did not always correspond to that of the *trans*-eQTL¹⁹⁰, indicating that *trans*-eQTL regions may still contain *cis*-eQTL genes that may not be TFs but are ultimately responsible for linkage of all *trans*-eQTL genes to their locus. Alternatively, given the recent report that up to 43% of the SNPs associating in *trans*- to gene expression levels also showed association with at least one local *cis*-eQTL¹⁷⁸, all *cis*-eQTL genes whose locus match that of the *trans*-eQTL could be considered as potential regulators of *trans*-eQTL genes, independently of their ability to function as TFs.

By analyzing gene expression profiles in mouse hearts, we detected previously a *trans*-eQTL connected to 192 gene transcripts¹⁹⁸. Within the eQTL hotspot, two transcripts (*Ypel5* and *Clip4*) corresponded to genes conforming to the definition of *cis*-eQTL genes (with transcription start sites at < 1Mbp from the SNP). On the base of preliminary data mining analyses, we prioritized *Ypel5* for further investigations. We found that *Ypel5* negatively regulates interferon- β (IFNB1) production via interactions with the TBK1/IKBKE kinases. In addition of their specific interest for the biology of the innate immune response, these findings illustrate the ability of this experimental strategy to discover genes coding for regulators of larger gene networks, without any prior knowledge of the biologic function of these genes.

2.3 Results

2.3.1 Characteristics of *cis*-eQTL and *trans*-eQTL genes

We previously profiled gene expression in hearts from 24 AxB/BxA mouse recombinant inbred strains (RIS) that originated from reciprocal crosses between the parental C57BL/6J and A/J strains¹⁹⁹, and further used that dataset to test the performance a newly developed Bayesian eQTL analysis tool¹⁹⁸. These tests allowed us to find that the new method detected ‘*trans*-eQTL hotspots’ containing large number of genes with greater sensitivity than more traditional univariate methods. In particular, the largest ‘*trans*-eQTL hotspot’ that was detected contained 192 genes linked in *trans*- to a locus (identified as “17@72.4”) at position 72.449 Mb on chromosome 17 (chr17), which corresponded to the rs33200140 single nucleotide polymorphism (SNP). Thus, the abundance of the *trans*-eQTL transcripts in RIS mouse hearts varied (in mice that all have a mixed C57BL/6J x A/J genetic background) according to whether rs33200140 originated from the parental A/J or C57BL/6J strains. *Yippee like 5* (*Ypel5*) and *Clip4* (at positions 72,851 Mb and 71,782 Mb, respectively) were the only eQTL genes having their transcription start site located at < 1Mbp from the rs33200140 SNP, and thus were the only two conforming to the definition of ‘*cis*-eQTL’ genes. The 190 *trans*-regulated transcripts in the hotspot originated from loci distributed among all chromosomes (Figure 2.1), with 177 of them corresponding to genes with official gene symbols.

The most statistically enriched categories of Gene Ontology (GO) terms for the genes within the hotspot were those corresponding to “immune responses”, “purine nucleotide binding” and “antigen processing and presentation” (Table S1a). As several of the hotspot genes corresponded to known “interferon-stimulated genes” (ISGs), we queried the “Interferome” database³¹³, and found that 110 out of the 177 ‘*trans*-eQTL’ with official gene symbols corresponded to ISGs (Table S2). Among these genes, 22 were identified as responding exclusively to type I IFN (IFN α /IFN β), 8 were identified as responding exclusively to type II IFN (IFN γ), whereas 80 are responding to both type I and type II IFNs. Thus, 102 out of 177 genes within the hotspot were responsive to type I IFN.

The eQTL cluster contained genes whose differences in expression in hearts from RIS mice (as a function of the allelic origin of the chr17 locus) were greater than 1.4 fold. We profiled the

pattern of expression of these genes across cells and tissues, using data from the BioGPS gene annotation portal (Figure S1). Euclidian clustering of the expression level values of these genes revealed 4 groups of cells and tissues. The highest levels of expression were found in stimulated macrophages, followed by 3 groups with progressively declining levels: the first one comprised most immune cells and tissues, the second one comprised mostly central nervous tissue (CNS) cells and testes, and lowest levels of expression were found in remaining tissues and cells (Figure S1). In comparison, expression of *Ypel5* itself was fairly ubiquitous, but highest levels were found in stimulated macrophages and mast cells, in bone marrow, and in tissues from the central nervous system.

The Ypel5 protein showed a high level of conservation across vertebrates, as there was 100% identity at the amino acid level in primates, rodents, bovine and xenopus (Figure S2a). Nonetheless, the only homologies detected by BLAST searches (either at the nucleotide or amino acid level) were for genes/proteins containing a *Yippee*-like domain (originally identified in *Drosophila*). To go beyond linear sequence analogies, we tested possible protein tridimensional (3D) homology/analogy, using the Phyre² recognition engine³¹⁴. Only 13 templates had scores for 3D structural match with YPEL5 higher than the suggested 80% threshold. Several of the templates were the C-terminal portion (i.e. the “regulatory domain”) of the RIG-I/DDX58, MDA5/IFIH1 and LGP2/DHX58 proteins (from multiple species), all of which being DExD/H box RNA helicases belonging to the sub-family of RIG-I-like receptor family (RLRs) that play important roles in the control of interferon- β (*IFNB1*)³¹⁵. YPEL5 shared with these proteins: 1) a similar succession of β strands and α helices, and 2) the positions of four zinc-coordinating cysteine residues that form a motif that is considered to be essential for the functions of these proteins³¹⁶ (Figure S2b). Three other templates corresponded to MSS4, a GDP/GTP exchange factor for small Rab-like GTPases which shows 3D similarity in both fold and zinc cluster with RIG-I/DDX58³¹⁶.

Expression of *Clip4* was more tissue-specific than that of *Ypel5*, with highest levels being found in skeletal muscle, mammary gland and cornea, medium levels in CNS, and very low levels in most other tissues. Likewise, the Clip4 protein showed only partial homologies (both by BLAST and Phyre² searches) for restricted ankb24 or Cap-Gly domains within other proteins. Given the predominance of ISGs among *trans*-eQTL genes and the greater expression

of both *Ypel5* and *trans*-eQTL genes in immune-related cells (and in particular macrophages), we considered *Ypel5* to constitute a more likely candidate regulator of the *trans*-eQTL genes than *Clip4*, and thus further tested whether *Ypel5* silencing affected the expression of *trans*-eQTL genes and/or *Ifnb1*, using mouse macrophage-like cell lines for initial experiments.

2.3.2 Effects of *Ypel5* knockdown on *trans*-eQTL genes in mouse macrophage cell lines

Among the ISGs showing the largest differences in abundance in hearts from mouse RIS (Table S2) (Figure 2.1), we randomly selected five genes (*Bst2*, *Usp18*, *Irf7*, *Ifit3*, and *Gbp2*) to test by RT-qPCR the effects of *Ypel5* silencing on their expression levels. Both in RAW 264.7 and J774 cells, transfection with specific *Ypel5* siRNA decreased *Ypel5* expression to less than 30 % of the levels found in cells transfected with scrambled siRNA, and significantly increased the expression of all 5 tested ISGs, in addition to that of *Ifnb1* itself (Figure S3a-b). Additional experiments in RAW 264.7 cells showed that the effects of *Ypel5*siRNA on all 5 ISGs were blocked when it was co-transfected with *Ifnb1*siRNA (Figure 2.2), suggesting that the responses of ISGs were in fact secondary to increased expression of *Ifnb1*. Beyond baseline conditions, we tested whether *Ypel5* silencing affected the induction of *Ifnb1* by either transfection with 25 ng poly (I:C) (which induces *Ifnb1* via RLRs acting as intracellular RNA sensors) or exposure of cells to 2000 ng extracellular poly(I:C) [which induces *Ifnb1* via Toll-like receptors 3 (TLR3)]. We verified previously that: 1) 25 ng of extracellular poly (I:C) had no effect on *Ifnb1*; 2) the dose of 2000 ng was slightly above the threshold required for extracellular poly(I:C) to affect *Ifnb1* expression; 3) neither poly(I:C) treatment affected endogenous *Ypel5* expression; and 4) *Ypel5* siRNA downregulated *Ypel5* expression to the same extent in all groups (Figure S3c-d-e). *Ypel5* silencing amplified the effects of both poly (I:C) treatments on *Ifnb1* induction (Figure 2.3). Specificity of *Ypel5* siRNA was further assessed in two different ways. First, we verified that the *Ypel5* siRNA had no effect on the expression of 2 control genes, i.e the *Gapdh* housekeeping gene, and *Gpr183*, a gene previously found to regulate expression of many inflammatory genes¹⁹⁰(Figure S3f). Secondly, we found that two additional *Ypel5* siRNAs affected endogenous *Ypel5* expression to the same extent as the first siRNA, and that they both increased *Ifnb1* production, either in basal conditions or after induction with intracellular poly (I:C) (Figure S3g-h).

To further explore the general effects of *Ypel5* silencing in macrophages, we compared the transcriptomes of RAW 264.7 cells transfected with either scrambled or *Ypel5* siRNA. Si-*Ypel5* affected the expression of a total of 944 gene transcripts (P -value < 0.005 and FDR < 0.1), including that of 50 genes belonging to the cardiac eQTL hotspot. Enrichment analyses for GO terms within the 944 genes revealed statistical enrichment for the “immune and defense responses” and “nucleotide and ribonucleotide binding” categories (i.e. similar categories as those showing enrichment in genes showing differential expression in mouse RIS hearts), as well as for the categories corresponding to “mitochondria” and “regulation of programmed cell death” (Table S1b). Among all gene transcripts affected by *Ypel5* silencing, 248 and 65 of them were either up- or downregulated by more than 30%, respectively, altogether corresponding to a total of 281 genes with official gene symbols. Among the latter, 160/281 genes (57%) were reported in the Interferome database as ISGs. When examining the 50 *trans*-eQTL genes showing $> 40\%$ fold differences in mouse RIS hearts and testing whether *Ypel5* silencing affected their expression in RAW264.7 cells, we found that this was the case for 26/50 of them (i.e. 52%). Altogether, these results showed that mouse RIS hearts expressed a core of IFNB1-responsive genes: 1) whose expression levels associated with differences in *Ypel5* expression, and 2) that overlapped with those genes whose expression was affected by *Ypel5* silencing in RAW264.7 cells.

2.3.3 Effects of YPEL5 on IFNB1 induction in human cells

We further tested whether *YPEL5* silencing would affect IFNB1 induction in human cells similarly as in mouse macrophage cell lines. We chose the HEK293T cell line, as it has been the preferred model in most previous studies on IFNB1-regulating pathways, and was previously used by ourselves to identify many potential modulators of the tightly regulated production of IFNB1³¹⁷. Sendai virus (SeV) is an RNA virus that is recognized by RIG-I, leading to IFNB1 production and up-regulation of ISGs^{318,319}. The effects of SeV infection on induction of RIG-I and MDA5 (tested as representative ISGs) were amplified by *YPEL5* downregulation (Figure 2.4a-c) and dampened by *YPEL5* overexpression (Figure 2.4d-f). After SeV infection, there was a time-dependent increase in the mRNAs of *IFNB1* and that of SeV nucleocapsid (NP) protein. These responses were either amplified (Figure 2.4g) or dampened (Figure 2.4h), respectively, by *YPEL5* downregulation. Si-*YPEL5* also amplified the effects of SeV on IFNB1-dependent

luciferase activity in transfected cells (Figure S4a); at no time after infection did we notice any change in endogenous *YPEL5* expression levels in scrambled siRNA-transfected cells (Figure S4b).

After sensing and binding to either non-self or 5'-ppp viral RNA, RLRs undergo conformational changes that allow them to interact with MAVS at the mitochondrial outer membrane³¹⁹. Binding of RIG-I and/or MDA5 to 5'-ppp RNA depends on the presence of several critical lysine residues in their regulatory domains³²⁰, but these residues were not conserved into the *YPEL5* sequence (Figure S2). Moreover, we verified experimentally that *YPEL5* did not share with RIG-I the ability to bind to either HCV RNA or poly (I:C) (a synthetic mimetic of double-stranded RNA) (Figure S5). Despite these differences, we tested whether *YPEL5* downregulation could affect signaling events initiated by activated RLRs. In HEK293T cells transfected with pIFNB1-LUC, *YPEL5* siRNA significantly amplified the induction of IFNB1-dependent luciferase activity by overexpression of either activated RIG-I (Figure 2.5a) or of effectors acting downstream of viral RNA sensing (i.e. MAVS, TBK1 or IKBKE) (Figure 2.5b-d). In addition, *YPEL5* downregulation enhanced the luciferase-inducing effects by two other nucleic acid sensing systems that activate the TBK1/IKBKE kinases independently of MAVS, i.e. TRIF, which mediates the effects of stimulated TLR3³²¹ and the cGAS/STING cytosolic DNA sensor system²¹⁶ (Figure 2.5e-f). One of the other reported effects of activated TBK1/IKBKE kinases is their ability to trigger phosphorylation events that lead to NF- κ B activation³²². Accordingly, we found that *YPEL5* downregulation enhanced the effects of TBK1/IKBKE overexpression on induction of endogenous *TNF*, a well-known target of NF- κ B (Figure S6c-d). Moreover, *YPEL5* silencing increased the effects of overexpressed TBK1 on its auto-phosphorylation and the effects of overexpressed IKBKE on its auto-phosphorylation and the cross-phosphorylation of TBK1 (Figure 2.6). To test possible physical interactions of *YPEL5*, we co-transfected HEK293T cells with both 3X-FLAG-*YPEL5* and constructs expressing either MAVS, TBK1 or IKBKE. Western blots performed on material immunoprecipitated with anti-FLAG antibody revealed that the precipitates contained immunoreactivity for IKBKE, but not for either MAVS or TBK1 (Figure 2.7). These results were confirmed in reverse experiments, where HEK293T cells were co-transfected with *YPEL5*-Myc and constructs expressing either Flag-tagged YFP, TBK1 or IKBKE (Figure S7). Altogether, the results showed that *YPEL5* interacts physically with IKBKE and functionally

with the activity of the TBK1/IKBKE kinases. However, *YPEL5* downregulation also increases IFN β 1-mediated ISG56luciferase activity in cells that are respectively exposed to rcIFN β and overexpression of IRF3 (5D), a phosphomimetic mutant of IRF3 that partially mimics virus-induced IRF3 activation (Figure S8).

The effects of *YPEL5* downregulation were validated in two different ways. We first verified that the effects of two additional *YPEL5* siRNAs on expression of endogenous *YPEL5* were similar to that of the first siRNA tested, and that amplified the effects of TBK1 on IFN β 1-dependent luciferase activity to a similar extent (Figure S9). We also ablated *YPEL5* in HEK293T cells by an RNA-guided nuclease approach and selected two clonal cell lines, with absence of *YPEL5* being verified by qRT-PCR (Figure S10a,b). *YPEL5* ablation was accompanied by a significant increase in IFN β 1- and NF- κ B -dependent activity (Figure S10c, d), along with increases in the expression of the endogenous *RIG-I* (a target of IFN β 1) and that of endogenous *TNF* (a target of NF- κ B) (Figure S10e,f). Likewise, TBK1 overexpression increased IFN β 1-LUC activity to a greater extent in the clones with *YPEL5* ablation (Figure S10c).

2.4 Discussion

In light of recent studies showing overlap of many ‘*trans*-eQTL hotspots’ with loci containing association signals for complex diseases^{105,128,177} a better understanding of the mechanisms controlling the expression of ‘*trans*-eQTL’ genes would be helpful to elucidate the causes of complex disorders. Up to now, the number of ‘*trans*-eQTL’ hotspots reported in the literature is small, with tentative explanations being that effect sizes are typically modest, that the multiple testing procedures create a particular statistical burden, and/or that *trans*-eQTLs are in general more sensitive to the effects of confounding factors^{105,165,323}. Nonetheless, three recent developments promise to facilitate the detection of ‘*trans*-eQTL hotspots’: 1) the pooling of the results of several studies, which may improve statistical power by allowing eQTL meta-analyses¹⁰⁵; 2) the design of studies where gene expression is studied in cells across several experimental conditions, in order to detect context-dependent eQTLs¹⁷⁷; and 3) improvements in the computational methods used to detect ‘*trans*-eQTLs’, including the use of Bayesian approaches^{196,197}. The eQTL hotspot reported herein was detected using such a multivariate computational tool¹⁹⁸. It contained 190 *trans*-regulated and two *cis*-regulated transcripts, among which we prioritized *Ypel5* as a possible candidate regulator on the base of data mining queries. The family of YPEL proteins has been reported in essentially all eukaryotes (including fungi, plants and unicellular organisms) and all members show very high levels of evolutionary conservation, which altogether suggest important functions for these proteins^{324,325}. However, functions of YPEL5 have been limited to a possible participation in events related to cell cycle progression^{324,325}.

Outside of the family of YPEL proteins, YPEL5 showed no homology with the linear amino acid sequences of any other known mammalian protein, but at the structural 3D level, it showed some homology with the regulatory domain of three RLRs. YPEL5 shared with these proteins the positions of four zinc-coordinating cysteine residues that form a motif considered as a typical structural element of this protein family³¹⁶. Although we found no evidence that YPEL5 could (in contrast to RIG-I) bind to viral RNA, there are precedents for RLR-related proteins being able to play roles other than just RNA sensing and to interact with further downstream signaling components. For instance, LGP2 (an homolog of RIG-I and MDA5)

inhibits IFN by acting both upstream and downstream of MAVS³²⁶. Likewise, DDX3 and SNRNP200 (which both belong to the family of DExD/H box RNA helicases) enhances the phosphorylation of IRF3 and the production of *IFNB1* via specific interactions with IKBKE and TBK1, respectively^{327,328}. Since most genes in the *trans*-eQTL hotspot were also known ISGs, we tested the hypothesis that *YPEL5* could play a role in *IFNB1* regulation. Our data indicate that downregulation of *Ypel5* by three different siRNAs enhances *Ifnb1* production in mouse macrophage cell lines. In human HEK293T cells: 1) downregulation of *Ypel5* by three different siRNAs enhances *Ifnb1* production, and the latter effect is accompanied by a decrease in SeV replication in infected cells; 2) the effects of CRISPR/Cas9-mediated deletion of *YPEL5* on *IFNB1*-dependent responses are similar to that of *YPEL5* downregulation; and 3) the effects of overexpressed *YPEL5* in HEK293T on ISGs are the reverse image of those obtained with *YPEL5* downregulation. It remains to be seen to which extent *YPEL5* affects the expression of inflammatory genes in other cell types and/or genetic backgrounds. For instance, several of the ‘*trans*-eQTL’ genes whose expression associated with the allelic origin of the chr17 locus in RIS mice (i.e. in mixed C57BL/6J x A/J genetic backgrounds) did not show significant differences in hearts from the parental inbred strains themselves (data not shown). The full effects of *YPEL5* may therefore either be modulated by additional factors and/or be partly dependent on them.

The TBK1 and IKBKE kinases are both non-canonical members of the IKK family that play a critical role in type 1 (and III) IFN responses via their ability to phosphorylate and activate IRF3³²⁹. In HEK293T cells, *YPEL5* silencing enhances the effects of overexpression of either: 1) TBK1 and IKBKE; 2) MAVS or RIG-I (which activates MAVS upstream of the TBK1/IKBKE kinases); or 3) TRIF or cGAS-STING (two additional adaptors that lead to activation of the TBK1/IKBKE kinases). Results of co-immunoprecipitation (co-IP) experiments indicated that the *YPEL5* protein interacted physically with material containing IKBKE immunoreactivity. That physical interaction had likely effects of the TBK1/IKBKE kinases themselves, since *si-YPEL5* enhanced the effects of overexpressed TBK1 and IKBKE on: 1) their auto- and cross-phosphorylation, i.e. two covalent modifications that participate in their activation³³⁰; and 2) the induction of endogenous *TNF*, a well-known target of NF- κ B which itself is activated by phosphorylation TBK1/IKBKE kinases³²². Other effects of *YPEL5*

may need further investigation, but are not incompatible with possible involvement of the TBK1/IKBKE kinases. For instance, the fact that the assembly of the multimeric STAT1/STAT2/IRF9 complex is facilitated by the effect of IKBKE on STAT1³³¹ might explain why *YPEL5* downregulation amplifies the effect of rcIFNB1. Likewise, the amplification of the effect of IRF3(5D) by si*YPEL5* might be explained by the fact that TBK1 directly phosphorylates MAVS, STING and MAVS to allow the latter to recruit IRF3 for phosphorylation on residues that are different from those whose phosphorylation are mimicked in IRF3(5D)³³².

Despite the difficulties of detecting ‘*trans*-eQTL hotspots’, most of the ones detected to date concern inflammatory genes^{105,128,177,190}. Since inflammatory responses typically implicate coordinated responses of many genes, such networks may constitute those where genetic variants are most likely to associate with ‘*trans*-eQTL hotspots’, and explain why the latter often overlap with loci containing association signals for complex diseases. In mouse RIS hearts (where the *Ypel5* locus is linked to the expression of the ‘*trans*-eQTL’ genes), QTL-dependent differences in gene expression may be due to differences in either cardiomyocytes themselves or in other cell types (including resident or infiltrating macrophages). Regardless of the cell type, the pathways regulated by *Ypel5* have been implicated in cardiac pathophysiology. For instance, several interferon regulatory factors play critical roles in the modulation of left ventricular remodeling^{313,315}. Likewise, invalidation of the IKBKE gene in mice leads to cardiac hypertrophy, fibrosis, and cardiac dysfunction, indicating a crucial role for IKBKE in regulating cardiac hypertrophy³³⁶.

In addition to their IFNB1 regulatory properties, both TBK1 and IKBKE promote oncogenesis^{317,319}. IKBKE is frequently overexpressed in a number of human cancers, in particular breast, ovarian and pancreatic cancer³⁴⁰, and inhibition of either kinase suppresses the growth of HER2-driven breast cancer³⁴¹. Of note, TBK1 is constitutively expressed in most cell types, whereas basal IKBKE expression is observed in specific tissues and cell types and is rapidly inducible by cytokines³⁴⁰. Although IKBKE and TBK1 show many overlapping activities that are relevant to innate immunity and inflammation, the more restricted expression of IKBKE might be of a particular interest in regards to its regulation by *YPEL5*. It will thus

be important to further determine the regulatory role of YPEL5 with the relative contribution of IKBKE and TBK1 in different pathways (e.g. NF- κ B versus IRF3) and pathophysiological processes.

In summary, biologic validation of genomic analyses allowed us to discover that *Ypel5* is a negative regulator of the IFNB1 production and innate immune responses, and that it acts primarily at the level of the TBK1/IKBKE kinases. Although the function of this gene in vertebrates was not known, its high level of evolutionary conservation suggested that it may play important roles. Interestingly it is an ortholog of *Yippee* in *Drosophila*, where it interacts with hemolin³⁰⁹, a protein that participates to anti-viral processes in insects³⁴². It is thus possible that the *Yippee* orthologs that have evolved in vertebrates have been co-opted for other innate immune responses.

2.5 Experimental Procedures

2.5.1 Animals and ethics statement

Results of gene expression studies were obtained in hearts from a panel of 24 male AxB/BxA recombinant inbred mice, as described previously^{198,343}. In these previous studies, animals were euthanized by cervical dislocation, following approval by the animal ethics committee of the Institut de Recherches Cliniques de Montréal (IRCM) and in agreement with the guidelines of the Canadian Council for Animal Care.

2.5.2 Cells lines and culture

RAW 264.7 and J774 cells (mouse monocytic/macrophagic cells) and HEK 293T (human embryonic kidney) cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM, Wisent), supplemented with 10% fetal bovine serum, 100 units/ml penicillin, 100 μ g/ml streptomycin and 2 mM glutamine (all from Wisent).

For siRNA transfections, the scrambled MISSION® siRNA SIC001 (Sigma-Aldrich, St-Louis, MO) was used as a negative control. For silencing mouse *Ypel5*, three separate specific siRNAs were obtained: SASI_Mm01_00066332, SASI_Mm01_00066328, and

SASI_Mm01_00066329, identified in the text as si-*Ypel5*-1, si-*Ypel5*-2, si-*Ypel5*-3, respectively. For silencing human *YPEL5*, three separate specific siRNAs were obtained: SASI_Hs02_00329464, SASI_Hs02_00329465, and SASI_Hs02_00329466. For mouse *Ifnb1*, we used the SASI_Mm01_00145086 siRNA.

Transfection of siRNAs was performed using either HiPerfect reagents (Qiagen, Canada) in mouse cells or Lipofectamine RNAi-Max (Life Technologies, Canada) in human cell lines, according to each manufacturer's instructions. Transfection of plasmids in human cells was performed with Lipofectamine 2000 (Life Technologies, Canada). For Western blot experiments, cells were infected with 120 HAU/ml of Sendai virus (Cantell Strain, Charles River Labs), as described previously³¹⁷.

2.5.3 Expression vectors

Eukaryotic expression pcDNA3.1 plasmids encoding cGAS or STING, as well as FLAG- and Myc-tagged eYFP, MAVS, TBK1, IKKBE and IRF3(5D), as described previously^{344,345} The plasmid for expression of the 3xFLAG-YPEL5 fusion protein was constructed by adding 3 consecutive FLAG repeats to the N-terminal end of the *Ypel5* coding sequence (amplified by PCR from a cDNA) and cloning the product into pcDNA3.1.

2.5.4 Structural protein homology analysis

Sequences were input into protein homology/analogy recognition engine (Phyre) version 0.2 (<http://www.sbg.bio.ic.ac.uk/phyre/>). Threading analysis was performed with the Phyre algorithm³¹⁴ to identify templates bearing structural homology to YPEL5.

2.5.5 Gene expression analyses

Total RNA was extracted from harvested cells using RNA Micro extraction kits (Qiagen, Canada). RNA concentrations were quantified by measuring absorbance at 260 nm and 280 nm with a Nanodrop apparatus (ThermoScientific, Wilmington, DE). Relative expression of specific genes vs. that of the normalizing *Rps16* housekeeping gene were performed by RT-qPCR. In short, we calculated ΔCt values ($Ct_{\text{gene}} - Ct_{\text{Rps16}}$), and values for normalized transcript abundance were expressed as $2^{-\Delta Ct}$. Primers used are listed in Table S4. Gene

expression profiling in transfected RAW 264.7 cells was performed using Illumina MouseRef-8 v2.0 BeadChips. Results were analyzed using the “two-sample Bayesian t-test for microarray data”³⁴⁶, using P -value < 0.005 and FDR < 0.1 as cut-off thresholds, and deposited under accession number GSE81150.

2.5.6 Luciferase assays

293T cells were cultured in white opaque 96-well plates with complete phenol red-free DMEM. The following day, cells were transfected with 1.8 pmol siRNA (either siYPEL5 or universal negative control). Twenty-four hours after siRNA transfection, cells were transfected with one of the several expression plasmids, along with one of the luciferase reporter constructs, i.e. pIFNB1-LUC, pISG56-LUC or p2xNF- κ B-LUC³¹⁷. Cells were lysed 24h after the second transfection, and luciferase activities were determined by a dual-luciferase reporter assay system (Promega, Canada).

2.5.7 Western immunoblot analyses

Protein extracts were boiled for 5 min. Samples were separated on SDS-PAGE gel and transferred to either nitrocellulose or polyvinylidene difluoride membranes. Protein signals were detected using the Western Lighting Chemiluminescence Reagent Plus (PerkinElmer). Each experiment was repeated at least 3 times. Primary antibodies used for Western blot analyses were as follows: anti-RIG-I (ALX-210-932-C100; Alexis Biochemicals), anti-MDA5 (210-935-c100; Alexis Biochemicals), anti-actin (MAB1501, Chemicon International), anti-FLAG (F3165; Sigma); anti-myc (SC-040; Santa Cruz); anti-TBK1 (SC-52957; Santa Cruz); anti-IKBKE (SC-365830; Santa Cruz); anti-phospho-TBK1 (Ser172) (5483; Cell Signaling); and anti-phospho-IKBKE (Ser172) (8766; Cell Signaling).

2.5.8 Co-immunoprecipitation

HEK 293T cells were co-transfected with 5 μ g of either 3xFlag-YPEL5 or YFP-FLAG, along with 5 μ g of an expression plasmid of Myc-tagged sequences of either IKBKE, TBK1, or MAVS. Thirty-two hours later, cells were infected (or not) with 100 μ l containing 600 HAU/ml of Sendai virus (Cantell Strain, Charles River Labs), as described previously³¹⁷. After another 16 hours, cells were harvested, lysed, and the concentration of total proteins in the

lysates was adjusted to 1 mg protein/ml. The lysates were precleared by incubation for one hour with 40 μ l of a slurry of immunoglobulin G-Sepharose (GE Healthcare), and incubated overnight with agitation at 4°C in the presence of 20 μ l of M2 anti-FLAG affinity gel (Sigma-Aldrich). After centrifugation, the immunoprecipitates were washed five times in lysis buffer, and elution was performed by incubation in the presence of purified FLAG peptide (250 ng/ μ l) (Sigma-Aldrich) for 45 min at 4 °C. The eluates were analyzed by Western immunoblotting. For reverse co-IP, the cells were transfected with YPEL5-Myc along with FLAG-tagged YFP, TBK1 and IKBKE.

2.5.9 Biotin-RNA pull down assay

Biotin labeled poly(I:C) RNA was purchased from InvivoGen, Biotin-HCV RNA was obtained by subjecting linearized HCV DNA to T7 reverse transcription (TranscriptAid T7 High Yield, Life Technologies) and biotin-dUTP (Enzo Life Sciences). HEK293T cells were transfected with either 3xFlag-YPEL5 or RIG-I expression plasmids or the control 3xFlag-MCS (empty plasmid) for 32 hours, then treated with either SeV or vehicle for 16 hours. For the RNA pulldown, Dynabeads M270 Streptavidin (Life Technologies) were incubated with the biotin-labeled RNA substrates for 15 min with agitation at room temperature, and then incubated for one hour with agitation at 4°C with aliquots of cell lysates containing a total of 100 μ g proteins. After magnetic isolation, the beads were washed three times, RNA-bound proteins were eluted by boiling in 0.1% SDS, and the eluates were analyzed by Western blot.

2.5.10 Ablation of the *YPEL5* gene in HEK293T cells by RNA-guided nuclease technology

Guide RNA sequences were designed to target two sites in 4 of h*YPEL5*, and cloned into the pSPCas9(BB)-2A plasmid, as described³⁴⁷. Transfection of the resulting plasmids into HEK 293T cells caused genomic deletions of 164 bp in exon 4 (between positions 30,158,609 and 30,158,772 in chr2). Single cell clones were isolated by clonal dilution of the transfected cells. To screen for clonal cell lines harboring the appropriate deletions, PCR amplification of the genomic DNA was performed using primers flanking the deletion sites. Invalidation of *YPEL5* was further confirmed by verifying (by RT-qPCR) that *YPEL5* was no longer expressed at detectable levels in the selected clonal lines.

2.5.11 Statistics

According to the specifics of each experiment, comparisons between groups were performed by either Student's t-tests, one-way or two-way ANOVAs, with the ANOVA analyses being followed by Tukey's multiple comparison tests.

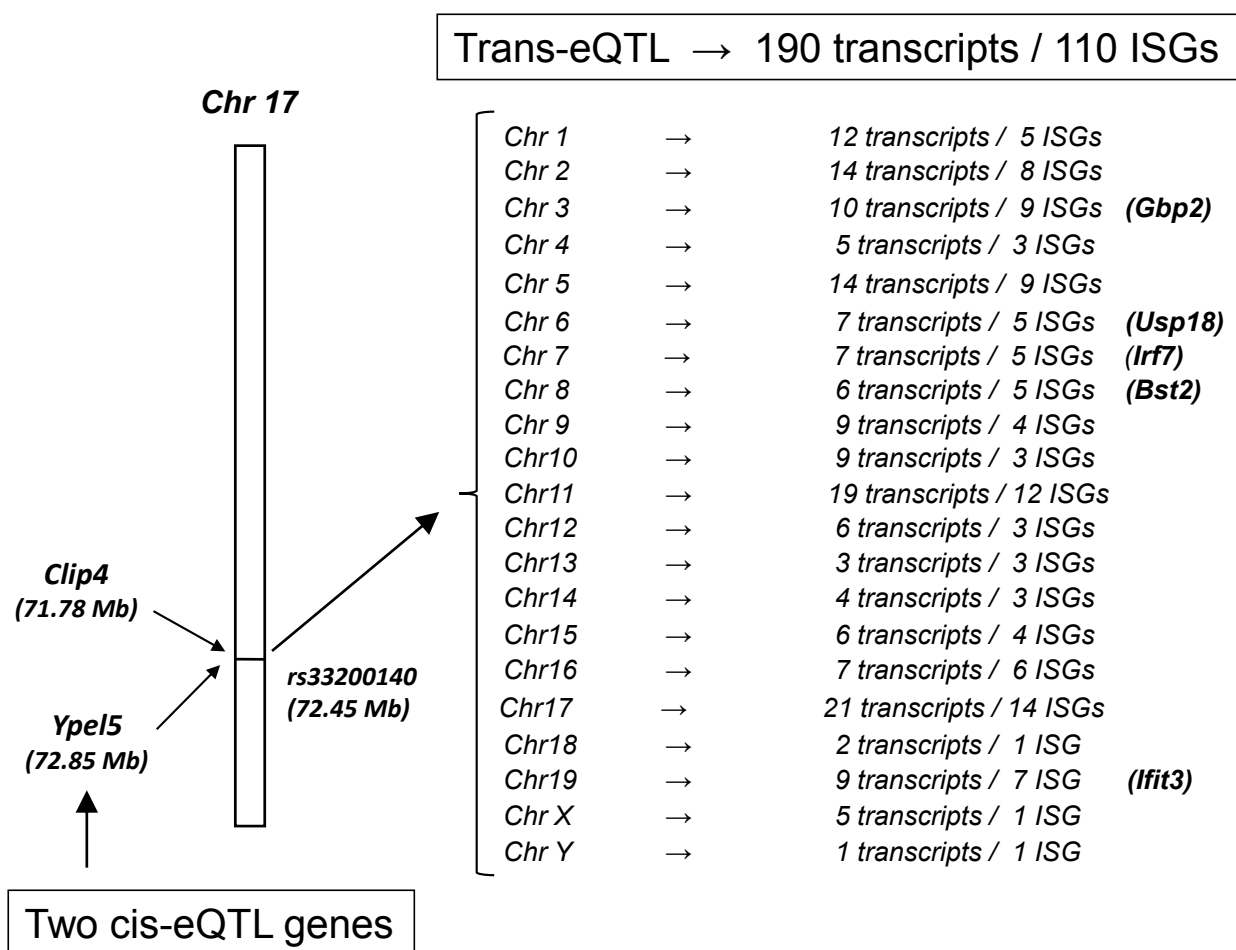


Figure 2.1: Chromosomal distribution of eQTL genes linked either in *cis*- or *trans*- to rs33200140 on chr17. Total transcript count is showed for each chromosome, many of which are Interferon Stimulated Genes (ISGs) transcripts according to the interferome v2.0 database. Genes in bold were those chosen for further functional characterization in subsequent experiments.

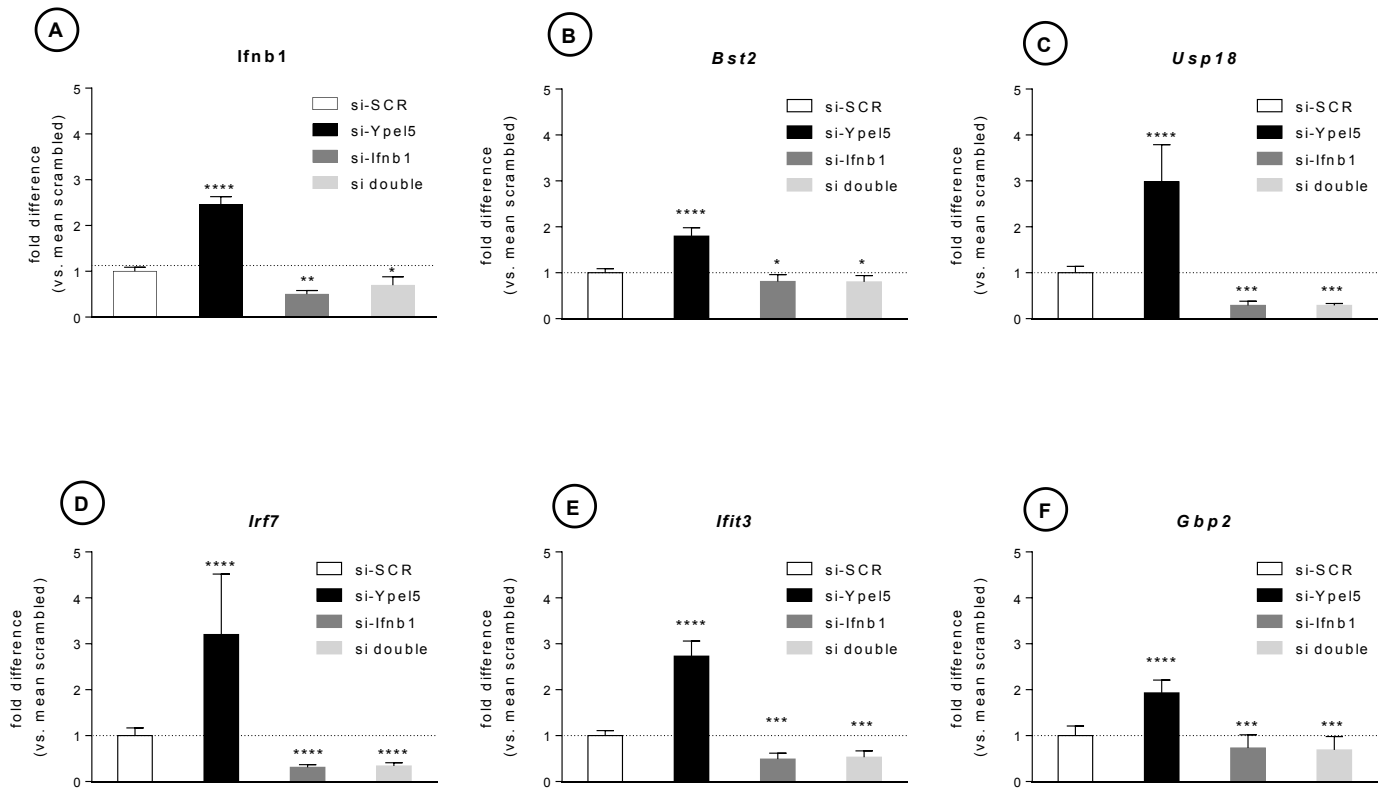


Figure 2.2: *Ypel5* negatively regulates *trans*-eQTL genes in RAW264.7 cells in an *Ifnb1*-dependent manner. (A-F)Relative abundance of mRNA transcripts of *Ifnb1*, *Bst2*, *Usp18*, *Irf7*, *Ifit3*, and *Gbp2* in RAW264.7 cells transfected for 24 hours with siRNA (either scrambled or specific for *Ypel5* or *Ifnb1*, either alone or in combination). In cells transfected with *Ifnb1* siRNA alone, *Ifnb1* expression levels decreased to an average of 45% of that observed in cells transfected with scrambled siRNA, and no further increase was observed after *siYpel5* transfection. All values represent the means (\pm SD) of results obtained in duplicate wells from 4 individual experiments ($n = 8$). Significance of differences was determined by one-way ANOVA followed by Tukey's multiple comparison post-hoc tests (**** $P < 0.0001$; *** $P < 0.001$).

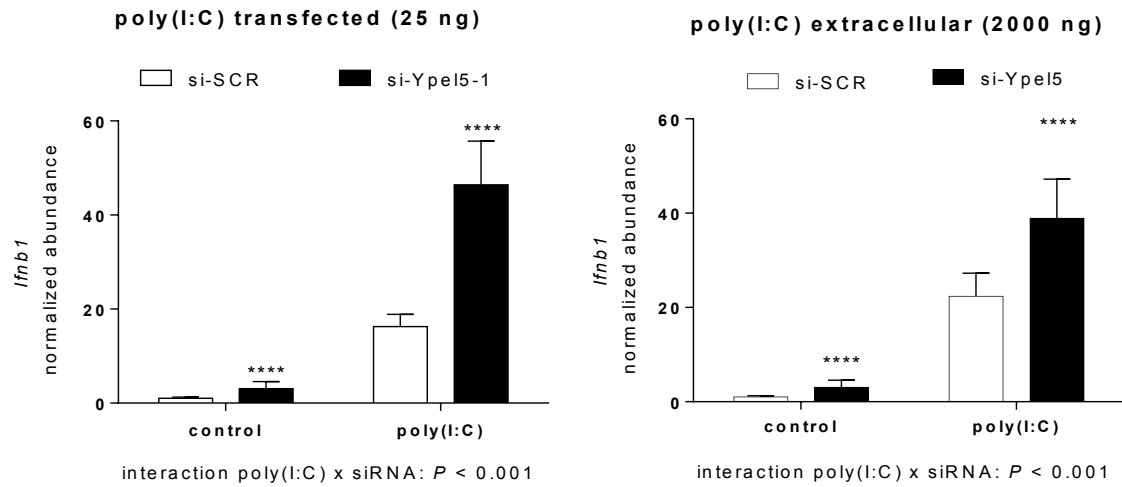


Figure 2.3: *Ypel5* downregulation amplifies the effects of either intracellular or extracellular poly (I:C) on *Ifnb1* induction. Expression of *Ifnb1* in RAW 267.4 cells transfected with either scrambled or *Ypel5* siRNA, either in basal conditions or after transfection with 25 ng poly (I:C) (left) or exposure to extracellular poly(I:C) (2000 ng/ml). All values represent the means (\pm SD) of results obtained in replicate wells from 2 individual experiments ($n = 14$). Significance of differences was determined by 2 way-ANOVA analysis. **** $P < 0.0001$ vs scrambled by Tukey's multiple comparison post-hoc tests.

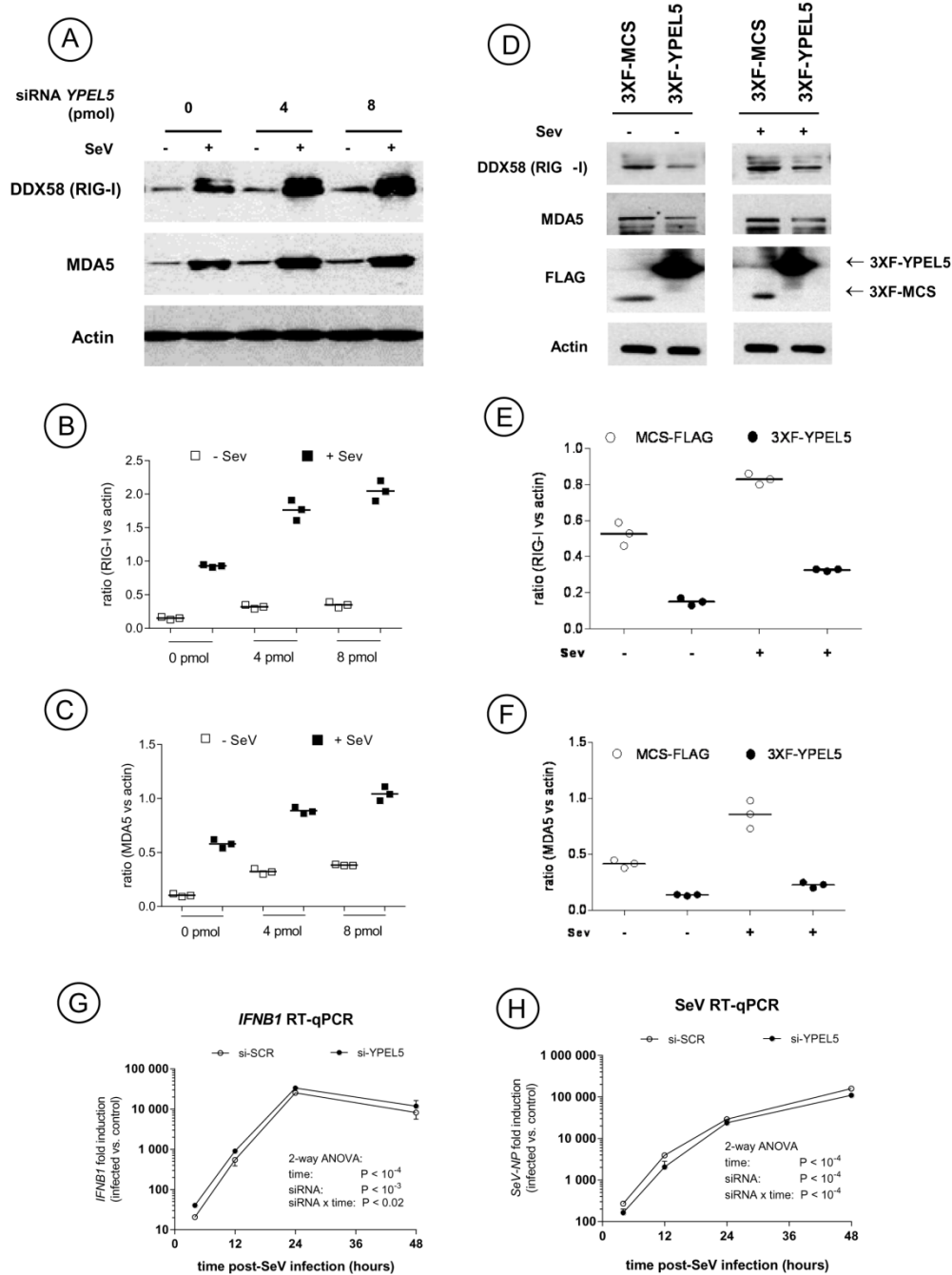


Figure 2.4: Effects of *YPEL5* down-regulation or overexpression in HEK293T cells with or without SeV infection. (A-C) Immunoblot analyses of HEK293T cells transfected with combinations of scrambled and *YPEL5* siRNA (either 8 pmol scrambled, 4 pmol scrambled and

4 pmol si-*YPEL5*, or 8 pmol si-*YPEL5*) for 32 hours, then infected (or not) with Sev for 16 hours. Immunoblotting for actin shows that equivalent amounts of proteins were loaded in each well. The blots are representative results of one experiment (Figure 2.4A), quantification of the results obtained in replication experiments are as shown in Figure 2.4B and Figure 2.4C (n = 3). (D-F): Immunoblot analyses of HEK 293T transfected with either 3XF-MCS (empty plasmid) or the 3XF-*YPEL5* expression plasmid for 32 hours, then infected (or not) with SeV for 16 hours. Immunoblotting with the FLAG antibody shows that the 3x-FLAG proteins are correctly expressed. The blots are representative results of one experiment (Figure 2.4D), quantification of the results obtained in replication experiments are as shown in Figure 2.4B and Figure 2.4C (n = 3). (G-H): RT-qPCR analyses of the mRNA abundance of either IFNB1 (G) or SeV NP (H). The extracts came from HEK 293T cells transfected with either scrambled or *YPEL5* siRNA for 24 hours, and further infected with SeV for 12, 24, 36, or 48 hours. The results of two-way ANOVA analyses are as indicated.

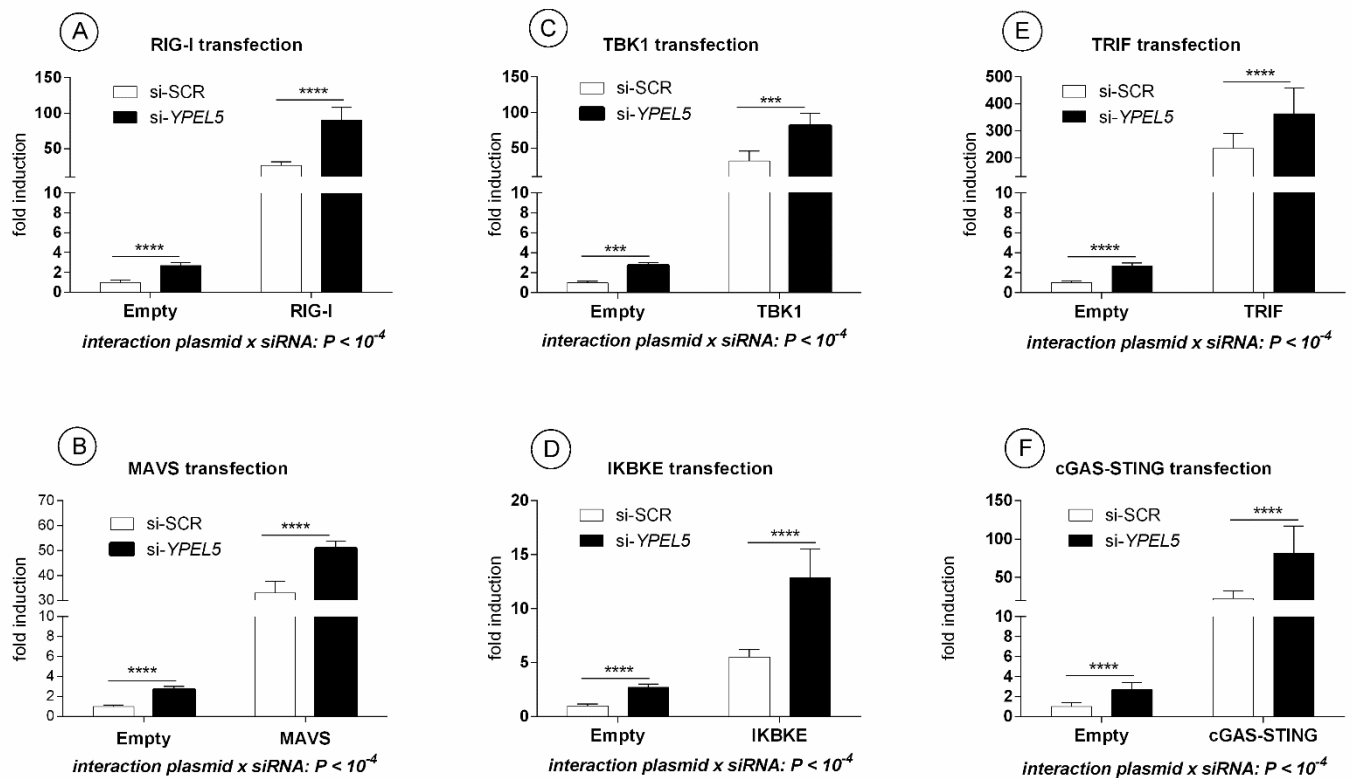


Figure 2.5: *YPEL5* siRNA potentiates the induction of IFNB1-dependent luciferase activity by signaling events that activate the TBK1/IKBKE kinases. Twenty four hours after transfection with *YPEL5*-specific siRNA, HEK293T cells were co-transfected with the pIFNB1-LUC reporter plasmid, along with either pcDNA3.1 (empty plasmid) or expression plasmids for RIG-I, MAVS, TBK1, IKBKE, TRIF or cGAS-STING. Luciferase activity was measured 24 h after transfection. All values represent the means (\pm SD) of results obtained in replicate wells from 3-4 individual experiments ($n = 16-24$). Interactions between the effects of the stimuli with that of the siRNAs were tested by two-way ANOVAs; significance values of the interactions between si-*YPEL5* and stimuli are as indicated. Asterisks indicate significant differences between cells transfected with either scrambled or *YPEL5* siRNA, as determined by Tukey's pot-hoc tests (**** $P < 0.0001$; ** $P < 0.01$).

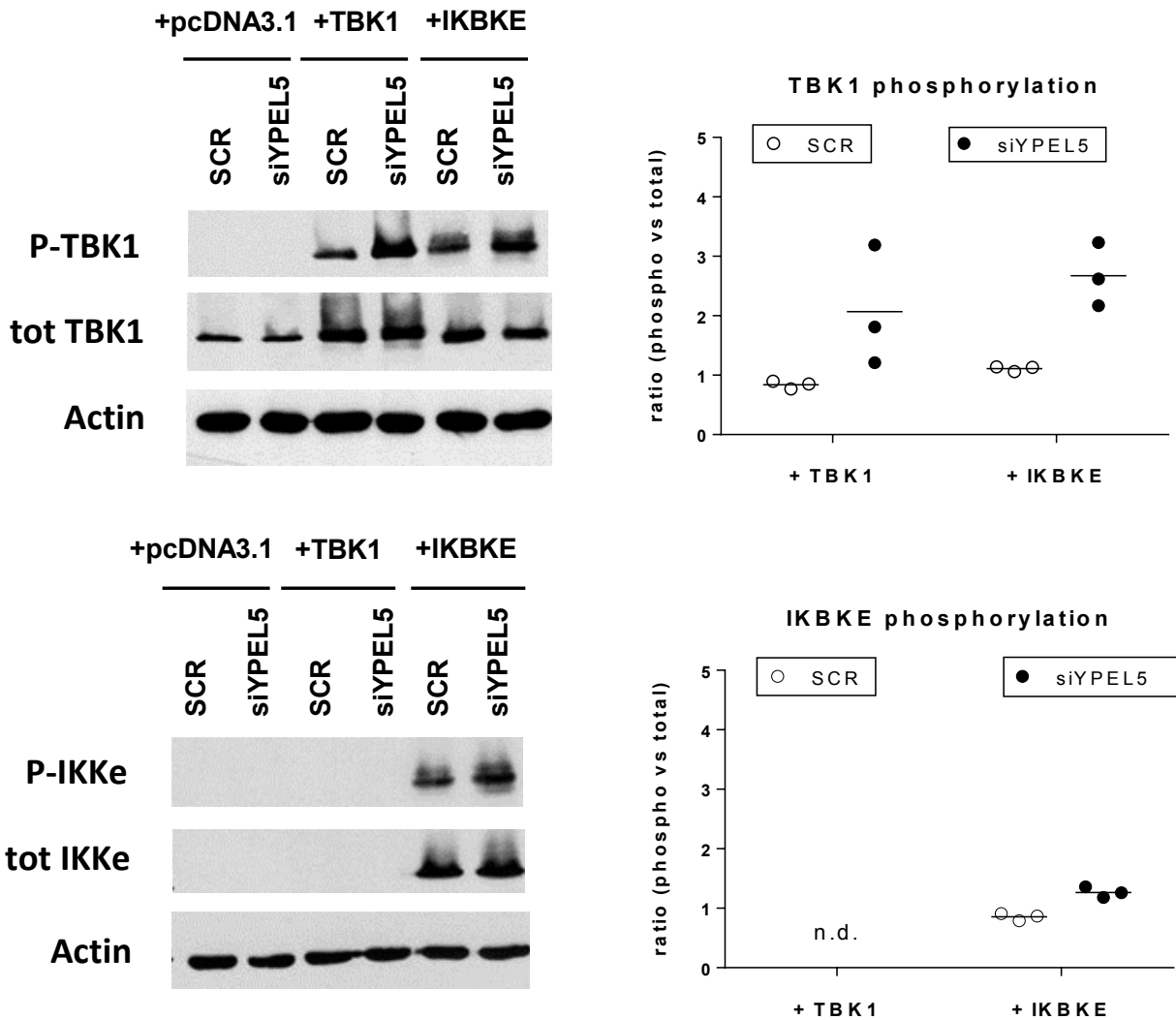


Figure 2.6: *YPEL5* silencing increased the effects of overexpressed IKKε and TBK1 on their respective auto- or cross-phosphorylation. HEK293T cells treated with either scrambled or *YPEL5* siRNA for 24 hours, then transfected for an additional 24 hours with either pcDNA3.1 (empty plasmid) or expression plasmids for either TBK1, IKKε. Phosphorylation levels at Ser172 were determined by calculating the ratio of the signals obtained with phospho-specific antibodies vs. that obtained with antibodies against total proteins. The levels observed in wells transfected with siYPEL5 showed no overlap with the lower levels observed in cells transfected with scrambled siRNA. The charts on the right show the results obtained in 3 replicate experiments.

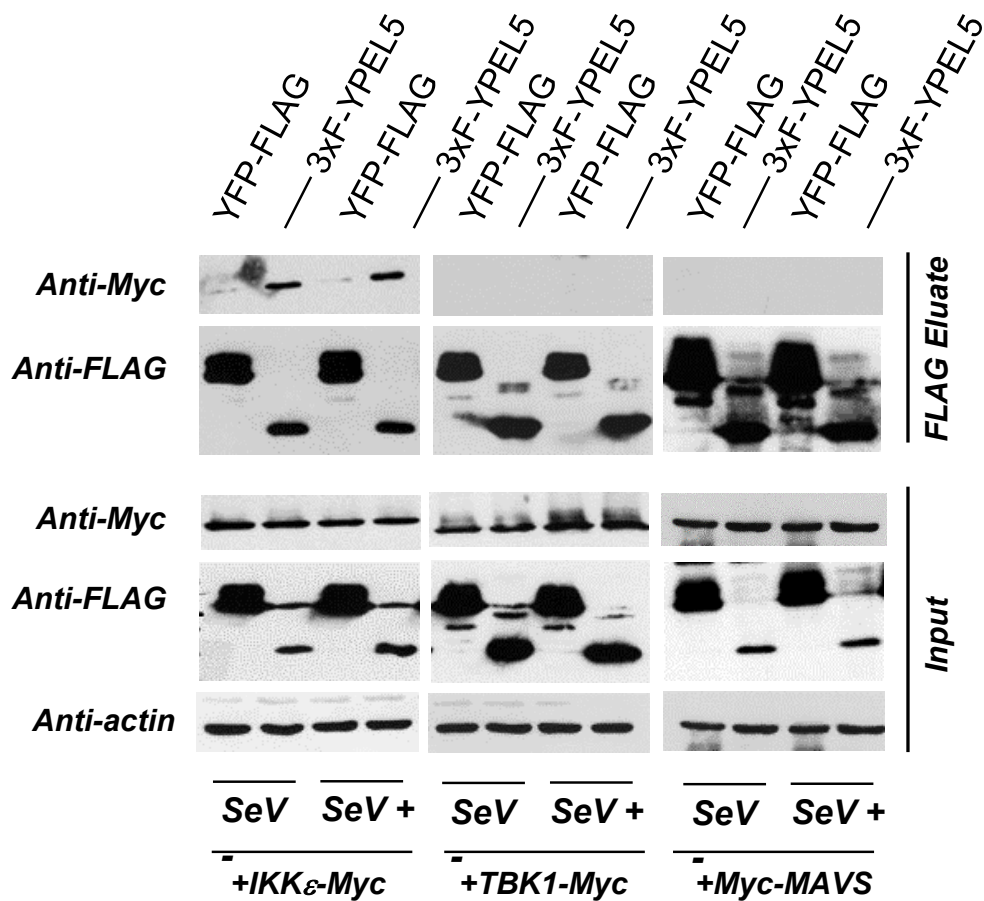


Figure 2.7: YPEL5 interacts physically with IKK ϵ in transfected cells. HEK293T cells were transfected with either FLAG-YFP (negative control) or 3xFLAG-YPEL5, as well as IKK ϵ -myc, TBK1-myc or myc-MAVS expressing plasmids for 32 hours. Cells were either uninfected (SeV -) or infected with SeV (SeV+). After another 16 h, immunoprecipitation was performed using anti-FLAG antibodies. Either cell extracts (lysate) or eluates from immunoprecipitated complexes (IP FLAG) were analyzed by Western blotting. The images are representative images of results obtained in three replicate experiments.

Table S1a (related to Figure 2.1)Enrichment of Gene Ontology (GO) terms within *trans*-eQTL genes

GO term	Term description	Gene count	Gene %	Fold enrich.	Bonferroni
GO:0006955	Immune response	23	2.5	8.84	1.96E-12
GO:0017076	purine nucleotide binding	26	2.8	2.4	0.004
GO:0019884	antigen processing and presentation	5	0.54	32.4	0.009

Table S1b (related to Figure 2.1)

Enrichment of Gene Ontology (GO) terms within genes whose expression is affected by Ypel5 silencing in RAW264.7 cells

GO term	Term description	Gene count	Gene %	Fold enrich.	P-value i
GO:0005739	mitochondrion	65	11.2	1.7	6.62E-08
GO:0000166	nucleotide binding	52	5.5	1.4	3.62E-06
GO:0045087	innate immune response	15	1.6	2.83	7.83E-04
GO:0002684	positive regulation of immune system	22	2.3	2.3	8.99E-04
GO:0043067	regulation of programmed cell death	150	16.1	1.7	2.60E-04

Table S2: (related to Figure 2.1)List of *trans*-eQTL genes reported in the “interferome database” as ISGs

Gene symbol	Ensembl Id	Entrez	A/B ratio
Adar	ENSMUSG00000027951	56417	1.26
Aif1	ENSMUSG00000024397	11629	1.20
Aim1	ENSMUSG00000019866	11630	1.10
ANTXR1	ENSG00000169604	84168	0.92
Apobec1	ENSMUSG00000040613	11810	1.19
Apo19b	ENSMUSG00000068246	71898	1.30
B2m	ENSMUSG00000060802	12010	1.47
BBS9	ENSG00000122507	27241	0.91
Bclaf1	ENSMUSG00000037608	72567	0.85
Bst2	ENSMUSG00000046718	69550	1.85
Ccl4	ENSMUSG00000018930	20303	1.26
Cd274	ENSMUSG00000016496	60533	2.29
Cd52	ENSMUSG00000000682	23833	1.43
Cggbp1	ENSMUSG00000054604	106143	0.88
Ch25h	ENSMUSG00000050370	12642	1.36
Csprs	ENSMUSG00000062783	100503923	1.21
CTDSPL	ENSG00000144677	10217	0.90
Ctss	ENSMUSG00000038642	13040	1.18
Cxcl10	ENSMUSG00000034855	15945	1.24
Ddah2	ENSMUSG00000007039	51793	0.86
Dhx58	ENSMUSG00000017830	80861	1.42
Dpy30	ENSMUSG00000024067	66310	0.89
Eif2ak2	ENSMUSG00000024079	19106	2.18
Fam92a	ENSMUSG00000028218	68099	0.91
Fbxw17	ENSMUSG00000037816	109082	1.20
Fcgr1	ENSMUSG00000015947	14129	1.17
Fcgr4	ENSMUSG00000059089	246256	1.79
Flywch1	ENSMUSG00000040097	224613	1.04
Gbp10	ENSMUSG00000054588	626578	2.01
Gbp2	ENSMUSG00000028270	14469	1.82
Gbp3	ENSMUSG00000028268	55932	1.82
Gbp5	ENSMUSG00000040264	229898	1.41
Gbp6	ENSMUSG00000079362	100702	1.46
H2-M3	ENSMUSG00000016206	14991	1.38
Hdc	ENSMUSG00000027360	15186	1.34
Helz2	ENSMUSG00000027580	229003	1.27
Idnk	ENSMUSG00000050002	75731	1.22
Ifi27	ENSMUSG00000064215	52668	1.75
Ifi35	ENSMUSG00000010358	70110	1.62
Ifi44	ENSMUSG00000028037	99899	1.10
Ifi47	ENSMUSG00000078920	15953	1.73
Ifit2	ENSMUSG00000045932	15958	1.93
Ifit3	ENSMUSG00000074896	15959	3.29
Ifitm3	ENSMUSG00000025492	66141	1.53
Igtp	ENSMUSG00000078853	16145	3.89
Irf1	ENSMUSG00000018899	16362	1.41
Irf7	ENSMUSG00000025498	54123	2.17
Irf9	ENSMUSG00000002325	16391	1.69
Irgm1	ENSMUSG00000046879	15944	1.52
Lgals3bp	ENSMUSG00000033880	19039	2.00
Lgals9	ENSMUSG00000001123	16859	1.43
Lims2	ENSMUSG00000024395	225341	1.13
LTBP1	ENSG00000049323	4052	1.08

Ly6a	ENSMUSG00000075602	110454	1.43
Ly6e	ENSMUSG00000022587	17069	1.40
LY86	ENSG00000112799	9450	1.18
MFAP5	ENSG00000197614	8076	1.17
Mkl1	ENSMUSG00000012519	74568	1.16
Ms4a6d	ENSMUSG00000024679	68774	1.31
Mx2	ENSMUSG00000023341	17858	1.80
MYOM1	ENSG00000101605	8736	0.76
Ncbp2	ENSMUSG00000022774	68092	0.92
Nlrc5	ENSMUSG00000074151	434341	1.62
Oas2	ENSMUSG00000032690	246728	1.46
Oasl1	ENSMUSG00000041827	231655	1.53
Oasl2	ENSMUSG00000029561	23962	1.92
Ogfr	ENSMUSG00000049401	72075	1.39
Parp12	ENSMUSG00000038507	243771	1.47
Parp14	ENSMUSG00000034422	547253	1.75
Phf11d	ENSMUSG00000068245	219132	1.84
PHTF2	ENSG00000006576	57157	0.84
PIK3CA	ENSG00000121879	5290	0.87
Plac8	ENSMUSG00000029322	231507	1.48
Pld4	ENSMUSG00000052160	104759	1.18
Pml	ENSMUSG00000036986	18854	1.21
PRPS2	ENSG00000101911	5634	0.81
Psmb10	ENSMUSG00000031897	19171	1.46
Psmb8	ENSMUSG00000024338	16913	1.46
Psme1	ENSMUSG00000022216	19186	1.12
Rnf31	ENSMUSG00000047098	268749	1.15
Rsad2	ENSMUSG00000020641	58185	2.19
Rtp4	ENSMUSG00000033355	67775	1.27
RXRG	ENSG00000143171	6258	1.19
Samhd1	ENSMUSG00000027639	56045	1.29
Sepw1	ENSMUSG00000041571	20364	1.16
Siglec1	ENSMUSG00000027322	20612	1.24
Slamf9	ENSMUSG00000026548	98365	1.27
SLC15A3	ENSG00000110446	51296	1.23
SLC2A6	ENSG00000160326	11182	1.17
Slfn1	ENSMUSG00000078763	20555	1.27
Slfn2	ENSMUSG00000072620	20556	1.18
SPCS3	ENSG00000129128	60559	0.90
Stat1	ENSMUSG00000026104	20846	2.13
Stat2	ENSMUSG00000040033	20847	1.69
Tap1	ENSMUSG00000037321	21354	1.47
Tap2	ENSMUSG00000024339	21355	1.39
Tcirg1	ENSMUSG00000001750	27060	1.13
Tdrd7	ENSMUSG00000035517	100121	1.22
Tmem106a	ENSMUSG00000034947	217203	1.15
TMEM30A	ENSG00000112697	55754	0.85
Trafd1	ENSMUSG00000042726	231712	1.20
TREX1	ENSG00000213689	11277	1.15
Trim21	ENSMUSG00000030966	20821	1.36
Trim26	ENSMUSG00000024457	22670	1.21
TRPS1	ENSG00000104447	7227	0.93
Tspo	ENSMUSG000000041736	12257	1.17
Ube2l6	ENSMUSG00000027078	56791	1.38
Usp18	ENSMUSG00000030107	24110	2.85
Wars	ENSMUSG00000021266	101056688	1.25
Xdh	ENSMUSG00000024066	22436	1.59

Table S2: The A/B ratios were calculated according to expression levels of each gene in hearts from 24 recombinant inbred AxB/BxA mouse strains, using 4 males from each strain and results from Illumina MouseRef-8 v2.0 BeadChips. The ratio corresponds to the mean value of expression in RIS strains carrying the allele from A/J compared to that carrying the allele from C57BL/6J mice, as determined from the genotype of SNP rs33200140, located on chr17 at position 72.449 Mb. Highlighted genes correspond to those whose expression was further tested in mouse macrophage cell lines by RT-qPCR.

Table S3 (related to Figure 2.1)

Comparisons of 50 *trans*-eQTL genes with largest fold differences in mouse hearts to genes affected by *Ypel5* silencing in RAW264.6 cells

Gene symbol	Fold difference in mouse hearts (A vs B)	Fold change after <i>Ypel5</i> siRNA in RAW264.6 cells	P-value	FDR
Igtp	3.89	1.31	1.25E-03	0.15
Ifit3	3.29	2.01	3.45E-04	0.102
Iigp2	2.90	1.42	8.23E-06	1.36E-03
Usp18	2.85	1.57	3.70E-08	1.73E-05
Cd274	2.29	1.39	0.122	0.596
Rsad2	2.19	3.79	4.44E-16	1.14E-11
Eif2ak2	2.18	1.17	1.90E-02	0.376
Irf7	2.17	1.57	7.39E-05	0.064
Stat1	2.13	1.27	6.53E-04	0.032
Gbp10	2.01	1.52	2.12E-01	0.714
Lgals3bp	2.00	1.16	9.61E-05	0.008
Ifit2	1.93	1.36	2.53E-04	0.017
Ube11	1.92	1.22	9.25E-05	0.008
Oasl2	1.92	1.68	3.46E-06	7.00E-04
Bst2	1.85	1.12	7.88E-02	0.508
Phf11d	1.84	1.32	1.63E-06	3.77E-04
Gbp2	1.82	1.43	2.72E-01	0.765
Gbp3	1.82	2.05	5.67E-03	0.257
Mx2	1.80	1.65	6.13E-04	0.030
Fcgr4	1.79	1.67	5.50E-06	0.001
Ifi27	1.75	1.13	5.26E-03	0.125
Parp14	1.75	1.48	1.15E-11	4.20E-08
Ifi47	1.73	1.36	1.49E-02	0.226
Irf9	1.69	1.03	5.69E-01	0.907
Stat2	1.69	1.28	6.62E-02	0.473
Cmpk2	1.64	2.22	2.72E-10	3.78E-07
Ifi35	1.62	1.19	3.60E-03	0.10
Nlrc5	1.62	1.04	3.80E-01	0.823
Xdh	1.59	1.94	3.65E-04	0.022
Ifitm3	1.53	1.00	9.15E-01	0.987
Oasl1	1.53	1.89	1.76E-05	0.003

Irgm1	1.52	1.41	3.49E-07	1.11E-04
Plac8	1.48	1.90	5.33E-03	0.126
B2m	1.47	0.95	3.20E-02	0.338
Parp12	1.47	1.25	9.43E-04	0.040
Tap1	1.47	1.11	5.53E-01	0.901
Gbp6	1.46	1.02	9.34E-01	0.988
Oas2	1.46	1.30	4.36E-05	0.005
Psmb10	1.46	0.97	3.21E-01	0.796
Psmb8	1.46	1.16	8.88E-02	0.532
Lgals9	1.43	1.34	1.18E-03	0.047
Ly6a	1.43	1.46	1.91E-01	0.691
Cd52	1.43	0.97	0.353	0.817
Dhx58	1.42	1.10	2.05E-01	0.71
Gbp5	1.41	1.30	2.80E-01	0.769
Irf1	1.41	1.30	1.94E-01	0.696
Ly6e	1.40	1.00	9.66E-01	0.994
Oasg1	1.40	1.14	1.14E-02	0.197
Ogfr	1.40	1.18	4.83E-04	0.026
Tap2	1.40	1.18	9.98E-03	0.183

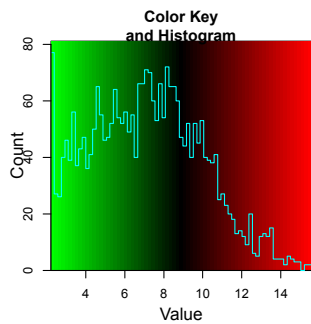
Table S3: List of the 50 trans-eQTL genes showing the largest fold differences in mouse hearts as a function of the allelic origin of the chr 17 locus (corresponding to a >0.4 fold difference), and comparison to the effect of *Ypel5* silencing on the same genes in RAW264.7 cells. A total of 26/50 genes corresponded to genes that were also affected significantly by *Ypel5* silencing in RAW264.7 cells (as highlighted). A total of 45/50 genes (formatted in bold) corresponded to genes listed as ISGs in the interferome database.

Table S4 (related to the “Experimental Procedures” section)

Primers used for RT-qPCR validation studies

Mouse genes	Forward	Reverse
<i>Ypel5</i>	CCT GAC CAA TCG CTC AGA AC	AGC ATG ACC CGA TCT TGA AC
<i>Bst2</i>	TGA AGT CAC GAA GCT GAA CC	ATG GAG CTG CCA GAG TTC AC
<i>Usp18</i>	TGA ACT CTT TGC CGT GAT TG	ACT GGA CAT CCT TCC AGG TG
<i>Irf7</i>	CCC ATC TTC GAC TTC AGC AC	TCA CCA GGA TCA GGG TCT TC
<i>Ifit3</i>	CCT TCC ACA GCT GAA GTG C	TAA ATG TTC GAC CTG GTT GC
<i>Gbp2</i>	ACT CAC AGA GGC AGC AAA GG	GCA TCA TCA GCT CGA ACT CC
<i>Ifnb1</i>	CAG CTC CAG CTC CAA GAA AG	TGC ACC CTC CAG TAA TAG CTC

Human genes	Forward	Reverse
<i>YPEL5</i>	CCA ACC GCT CAG AAC TCA TC	AGC ATG ACC CGA TCT TGA AC
<i>IFNBI</i>	CGA GCA CAA CAG GAG AGC AA	GAA GCA CAA CAG GAG AGC AA
<i>TNF</i>	CAG CCT CTT CTT CTC CTT CCT GAT	GCC AGA GGG CTG ATT AGA GA
<i>RIG-I</i>	TCA GCA CAA ACC ACA ACCTG	AAG CGT CTC CAA GGA CAGTG



Heatmap of expression levels of selected *trans*-eQTL genes across different mouse cell types and tissues (data from BioGPS)

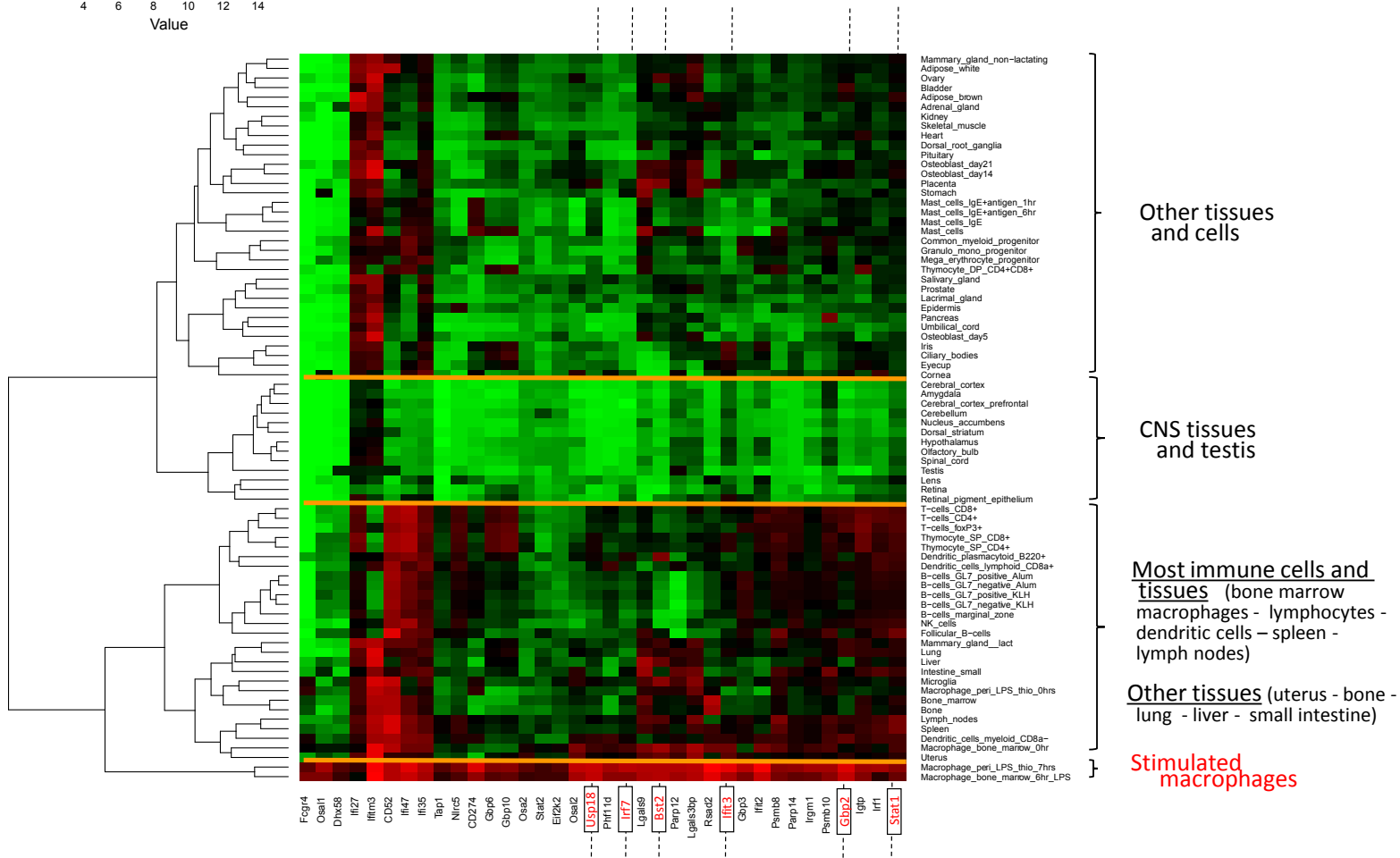


Figure S1 (related to Figure 2.1): Heatmap of expression levels of *trans*-eQTL genes across mouse cell types and tissues. Data were obtained from the BioGPS database. Expression levels of genes were color-coded as a function of strength of the hybridization signal, from lowest (green) to highest (red).

A

Homo sapiens	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVML
Macaca mulatta	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVML
Mus musculus	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVML
Rattus norvegicus	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVML
Bos Taurus	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVML
Ictalurus furcatus (fish)	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSDVQDRVML
Xenopus laevis	MGRIFLDHIGGTRLFSCANCNTILTNRSELISTRFTGATGRAFLFNKVVNLQYSEVQDRVML
Homo sapiens	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS
Macaca mulatta	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS
Mus musculus	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS
Rattus norvegicus	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS
Bos Taurus	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS
Ictalurus furcatus	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS
Xenopus laevis	TGRHM VRDVSCNKNCKLGIWYEFATEDSQRY KEGRVILERALVRESEGFEEHVPSDNS

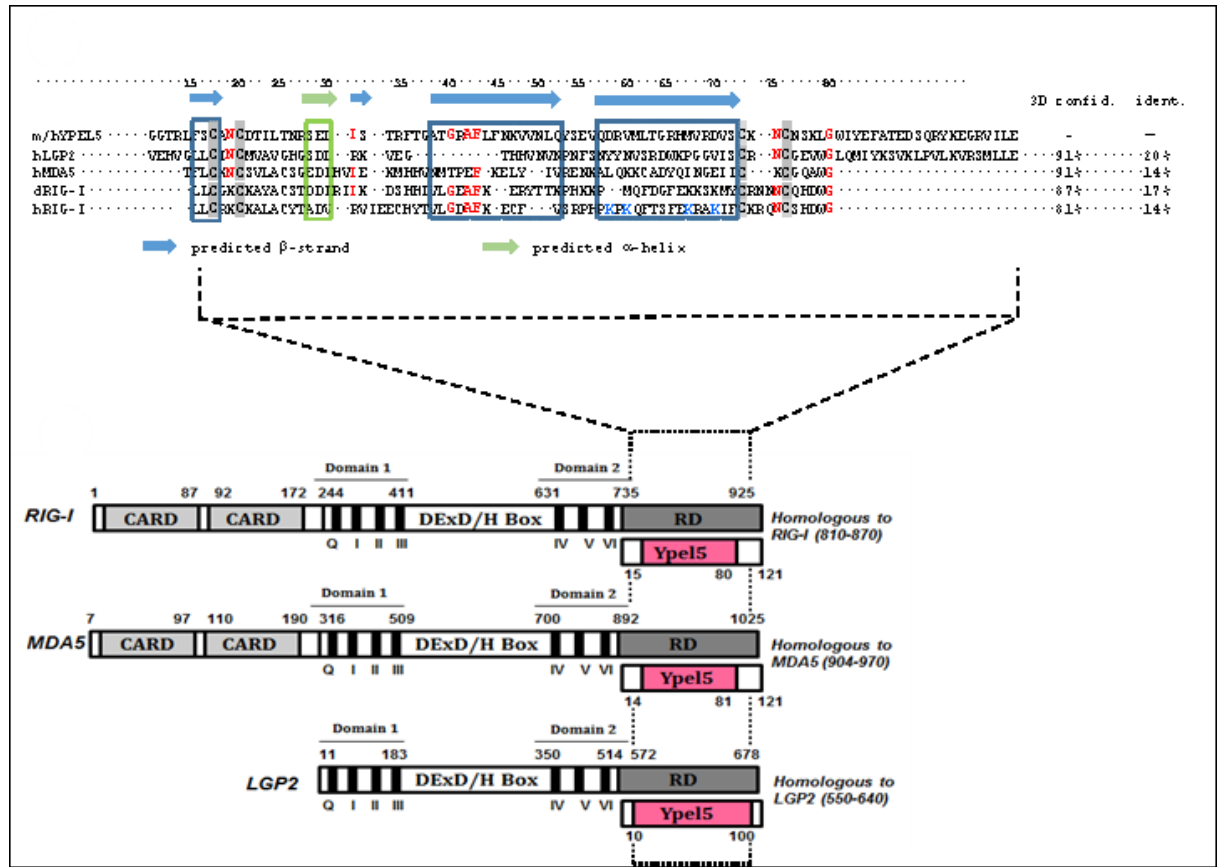
B

Figure S2 (related to Figure 2.1): Characteristics of the YPEL5 protein. Figure S2a: BLAST alignment of the amino acid sequence of the YPEL5 protein in 7 vertebrates. All amino acids are identical, with the exception of the neutral substitution of one glutamic acid residue (E) to a similarly negatively charged aspartic residue (D) in fish. Figure S2b: 3D structural homologies of mYpel5/hYPEL5 with members of the RLR family. The A.A. sequence alignment the regions of RIG-I/MDA5/LGP2 (either human or duck sequences) showing homology with YPEL5 (either mouse or human, as both are identical). Residues whose positions are conserved at the

linear level are shown in red. Zinc-coordinating cysteines are highlighted in grey. Lysine (K) residues shown in blue in the hRIG-I sequence correspond to those reported as critical for 5'-ppp RNA binding³²⁰ (and not conserved in the YPEL5 sequence) The full YPEL5 protein is 121 amino acids long; portions in the body of the YPEL5 protein (in pink) show 3D homologies with the regulatory domain (RD) of either RIG-I, MDA5 or LGP2; the graphic representation of the 3 proteins is adapted from Bruns and Horvath³¹⁵.

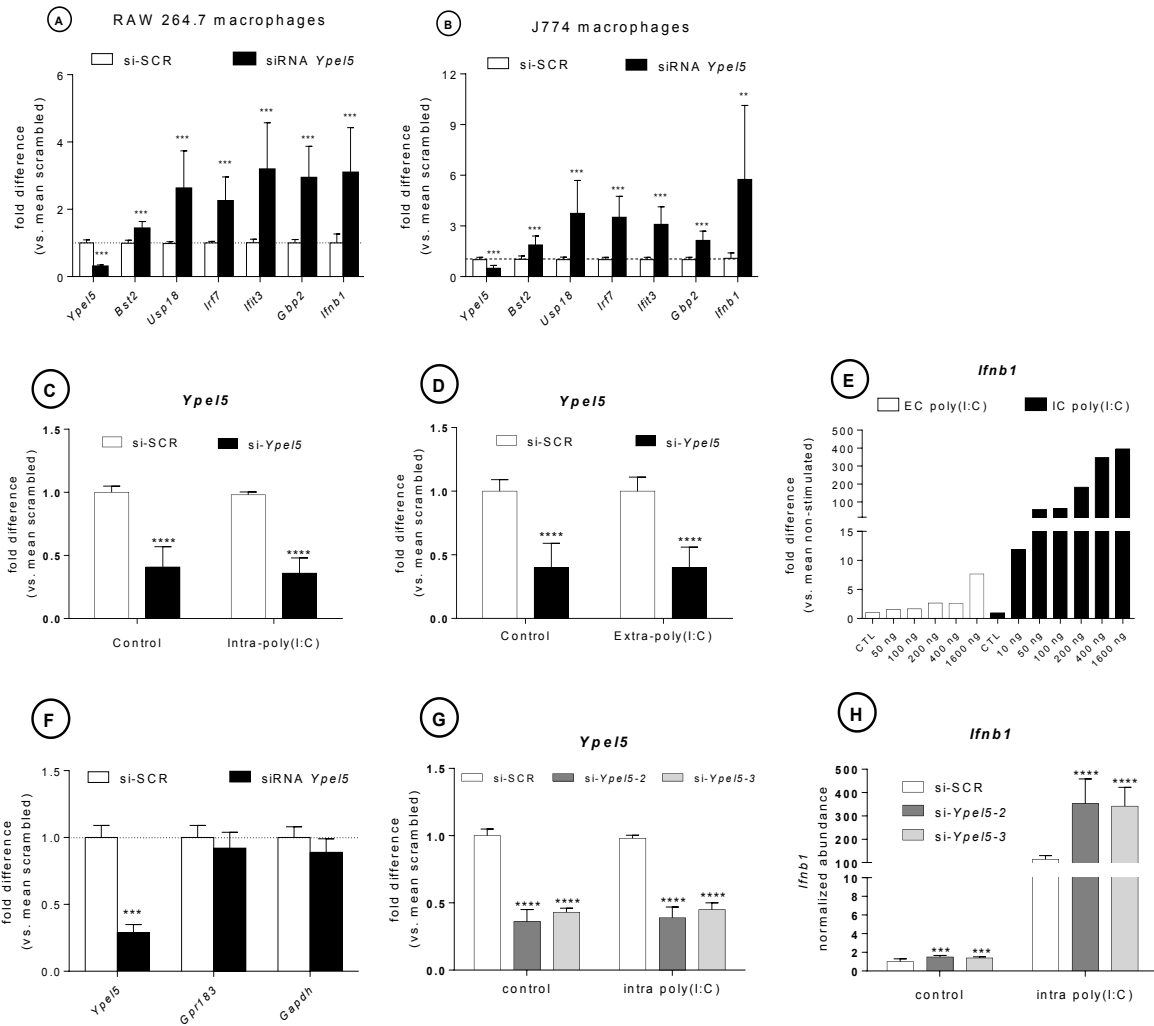


Figure S3 (related to Figure 2.2): *Ypel5* in mouse macrophage cell line. Figure S3a,b: Relative abundance of mRNA transcripts of *Ypel5*, *Ifnb1* and 5 ISGs (*Bst2*, *Usp18*, *Irf7*, *Ifit3*, *Gbp2*) in either RAW264.7 cells (Figure S3a) or J774 cells (Figure S3b), treated for 24 hours with either scrambled or *Ypel5* siRNA. All values represent mean \pm SD of results obtained in duplicate wells from 3-6 individual experiments (n = 6-12). All differences are significant by Student's t-test (** $P < 0.01$; *** $P < 0.001$). Figure S3c,d: RAW264.7 cells were transfected with either scrambled or *Ypel5*-specific siRNA, then either transfected with 25ng/ml poly (I:C) (Figure S3c) or treated with external poly (I:C) (2000 ng/ml) (Figure S3d). In both cases, *Ypel5* siRNA downregulated *Ypel5* expression to the same extent in all groups, and neither poly (I:C) treatment affected endogenous *Ypel5* expression. Values are mean \pm SD, replicate wells from three separate experiments (n = 12). Asterisks indicate significant differences between

cells transfected with either scrambled or *YPEL5* siRNA, as determined by Tukey's pot-hoc tests (**** $P < 0.0001$). Figure S3e: dose-response effects of either extracellular or intracellular (transfected) poly (I:C) on *Ifnb1*. Induction of *Ifnb1* by transfection of poly (I:C) was observed after using doses as small as 10 ng; in contrast, extracellular poly(I:C) required doses at least > 200 ng/ml prior to observing any stimulatory activity on *Ifnb1*. Figure S3f: Transfection of RAW264.7 cells with si-*Ypel5*-1 (the first siRNA tested) significantly decreases *Ypel5* expression, but has no significant effect on *Gpr183* and *Gapdh* expression; the values represent mean \pm SD of results, replicate wells from 3 individual experiments (n =12). Differences were assessed by Student's t-tests (*** $P < 0.001$). Figure S3g: Effects of two other *Ypel5*-specific siRNAs (si-*Ypel5*-2 and si-*Ypel5*-3) on *Ypel5* expression levels. Both in unstimulated cells or in cells stimulated by transfection with 25 ng poly(I:C), *Ypel5* expression was about 30-40% of levels observed in cells transfected with scrambled siRNA; by itself, transfection with 25 ng poly(I:C) had no effect on endogenous *Ypel5* expression. Figure S3h: Effects of si-*Ypel5*-2 and si-*Ypel5*-3 on *Ifnb1* expression levels in either unstimulated cells or in cells stimulated by transfection with 25 ng poly(I:C). After poly (I:C) transfection, *Ifnb1* expression was higher in cells transfected with either si-*Ypel5*-2 or si-*Ypel5*-3 than in cells transfected with scrambled siRNA. Figure S3g,h: Differences were assessed by one-way ANOVA followed by Tukey's multiple comparison post-hoc tests. *** $P < 0.001$; **** $P < 0.0001$.

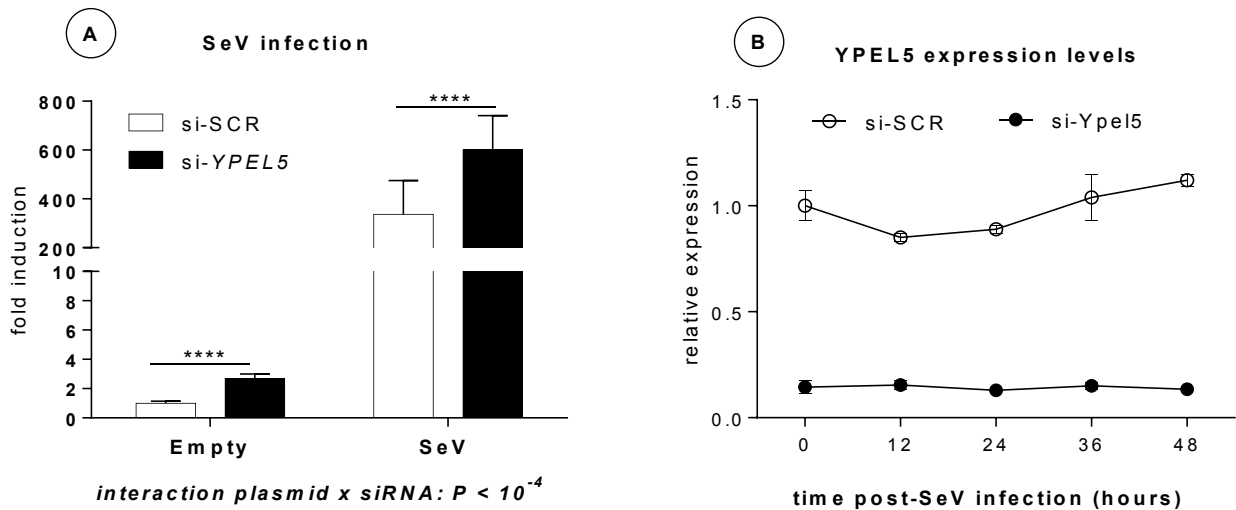


Figure S4 (related to Figure 2.4): Effects of SeV infection on HEK293T cells transfected with either scrambled or *YPEL5*-specific siRNA. Figure S4a: Twenty four hours after transfection with either scrambled (si-SCR) or *YPEL5*-specific (si-*YPEL5*) siRNA, HEK293T were transfected with the pIFNB1-LUC reporter plasmid, then exposed to either vehicle or SeV. After 16h of incubation, the cells were used for measuring luciferase activity. All values represent the means (\pm SD) of results obtained in replicate wells from 3 individual experiments ($n = 18$). SeV infection increased IFNB1-LUC activity in both si-SCR and si-*YPEL5* transfected cells; two-way ANOVA tests showed that interaction of the si-RNA treatments with the effects of SeV was significant ($P < 0.001$). Figure S4b: Quantification of the abundance of *YPEL5* mRNA in cells transfected with either si-SCR or si-*YPEL5*. At all times after SeV exposure, the levels of expression of *YPEL5* in si-*YPEL5*-transfected cells were about 20% of that in their si-SCR-transfected counterparts. In si-SCR-transfected cells, the abundance of *YPEL5* mRNA did not vary significantly after SeV infection.

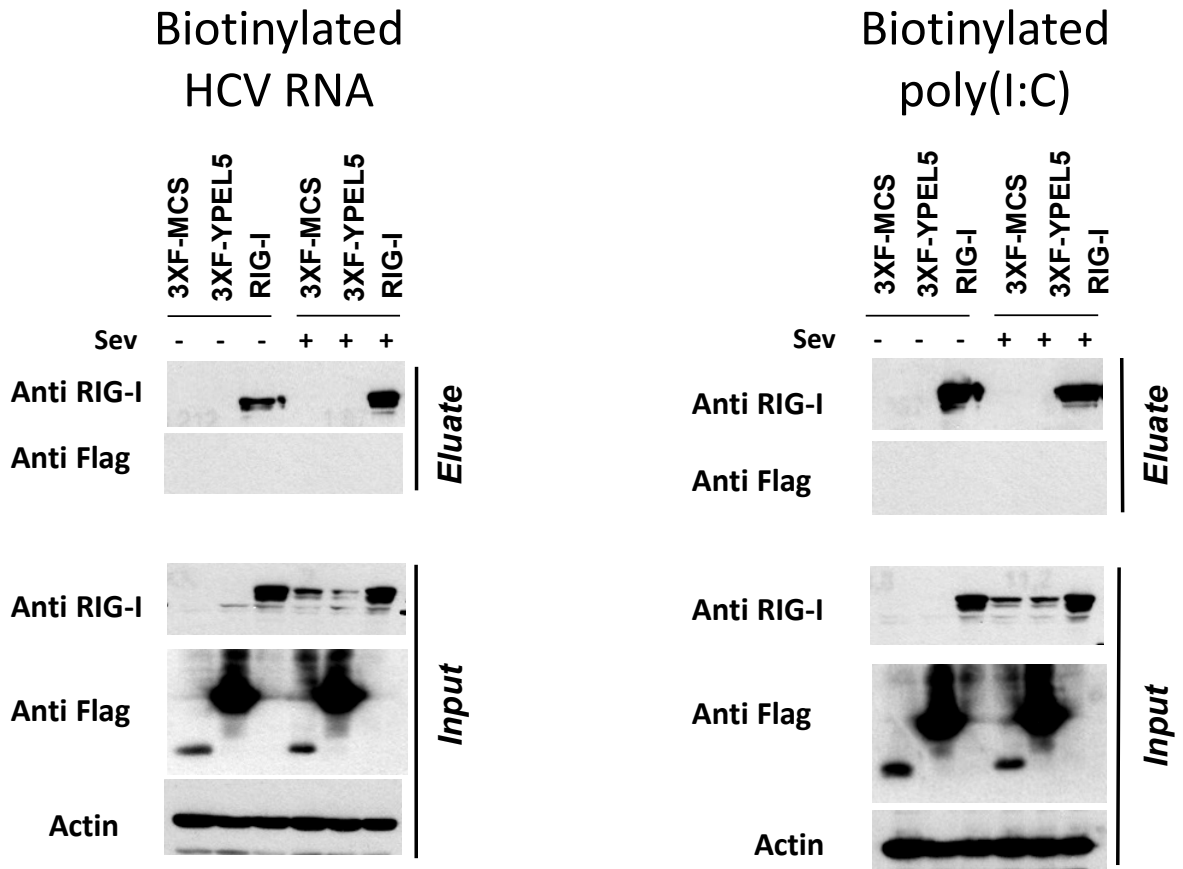


Figure S5 (related to Figure 2.4): *In vitro* interaction of biotinylated HCV RNA and poly (I:C) with cell lysates of transfected cells. HEK293T cells were transfected with either 3xF-YPEL5 or RIG-I expression plasmids or the control 3xF-MCS empty plasmid for 32 hours, and then treated with either SeV or vehicle for 16 hours. The presence of overexpressed proteins in the lysates was verified by Western blot analysis with either anti-Flag or anti-RIG-I antibodies (input). RNA pull-down with streptavidin beads revealed that the biotinylated RNA bound to proteins in lysates from cells transfected with RIG-I expression plasmids, but not in that from cells transfected with plasmids overexpressing 3xF-YPEL5 (eluate results).

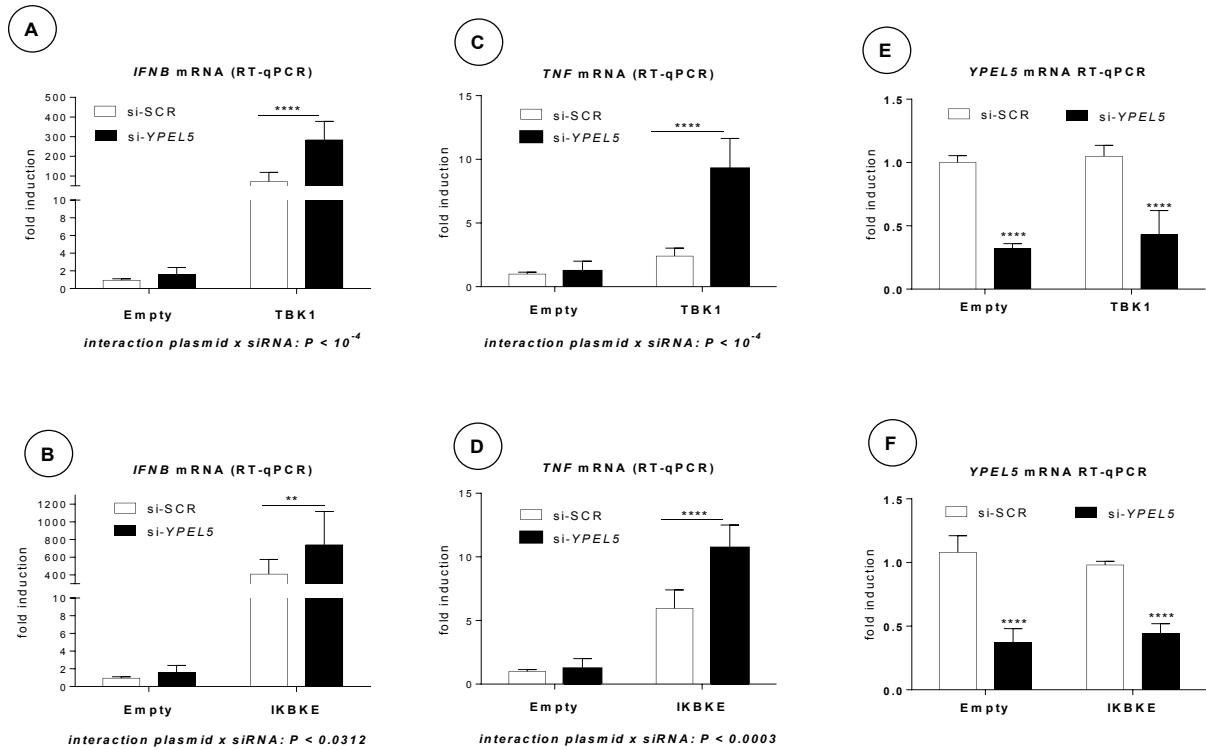


Figure S6 (related to Figure 2.5). *YPEL5* siRNA potentiates the effects of *TBK1/IKBKE* overexpression on endogenous *IFNB1* and *TNF* expression in HEK293T cells. Cells were transfected with siRNA for 24 hours and further stimulated for an additional 24 hours with either pcDNA3.1 (empty plasmid) or expression plasmids for either *TBK1* (A,C) or *IKBKE* (B,D). The abundance of endogenous *IFNB1* (A, B) or *TNF* mRNA (C, D) was assessed by RT-qPCR. Interactions between the effects of the stimuli with that of the siRNAs were tested by two-way ANOVAs; significance values of the interactions between si-*YPEL5* and stimuli are as indicated. All values represent the means (\pm SD) of results obtained in replicate wells from several individual experiments ($n = 6-10$). Asterisks indicate significant differences between cells transfected with either scrambled or *YPEL5* siRNA, as determined by Tukey's post-hoc tests (**** $P < 0.0001$). The abundance of endogenous *YPEL5* mRNA was also measured in all conditions, with all values being normalized to that obtained in cells transfected with scrambled siRNA and empty plasmid (E, F). Overexpression of either *TBK1* or *IKBKE* had no effect by themselves on endogenous *YPEL5* expression; in all groups, the levels of *YPEL5* expression were about 30-40% to those observed in cells transfected with scrambled siRNA.

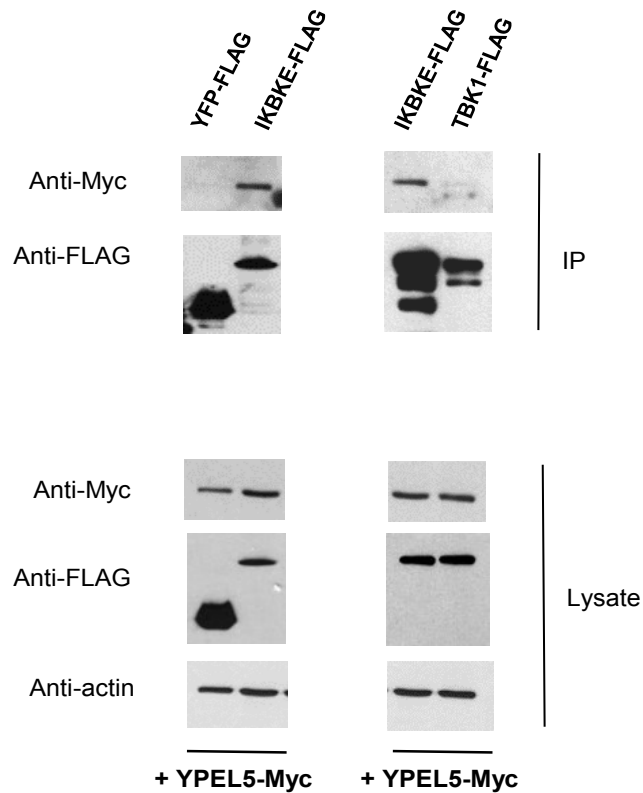


Figure S7: Reverse co-IP (related to Figure 2.7). HEK293T cells were transfected with 5 μ g of YPEL5-Myc along with either FLAG-tagged YFP (in the first experiment, left part) or FLAG-tagged TBK1 (in the second experiment, right part). The tags of these constructs were the inverse of what had been used for the experiments whose results presented in Figure 2.

7. Immunoprecipitation was performed using anti-FLAG antibodies, and either cell extracts (lysate) or eluates (IP) were analyzed by Western blotting. Western blots revealed that when the FLAG-tagged transfected protein was IKBKE, the immunoprecipitated material also contained Myc immunoreactivity (from YPEL5-Myc). No clear Myc signal was obtained when the FLAG-tagged transfected protein were either YFP or TBK1.

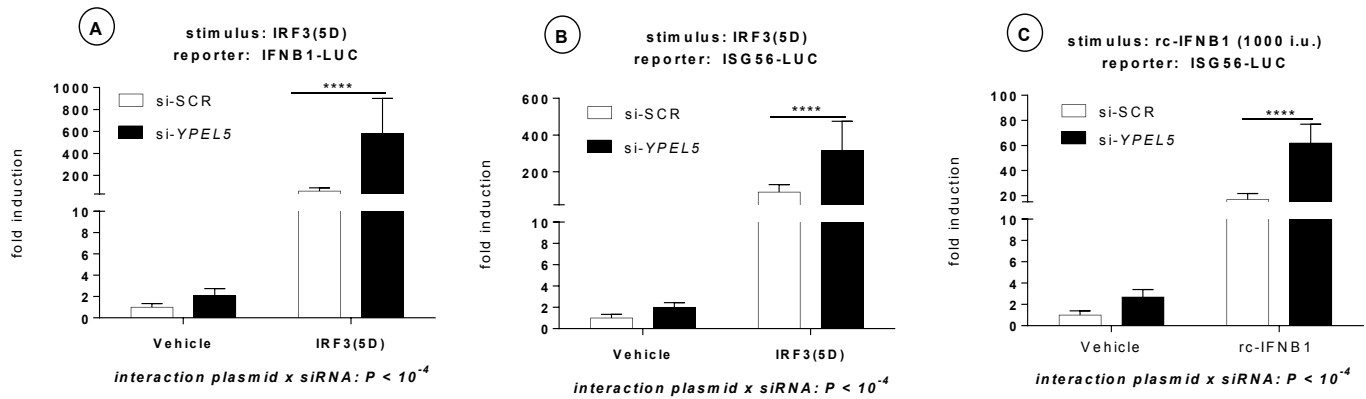


Figure S8 (related to Figure 2.5): Interactions of *YPEL5* siRNA with the effects of IRF3 (5D) and rcIFNβ1. Twenty four hours after transfection with siRNA, HEK293T cells were transfected with either pIFNβ1-LUC (A) or pISG56-LUC reporter plasmid (B, C). Then, cells were either transfected with IRF3 (5D) (A, B) or exposed to 1000 i.u. of rcIFNβ1 (C). Luciferase activity was measured 24 h after stimulation. All values represent the means (\pm SD) of results obtained in replicate wells from 3 individual experiments ($n = 12-18$). Interactions between the effects of the stimuli with that of the siRNAs were tested by two-way ANOVAs; significance values of the interactions between si-*YPEL5* and stimuli are as indicated. Asterisks indicate significant differences between cells transfected with either scrambled or *YPEL5* siRNA, as determined by Tukey's pot-hoc tests (**** $P < 0.0001$).

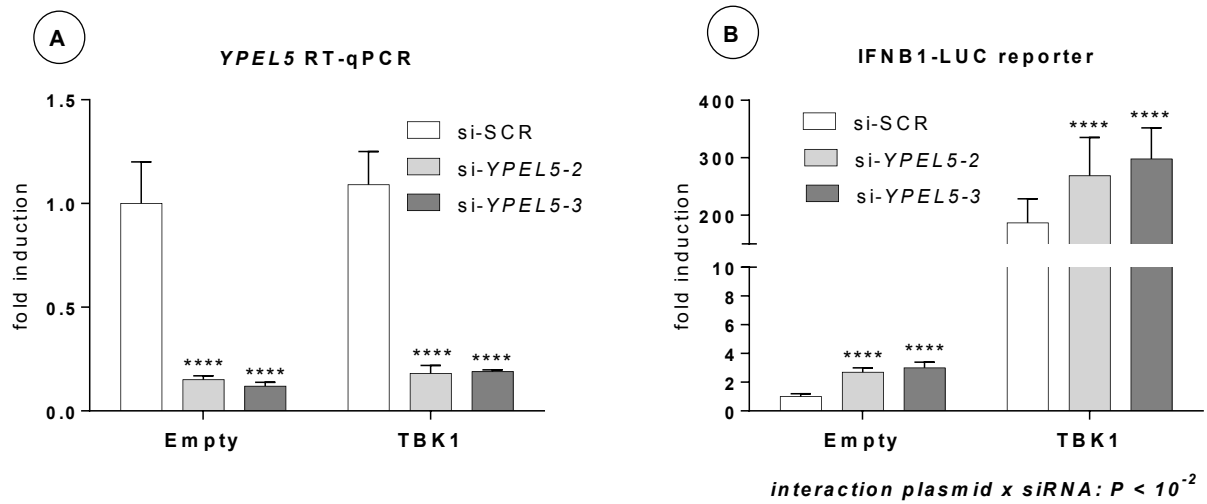


Figure S9 (related to Figure 2.5): Two additional *YPEL5*-specific siRNAs have effects that are similar to that of si-*YPEL5-1*. (A) In HEK293T cells stimulated with either pcDNA3.1 (empty plasmid) or expression plasmids for TBK1, endogenous *YPEL5* expression levels were significantly lower than those observed in cells transfected with scrambled siRNA; (B) both *YPEL5-2* and *YPEL5-3* siRNAs amplified significantly the effects of *TBK1* overexpression on IFNB1-dependent luciferase activity. All values represent the means (\pm SD) of results obtained in replicate wells from three individual experiments ($n = 12$). Asterisks indicate significant differences between cells transfected with either scrambled or *YPEL5* siRNA, as determined by Tukey's pot-hoc tests (**** $P < 0.0001$).

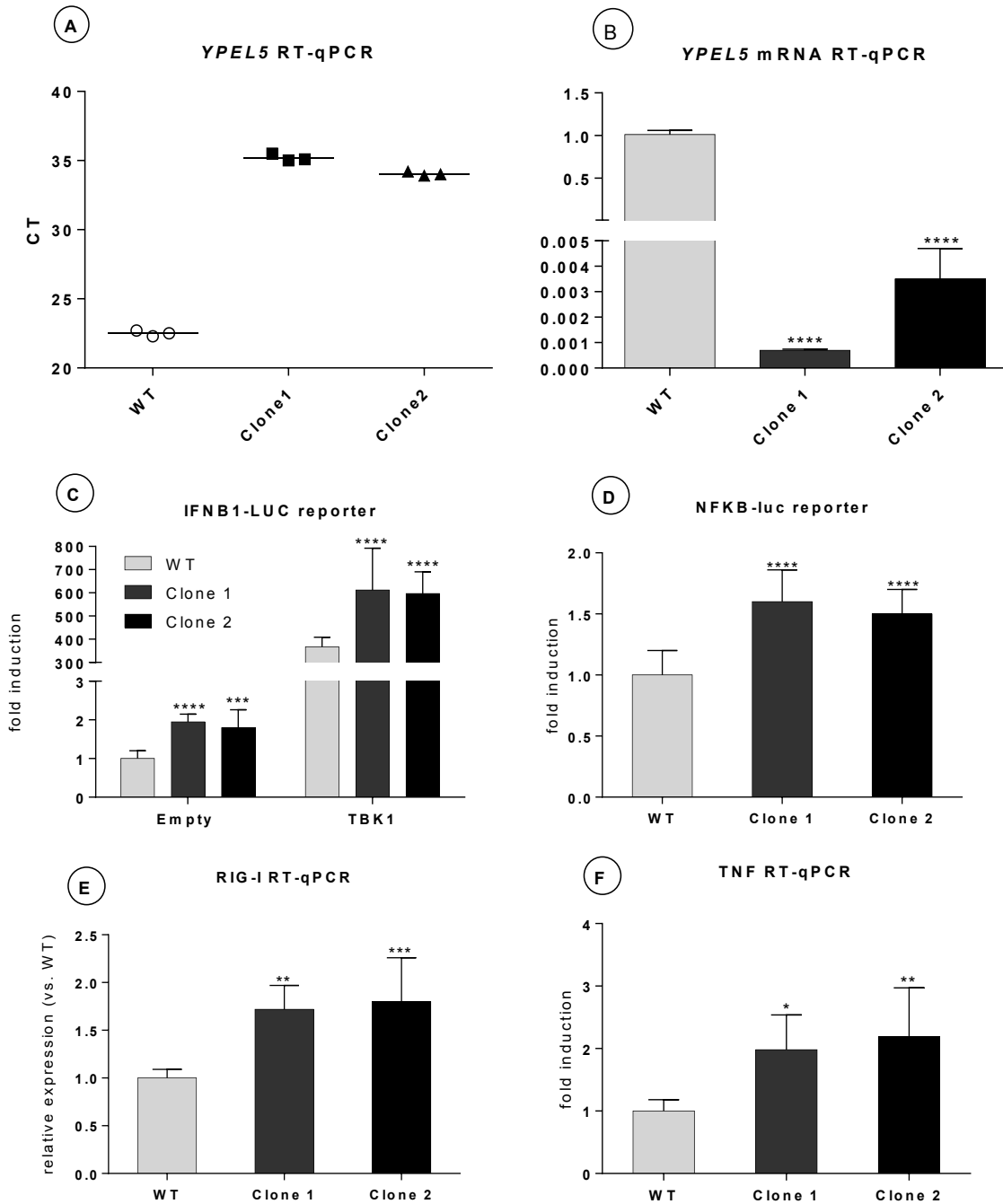


Figure S10 (related to Figure 2.5): Ablation of the *YPEL5* gene by RNA-guided nuclease in HEK293T cells. Two individual clonal cell lines carrying a deletion within exon 4 of *YPEL5* were compared to wild-type (WT) HEK293T cells. Ablation caused a large increase in the qRT-PCR Ct values (A), so that *YPEL5* mRNA abundance was barely measurable in the 2 clonal cell

lines (B). Both clonal cell lines displayed increased IFN β 1-dependent (C) and NF- κ B-dependent luciferase (D) activity, as well as increased expression of *RIG-I* (chosen as a canonical ISG) (E) and increased *TNF* expression (a target gene of NF- κ B) (F). Moreover, TBK1 overexpression increased IFN β 1-dependent luciferase activity, but the effect was significantly higher in the two clonal cell lines than in the WT cells (C). Differences vs WT cells were tested by one-way ANOVA followed by Tukey's multiple comparison post-hoc tests; ****P < 0.0001; ***P < 0.001).

CHAPITRE 3 : LE ROLE DE YPEL5 DANS LA REGULATION DU CYCLE CELLULAIRE ET/OU DE LA SENESCENCE

3.1. Introduction

Les génomes de vertébrés contiennent 5 gènes *YPEL* (*YPEL1* à *YPEL5*) appartenant à une même famille caractérisée par un degré élevé d'homologie avec le gène *Yippee* de *Drosophila melanogaster*. Des orthologues de ces gènes se retrouvent chez tous les organismes eukaryotes, incluant les protozoaires, les végétaux et les champignons³²⁴. *YPEL5* est la protéine qui montre le plus haut niveau d'homologie par rapport à la protéine *Yippee* de la drosophile (70.8% d'identité), et le niveau de conservation de sa séquence parmi les vertébrés est de 100%^{310,324}. Les protéines *YPEL1-4* montrent un niveau d'homologie plus faible avec la séquence de *Yippee* (45% d'identité), mais un fort pourcentage d'identité les unes par rapport aux autres³²⁴.

Bien que le haut degré de conservation suggère des fonctions importantes pour ces protéines, il n'existe que très peu d'informations fonctionnelles à leur égard. Dans les cellules COS-7, toutes les protéines *YPEL* sont localisées préférentiellement au niveau du centrosome lors de l'interphase, et sont associées avec le fuseau mitotique pendant la mitose³²⁴. Lorsque le niveau d'expression de *YPEL5* est diminué dans des cellules COS-7, le taux de croissance de ces dernières est ralenti en raison d'une augmentation du nombre de cellules arrêtés au niveau de la phase G1³²⁵. Ceci suggère que les protéines *YPEL* possèdent une fonction étroitement liée au cycle de division cellulaire³²⁴. Par ailleurs, certaines études ont suggéré des relations étroites du gène *YPEL3* avec les processus d'oncogenèse et/ou de sénescence cellulaire, deux processus étroitement liés au cycle cellulaire, tel que détaillé ci-dessous.

L'activation de la protéine p21 [encodée par le gène *CDKN1A* « *cyclin-dependent kinase inhibitor1* »] représente l'une des voies qui induit la sénescence et l'arrêt de la croissance

cellulaire^{348,349}. Le gène *CDKN1A* est lui-même régulé par la p53³⁵⁰⁻³⁵². En principe, la p21 activée empêche les complexes cycline-CDK2 de phosphoryler la famille de protéines RB³⁵³⁻³⁵⁵. Dans son état non phosphorylé, la protéine RB est active et peut fixer les facteurs E2F, ce qui empêche la progression du cycle cellulaire et provoque un blocage de la transition G1/S.³⁵⁶⁻³⁶⁰ La p21 peut également arrêter la progression du cycle cellulaire en inhibant les complexes cycline A/B-CDK1 et cyclinD-CDK4³⁶¹. Plusieurs études indiquent que l'activité transcriptionnelle de *YPEL3* est augmentée par la surexpression de p53³⁶² ou des formes apparentées p73 et p63³⁶³. La surexpression de p53 augmente aussi l'expression de *YPEL3* dans des cellules du cancer du sein (MCF7)^{364,365}. L'induction physiologique de *YPEL3* entraîne elle-même une diminution importante de la viabilité cellulaire associée à une augmentation de la sénescence cellulaire^{362,366}. En plus de son rôle dans le processus de sénescence, p53 est connu comme un « *tumor suppressor gene* » dont la fonction est inactivée dans la majorité des cancers humains, soit suite à des mutations inactivatrices, soit en raison de la surexpression de deux régulateurs négatifs de p53, *HDMX* et *HDM2*^{367,368}.

L'implication possible de *YPEL3* dans le cancer est également supportée par des données montrant que : son expression est diminuée dans un large éventail de tumeurs humaines solides^{366,369,370}, et 2) les estrogènes diminuent l'expression de *YPEL3* dans des cellules de carcinome mammaire MCF-7, en parallèle avec une augmentation du taux de croissance de ces cellules³⁶⁹. Prises collectivement, ces données suggèrent que certaines fonctions de *YPEL3* pourraient être compatibles avec celles de « *tumor suppressor genes* ». Le fait que ces propriétés ont été démontrées en partie dans des cellules de cancer mammaire pourrait être également avoir un intérêt particulier car : 1) les kinases IKBKE et TBK1 sont fréquemment surexprimées dans divers cancers mammaires^{337,338} ; 2) l'inhibition de ces kinases diminue la viabilité de cellules humaines de cancer mammaire³⁷¹; et 3) nous avons montré qu'une des nouvelles fonctions de *YPEL5* serait d'inhiber l'activation des kinases IKBKE et TBK1³⁷². Finalement, les liens possibles des gènes *YPEL* avec p53 et les processus de sénescence seraient également compatibles avec la localisation centriolaire des protéines *YPEL*, car il est connu que la perturbation de l'expression de nombreuses protéines du centriole active des mécanismes

dépendant de p53 qui causent les cellules à adopter un état proche de la sénescence cellulaire avec un arrêt du cycle cellulaire en G1³⁷³⁻³⁷⁶.

Nous avons initié des expériences additionnelles pour répondre à plusieurs types de questions. Pour analyser en plus de détail de quelle manière les gènes *YPEL* pourraient être associés aux processus de sénescence et/ou de croissance cellulaire, nous avons comparé les effets du « *knockdown* » des gènes *YPEL5* et *YPEL3* dans des cellules humaines HEK293T et MCF7. En particulier, nous avons testé si ces gènes interféraient avec les effets de H₂O₂, *i.e.* un stimulus connu en général pour induire le processus de sénescence³⁷⁷ et pour induire le stress oxydatif dans les cellules MCF7³⁷⁸. En plus d'analyser les effets sur l'expression de *p21*, nous avons testé les effets de diverses conditions sur l'expression de *TNFA* car : 1) les cellules sénescents présentent un phénotype appelé '*senescent associated secretory phenotype*' (SASP), qui s'accompagne de la sécrétion de nombreuses cytokines (incluant TNF α); et 2) nous avons montré précédemment que *YPEL5*, via son effet inhibiteur sur les kinases TBK1/IKBKE, inhibe l'expression de *TNF α* , une cible bien connue de NF-kB³⁷².

3.2 Procédures expérimentales

Lignées cellulaires et culture

Nous avons cultivé des lignées cellulaires MCF7 (du cancer du sein humain) dans du milieu Eagle modifié par Dulbecco (DMEM, Wisent), supplémenté avec 10% FBS, 100 unités / ml de pénicilline, 100 ug / ml de streptomycine et 2 mM de glutamine (tous de Wisent). Les cultures ont été cultivées et maintenues à 37 ° C dans une atmosphère humidifiée contenant 5% de CO₂.

Pour les transfections siARN, on a utilisé le contrôle négatif MISSION® siARN SIC001 (Sigma-Aldrich, St-Louis, MO). Pour le knockdown de *YPEL5* humain, nous avons utilisé le SASI_Hs02_00329464. Pour le knockdown de *YPEL3* humain, nous avons utilisé le siRNA SASI_Mm01_00145086. La transfection des siARN a été effectuée en utilisant la Lipofectamine RNAi-Max (Life Technologies, Canada), selon les instructions du fabricant.

Pour le traitement aigu, les cellules MCF-7 ont été exposées à 250 μM H_2O_2 pendant 24 heures et ont ensuite été utilisées pour l'analyse. Des cultures parallèles ont été cultivées et maintenues en tant que témoins de passage identiques.

Analyses de l'expression génétique

L'extraction d'ARNm et la RT-qPCR a été effectuée selon la procédure décrite dans la section « matériels et méthodes » du chapitre 2.

3.3 Résultats

3.3.1 *YPEL3* et *YPEL5* sont les deux gènes *YPEL* les plus abondamment exprimés dans les cellules HEK293T et MCF7.

Dans un premier temps, nous avons voulu vérifier le niveau d'expression des 5 gènes *YPEL* dans les cellules humaines HEK293T et les cellules du cancer du sein (MCF7). En effet, il a été rapporté précédemment que *YPEL3* et *YPEL5* sont exprimés de manière ubiquitaire dans des tissus provenant d'humains et de souris, alors que les autres membres de la famille ont des profils d'expression plus restrictifs³²⁴. De façon compatible avec ces études précédentes, nous avons constaté que *YPEL3* et *YPEL5* étaient effectivement les 2 gènes *YPEL* dont les niveaux d'expression sont les plus abondants dans les cellules humaines HEK293T (figure 3.1A) et les cellules du cancer du sein (MCF7) (figure 3.1B).

3.3.2 H_2O_2 induit l'expression de *YPEL5*, *p21*, et *TNF α* , mais pas celle de *YPEL3* dans les cellules MCF7.

L'exposition à H_2O_2 (250 μM pour 24h) cause une augmentation significative ($p < 0,0001$) de l'expression de *YPEL5* (1,9 fois), de *p21* (4,1 fois) et de *TNF α* (1.9 fois) (Figure 3.2A). Par contre, ce même traitement n'affectait pas l'expression de *YPEL3*.

3.3.3 Le KD de *YPEL5* augmente l'expression de *YPEL3*, *p21*, et *TNF α* dans les cellules MCF7.

Nous avons testé les effets du KD de *YPEL5* sur l'expression de *YPEL3*, *p21*, et *TNF α* , soit dans des conditions basales, soit après traitement avec H₂O₂ dans les MCF-7. Le KD fut induit par transfection avec un siARN spécifique à *YPEL5*, ce qui diminue l'expression de *YPEL5* à moins de 10% du niveau observé dans des cellules transfectées avec un siARN contrôle (scrambled siRNA) (Figure 3.2B). Dans les conditions basales (sans traitement de H₂O₂), le KD de *YPEL5* augmente l'expression du *TNF α* endogène (2,35 fois), de la *p21* (3,2 fois) et de *YPEL3* (2 fois) de façon significative (p <0,0001) (Figure 3.2B). Par contre, le knockdown de *YPEL3* dans les MCF7 n'affectait pas l'expression de *YPEL5*, *p21*, et de *TNF α* (données non montrées).

Nous avons montré précédemment que le KD de *YPEL5* augmentait les effets de la surexpression de TBK1 sur son autophosphorylation, sur la phosphorylation croisée de IKBKE, et l'induction de IFN β et *TNFA*³⁷². Cependant, comme *YPEL5* inhibe *YPEL3* (mais non l'inverse), il devient important de déterminer quels seront au final les effets du KD de ces 2 gènes. En comparant les effets de si*YPEL3* et si*YPEL5* (transfectés soit seuls, soit de façon conjointe), nous avons constaté que tous ces traitements affectaient de manière similaire les effets de la surexpression de TBK1 sur l'induction d'IFN β et *TNFA* et la phosphorylation de TBK1 et IKBKE (Figure 3.3C).

3.3.4 Le KD de *YPEL5* amplifie l'induction de *p21* par H₂O₂

24 heures après la transfection de si*YPEL5*, les MCF7 furent traitées pour 24 heures supplémentaires avec une dose de 250 μ M de H₂O₂. Tel que montré précédemment, le KD de *YPEL5* augmente l'expression de *p21* de 2 fois. L'exposition à H₂O₂ provoque une induction d'expression de *p21* de 3,8 fois, alors que le KD de *YPEL5* amplifie l'effet de H₂O₂ sur l'expression *p21* en augmentant son expression de 6 fois (Figure 3.4A). L'effet additif (ou synergique) de *YPEL5* et H₂O₂ suggère que les deux agissent sur l'expression de *p21* à travers la même voie de signalisation. Par contre, le KD de *YPEL5* n'a pas amplifié l'expression de *TNF α* induite par H₂O₂ (Figure 3.4B).

3.3.5 Le KD de *Ypel5* augmente l'expression de gènes associés à la sénescence et/ou l'apoptose dans les cellules RAW264.7.

Nous avons étudié précédemment l'effet du KD de *Ypel5* sur le profil transcriptomique de cellules macrophagiques de souris RAW 264.7. En plus des gènes associés à la réponse immune, une des classes de gènes affectée de la façon la plus significative incluait ceux associés à la régulation de la mort cellulaire programmée³⁷². Le taux d'induction et la significativité de l'effet sont présentés de façon plus détaillée dans le Tableau 3.1. En particulier, on peut observer que (de façon similaire aux cellules MCF7), si *Ypel5* augmente l'expression de *Ypel3*. Les autres gènes répertoriés dans le tableau ont tous été décrits comme ayant des implications possibles dans les processus d'apoptose et/ou de sénescence cellulaire.

3.4 Discussion

Au sein de la famille YPEL, les gènes *YPEL5* et *YPEL3* semblent être ceux qui sont exprimés de la façon la plus prédominante et ubiquitaire³²⁴. Cependant, nous ne savons pas à ce jour si ces 2 gènes se régulent mutuellement, et si les fonctions des protéines encodées par ces gènes sont similaires ou différentes. De plus, les données publiées à ce jour suggèrent que tant *YPEL5* et *YPEL3* peuvent être associés à la division cellulaire et/ou à la sénescence sans qu'il soit possible de départager les contributions relatives de ces deux gènes. Bien que nos expériences ne soient encore que très préliminaires, nos résultats apportent un début de réponse à ces questions.

Tel que déjà mentionné, les résultats publiés à ce jour indiquent que les protéines YPEL ont une localisation principalement centrosomale. Le centrosome est une composante cellulaire importante qui contrôle l'intégrité génomique, en partie grâce au contrôle de la progression du cycle cellulaire (de la phase G1 à la phase S et de la phase G2 à la phase M)³⁷⁹. La dérégulation des protéines associées au centrosome est liée à une mauvaise régulation du cycle cellulaire ainsi qu'au cancer³⁸⁰. Il a été rapporté que la variation héréditaire des gènes impliqués dans la structure et la fonction du centrosome peut contribuer au développement du cancer du sein³⁸¹.

Concernant les gènes YPEL, plusieurs membres de cette famille ont été associés aux phénomènes de régulation de la progression du cycle cellulaire et/ou du cancer. Il a été suggéré que YPEL1 puisse être impliqué dans la transition épithéliale-mésenchymateuse '*mesenchymal to epithelial-like transition*' au cours du développement tissulaire ³⁸², l'un des événements caractéristiques de la transformation maligne dans les cellules cancéreuses pancréatiques. D'autres études ont montrées la dérégulation de l'expression de YPEL1 dans les cellules péritonéales invasives du cancer pancréatique ^{383,384}. La réduction d'expression de *YPEL1* corrèle avec une meilleure survie des patients atteints de cancer du pancréas. De plus, il a été montré que la région 17q23 (précédemment identifiée comme région fréquemment amplifiée dans les tumeurs mammaires ³⁸⁵) contient un polymorphisme (rs16943468) dans le gène YPEL2 qui a été significativement associé au cancer du sein post-ménopausique ³⁸⁶. Une étude récente a montré également que la surexpression de *YPEL4* provoquait une augmentation significative de la production d'aldostérone dans les milieux de culture cellulaire associés à une prolifération accrue de cellules adrénocorticales humaines HAC1 ³⁸⁷.

En plus des associations possibles avec le cancer, plusieurs études suggèrent qu'il existe des interactions bi-directionnelles entre YPEL3 et p53. En plus du cancer, p53 est une molécule centrale dans le processus de sénescence cellulaire ³⁶². Par ailleurs, plusieurs études récentes ont montré que l'induction de la sénescence pouvait prévenir l'initiation et ralentir la progression des tumeurs chez les souris ^{388,389}. En plus de freiner la tumorigénèse, la sénescence contribue à la cytotoxicité de certains agents anticancéreux comme les traitement de chimiothérapie ou de radiothérapie ^{390,391}. Bien que la sénescence puisse être un phénomène spontané limitant la durée de vie répliquative de cellules en culture en raison de la perte progressive de l'ADN ^{357,392}, elle peut également être induite par divers facteurs de stress cellulaires et environnementaux, incluant les rayonnements ionisants ou les agents induisant un stress oxydatif ^{393,394}. L'administration de H₂O₂ exogène tant *in vitro* que *in vivo* représente un stimulus qui induit un stress oxydatif ^{395,396}, une condition qui cause aussi des dommages à l'ADN ³⁹⁷. Dépendamment de la dose et de temps d'exposition utilisé, H₂O₂ peut provoquer un arrêt transitoire ou permanent de la croissance cellulaire ^{398,399}. Il a été montré précédemment qu'une exposition courte de cellules MCF7 à des doses élevées de H₂O₂ ralentissait le taux de croissance de ces

cellules et diminuait l'expression de « *pro-survival genes* » du cycle cellulaire, incluant *cyclinD1*, *survivin*, et *Bcl2* ⁴⁰⁰.

Nos résultats préliminaires obtenus dans des cellules MCF7 indiquent que : 1) si *YPEL5* augmente l'expression des gènes *P21* et *TNFA*; et 2) H_2O_2 , en plus d'augmenter l'expression des deux même gènes *P21* et *TNFA*, augmente l'expression du gène *YPEL5* lui-même. Ces résultats suggèrent des interactions réciproques entre les gènes *YPEL5* et *P21*, ce dernier jouant un rôle central dans le processus de sénescence cellulaire. Il est à noter que contrairement à nos propres résultats, les études précédentes concernaient principalement les interactions du gène *YPEL3* avec les phénomènes de sénescence ³⁶²⁻³⁶⁵. Bien qu'il n'était pas connu jusqu'à ce jour si *YPEL3* et *YPEL5* avaient des effets similaires, certains de nos résultats indiquent que tel serait le cas concernant leur effets sur l'expression de *IFN β* et *TNFA*, ainsi que sur certains effets de la surexpression de *TBK1*. Cependant, nous avons également observé certaines différences : 1) H_2O_2 augmente l'expression de *YPEL5*, mais pas celle de *YPEL3* ; 2) si *YPEL5* amplifie l'effet de H_2O_2 sur l'induction de *p21*, alors que si *YPEL3* ne montre pas d'effet similaire ; et 3) si *YPEL5* augmente l'expression de *YPEL3*, mais si *YPEL3* n'affecte pas l'expression de *YPEL5*. Des études supplémentaires seront donc nécessaires pour comprendre les interactions de ces 2 gènes avec le processus de sénescence cellulaire et la manière dont ils peuvent se réguler réciproquement.

Des données concernant les effets des gènes *YPEL* dans les cellules MCF7 pourraient être pertinents pour plusieurs phénomènes concernant le cancer du sein. Ce dernier est le cancer le plus couramment diagnostiqué chez les femmes dans la grande majorité des pays du monde entier, représentant un quart de tous les cancers diagnostiqués chez les femmes ⁴⁰¹. C'est aussi la principale cause de décès liés au cancer chez les femmes ⁴⁰², en raison des métastases précoces, et de la résistance aux agents chimiothérapeutiques existants. Bien que le temps de survie sans récurrence et la survie globale des patients atteints d'un cancer du sein ait été amélioré grâce au perfectionnement des traitements, ce cancer reste encore un problème de santé publique important. Par ailleurs, le stress oxydatif peut affecter le comportement des cellules cancéreuses de diverses façons. Dans le cancer du sein, plusieurs causes peuvent augmenter les concentrations des radicaux libres, incluant les œstrogènes ^{403,404}, les xœstrogènes (qui miment les œstrogènes) ^{405,406}, le dysfonctionnement mitochondrial ⁴⁰⁷, l'activité des

peroxysomes et l'activité accrue de diverses oxydases ^{408,409}. Alors que des niveaux élevés de radicaux libres peuvent endommager diverses composants cellulaires et ainsi induire le phénomène de sénescence ^{389,410}, de faibles concentrations pourraient au contraire stimuler la prolifération cellulaire ⁴¹¹⁻⁴¹³. Par ailleurs, de nombreuses stratégies chimiothérapeutiques sont conçues pour provoquer l'arrêt du cycle cellulaire par le biais de l'induction d'un stress oxydatif ^{414,415}. Par exemple, la doxorubicine (un traitement de chimiothérapie utilisé contre le cancer du sein) arrête la croissance cellulaire par un mécanisme médiée par les ROS dans diverses cellules tumorales ⁴¹⁶⁻⁴¹⁸. Des études supplémentaires permettraient ainsi de déterminer de quelle manière les gènes YPEL peuvent affecter la sénescence cellulaire et/ou les effets du stress oxydatif dans des cellules de cancer du sein.

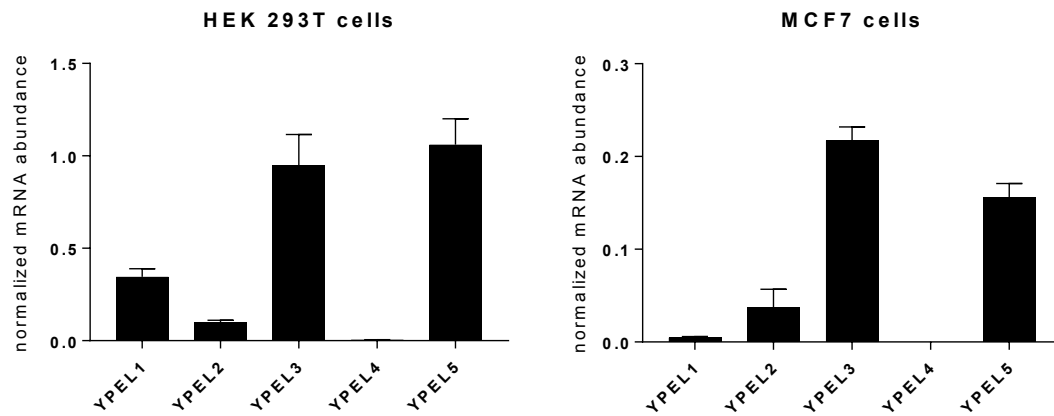


Figure 3.1: L'expression relative de *YPEL1-5*. L'abondance normalisée (telle que déterminée par RT-qPCR, avec normalisation par rapport au Rps16) des ARNm des gènes YPEL1-5 dans les cellules HEK 293T (gauche) ou MCF-7 (droite). Dans les deux types de cellules, YPEL3 et YPEL5 sont les gènes les plus abondamment exprimés.

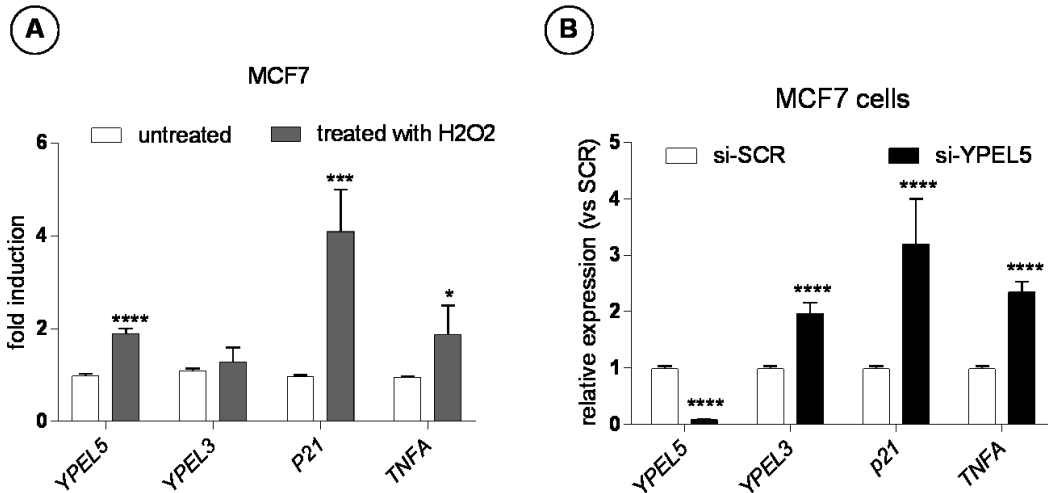


Figure 3.2: (A) H₂O₂ induit YPEL5, YPEL3, p21, et TNFA dans des cellules MCF7. Les cellules MCF7 été traitées ou non pendant 24 heures avec une forte concentration de H₂O₂ (250μM). **(B) YPEL5 régule négativement les gènes de sénescence dans des cellules MCF7.** Les cellules MCF7 été transfectées avec si-SCR ou si-*YPEL5* pendant 48 heures ; les niveaux d'expression de YPEL5 ont diminué à une moyenne de 10% par rapport aux cellules transfectées avec si-SCR. L'abondance relative des transcrits d'ARNm a été quantifiée par RT-qPCR. Toutes les valeurs représentent les moyennes (± SD) des résultats obtenus dans des duplicatas à partir de 4 expériences individuelles (n = 8).****P< 0.0001

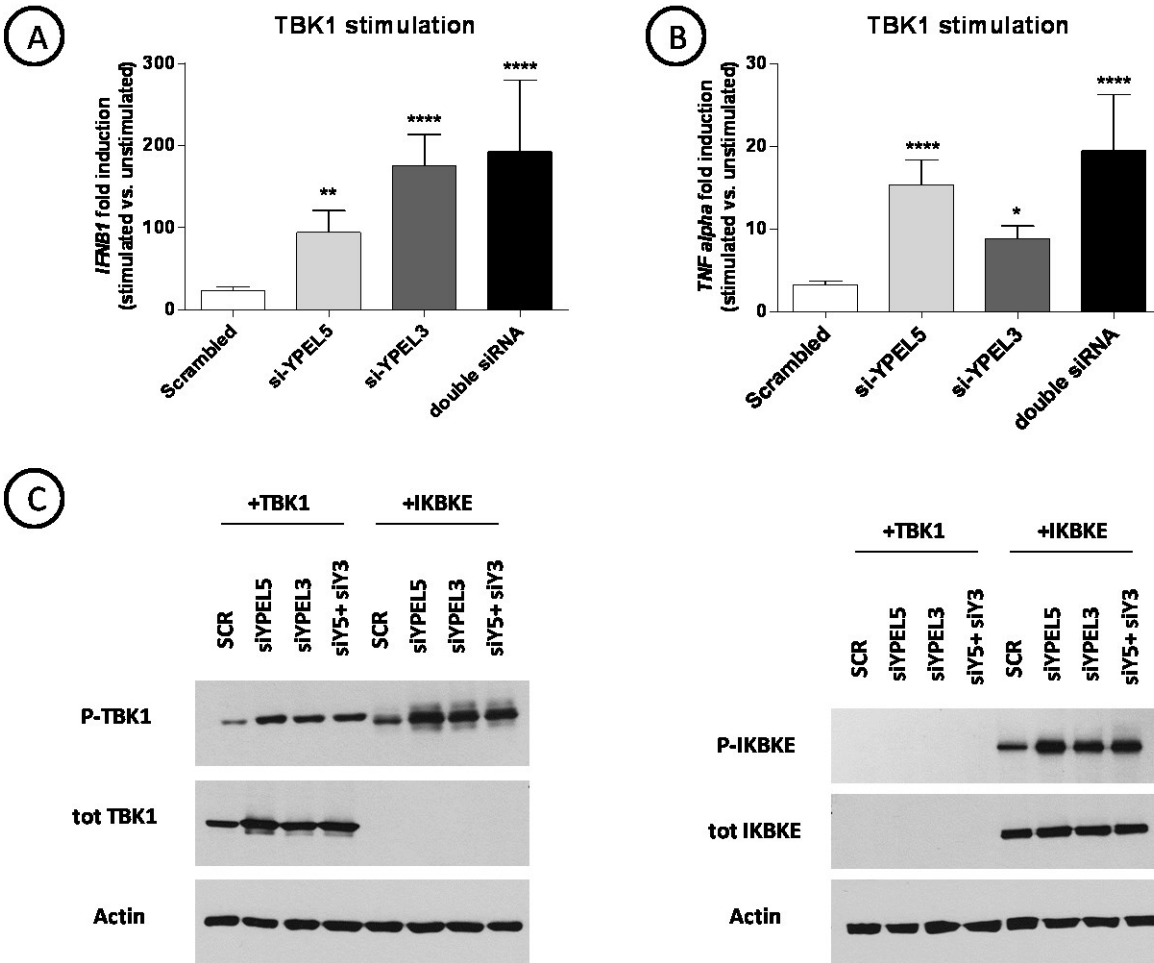


Figure 3.3: Les si-*YPEL5* et si-*YPEL3* amplifient l'induction d'IFNβ1 (A) et de TNFα (B) dans des cellules HEK 293T, avec une transfection des deux si-ARN ayant un effet additif. Les cellules ont été transfectées avec du si-ARN pendant 24 heures, puis stimulées pendant 24 heures supplémentaires soit avec pcDNA3.1 (plasmide vide) soit avec des plasmides d'expression pour TBK1. L'abondance d'*IFNβ1* (A) ou de *TNFα* (B) a été évaluée par RT-qPCR. Les interactions entre les effets des stimuli et ceux des siRNAs ont été testées par des ANOVA bidirectionnels. Toutes les valeurs représentent les moyennes (\pm SD) des résultats obtenus dans les puits répliqués à partir de plusieurs expériences individuelles (n = 6-10). **** P < 0,0001). (C) Si-*YPEL5* et si-*YPEL3* ont augmenté les effets de l'IKKε et de TBK1 surexprimé sur leur auto-ou cross phosphorylation. Les cellules HEK293T ont été traitées soit par si-*YPEL5*, soit par si-*YPEL3*, soit par les deux pendant 24 heures, puis elles ont été transfectées pendant 24 heures

supplémentaires soit avec pcDNA3.1 (plasmide vide) soit avec des plasmides d'expression pour TBK1 ou IKBKE.

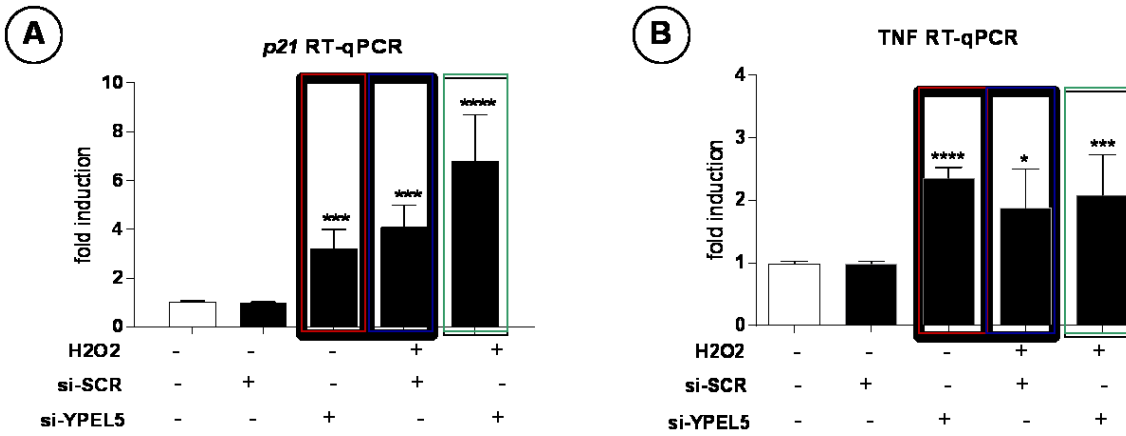


Figure 3.4: le Knockdown de *YPEL5* amplifie l'induction de *p21* par le traitement de H_2O_2 . L'expression de *p21* (A) ou de *TNFA* (B) a été quantifiée par RT-qPCR dans des cellules MCF7 transfectées avec un si-SCR ou un si-*YPEL5* pendant 24 heures, puis traitées ou non pendant 24 heures supplémentaires avec un milieu de culture contenant 250 μ M de H_2O_2 . Toutes les valeurs représentent les moyennes (\pm SD) des résultats obtenus dans des duplicatas à partir de 4 expériences individuelles (n = 8). ****P < 0.0001

Gene Symbol	Fold change	P-value	FDR
<u>Ptpn14</u>	2.99	1E-05	2E-03
Ypel3	1.55	2E-06	5E-04
<u>Ddit4</u>	1.47	1E-03	5E-02
Casp4	1.40	2E-04	2E-02
<u>Rhob</u>	1.37	3E-03	9E-02
Gch1	1.29	2E-09	2E-06
Cdkn1a	1.19	5E-04	3E-02
Stambp	1.21	2E-03	6E-02
Eif2ak2	1.17	7E-06	1E-03
Cd74	1.21	4E-05	5E-03
Casp8	1.17	1E-03	5E-02
Fadd	1.16	2E-03	8E-02
Ercc3	1.17	9E-06	1E-03
Tmbim6	1.19	4E-05	4E-03
Pik3cg	1.16	7E-04	3E-02

Tableau 3.I : Si-*Ypel5* dans les cellules RAW164.7 régule de nombreux gènes associés à la voie de sénescence cellulaire et ou aux voies de la mort cellulaire programmée (Illumina screen)

CHAPITRE 4 : DISCUSSION ET CONCLUSION

4.1 Discussion

L'identification de variants de gènes causalement liés à des caractères phénotypiques complexes demeure difficile. Les études de génomique fonctionnelle, qui évaluent les conséquences fonctionnelles des variations génétiques sur les caractères moléculaires intermédiaires, ont été proposées comme moyen d'améliorer la puissance de détection de telles variations génétiques^{419,420}. L'importance de l'expression génique pour les caractères complexes est illustrée par des études montrant que les polymorphismes associés à des variations de traits phénotypiques sont plus susceptibles d'être associés à l'expression de gènes particuliers⁴²¹ et que les variants associés à des maladies humaines courantes impliquent principalement des séquences d'ADN régulatrices (plutôt que des régions codantes pour une protéine)⁸⁹. En conséquence, l'expression des gènes constitue un type de phénotype moléculaire intermédiaire qui a été souvent étudié. Des études antérieures sur la génétique de l'expression ont principalement porté sur les eQTLs ('*cis*-eQTLs' et '*trans*-eQTLs') liés à des gènes uniques.^{22,125,132,422} Dans les cas particuliers où les effets des variants alléliques uniques étaient très pénétrants, cette approche a permis d'identifier des variants alléliques de gènes qui sont causaux de plusieurs caractères quantitatifs, parmi lesquels l'hypertension⁴²³, la masse ventriculaire gauche^{424,425}, et la calcification cardiaque dystrophique⁴²⁶. Néanmoins, dans les cas plus fréquents, où les traits complexes impliquent l'interaction entre divers gènes (qui opèrent dans des réseaux), l'identification d'une variation génétique qui affecte un seul gène n'est pas suffisante. Il a été observé que les '*trans*-eQTLs' forment des «hotspots», où les niveaux d'ARNm des transcrits à travers le génome montrent une liaison avec le même locus génétique. De tels '*trans*-eQTL hotspots' sont également souvent enrichis en gènes fonctionnellement apparentés^{172,174,188}, qui influent le même trait phénotypique. L'utilisation d'un outil informatique permettant l'analyse multivariée des eQTLs (qui tient compte des interactions possibles entre les gènes et/ou d'autres effets combinatoires possibles), nous a permis de détecter un '*trans*-eQTL hotspot' (dans des données d'expression de tissu cardiaque d'une population de souris provenant de 24 RIS) qui contenait un nombre très élevé de gènes associés en *trans*- (200 gènes). Donc, le but de cette thèse fut de : 1) identifier le/ou les régulateurs

causaux de ce réseau de gènes, et 2) tester et valider expérimentalement le mécanisme de fonctionnement du gène régulateur.

4.1.1 La validation biologiquement du ‘*trans*-eQTL hotspot’ au ch17

Pour la majorité des ‘*trans*-eQTL hotspots’ rapportés dans la littérature jusqu’à présent, un ‘*cis*-eQTL’ a été détecté au même locus (voir tableau 1.I). Ceci est intéressant, car les ‘*cis*-eQTLs’ pourraient être des régulateurs de l’expression de plusieurs gènes ‘*trans*-eQTLs’. Ce gène ‘*cis*-eQTL’ peut être un régulateur transcriptionnel direct (un TF qui cible et se lie directement à la région régulatrice de plusieurs gènes ‘*trans*-eQTLs’) ^{105,128,177,179,180,427}, comme il peut être un gène qui cible un TF ou un régulateur transcriptionnel intermédiaire ^{139,168,174,182,191,192,194}. Alors que l’identification de QTL d’expression a le potentiel de découvrir des gènes qui n’étaient pas précédemment connus comme étant des régulateurs communs, ces études précédentes se sont surtout appuyées sur des analyses bio-informatiques pour suggérer des rôles régulateurs pour les gènes ‘*cis*-eQTL’. Dans cette étude, le ‘*trans*-eQTL hotspot’ détecté au chr 17 contenait un ‘*cis*-eQTL’ qui correspondait au gène *Ypel5*. Des tests expérimentaux effectués par nous-même ont confirmé le rôle du gène *Ypel5* dans la régulation d’*IFNB1* et des ‘*trans*-eQTLs’ appartenant au ‘hotspot’. Néanmoins, il faut garder à l’esprit que ce *cis*-eQTL n’est pas le seul régulateur possible. Des éléments comme les marques épigénétiques (qui n’impliquent pas nécessairement de changements dans la séquence d’ADN mais qui sont potentiellement héréditaires, tel que la méthylation de l’ADN et l’acétylation des histones) et/ou les microARN ont également la capacité de réguler l’expression de plusieurs gènes.

4.1.2 Le choix du gène régulateur du ‘*trans*-eQTL hotspot’

Le choix de *Ypel5* comme gène régulateur a été basé sur des analyses bioinformatiques préalables et sur l’hypothèse selon laquelle la régulation *trans*- peut être médiée par des effets locaux. Aucun rôle de ce gène dans la régulation des ISGs n’avait été documenté à ce jour. Dans les études précédentes de validation biologique des ‘*trans*-eQTL hotspot’, le choix du gène régulateur candidat avait souvent été basé sur d’autres informations préalables indiquant que le gène était : 1) soit un gène ayant une fonction connue en relation avec le phénotype étudié ou

ayant une fonction prédite des ‘*trans*-eQTLs’^{168,191}, 2) soit un TF ayant un motif de liaison dans des régions promotrices des gènes cibles^{105,177,179}.

4.1.3 La fonction ou l’intérêt biologique du ‘*trans*-eQTL hotspot’ au chr 17

Le ‘*trans*-eQTL hotspot’ détecté au chr17 est enrichi de façon très significative pour des gènes immunitaires, dont la dérégulation peut être impliquée dans la pathogenèse d'un large éventail de phénotypes courants, y compris la maladie inflammatoire de l'intestin, l'athérosclérose, la polyarthrite rhumatoïde et le cancer. De plus, une proportion significative de loci de risque de maladies communes identifiées dans les études GWAS impliquent également des gènes immunitaires. Toutefois, nous n’avons pas identifié à ce jour de phénotype cardiaque dont le QTL chevauche avec le locus lié à *Ypel5* et au ‘*trans*-eQTL hotspot’. Aucun chevauchement entre le locus au chr17 et les loci de maladies rapportées dans le catalogue GWAS (www.genome.gov/gwastudies/) (données non présentées) n’a été détecté.

Une explication possible serait que, d'une manière générale, les SNPs détectés par les études GWAS n’expliquent qu’une faible partie de l’héritabilité. L’héritabilité manquante des études d’association peut être due en partie 1) au fait que la plupart des variants détectés par les GWAS sont le plus souvent des variants communs, et ne sont pas des variants fonctionnels qui causent la maladie; 2) au phénomène de ‘*phenotypic buffering*’, où de nombreux SNPs avec de faibles effets phénotypiques sont indétectables par les analyses du GWAS. Dans de tels cas, l'effet de ces SNPs (qui peuvent être associés à des traits) semble être plus important sur l'expression génique que sur le phénotype éventuellement associé. Ceci soutient l'idée que l'étude de l'expression génique comme phénotype intermédiaire pourrait aider à découvrir une partie de l'hérédité manquante dans les maladies complexes, de même qu'elle peut permettre d'identifier les gènes de la maladie affectée en aval (qui n'étaient pas impliqués du tout dans les études GWAS), et de révéler des voies précédemment inconnues. Toutefois, ce genre d'étude demeure incomplète et n’explique pas encore la totalité de l’héritabilité des traits complexes, car très peu d’études prennent en compte la possibilité de présence de plusieurs perturbations géniques (comme les interactions épistatiques), ou non géniques (comme les facteurs épigénétiques et les interactions entre les gènes et l’environnement). Par ailleurs, pour savoir de façon plus complète comment un génotype détermine un phénotype de trait quantitatif complexe, la détermination de la relation biologique réelle entre la variation du niveau de

transcription et d'autres phénotypes, tels que les taux de protéines et de métabolites, qui influencent à leur tour les phénotypes caractéristiques plus en aval, est une tâche nécessaire mais très complexe, qui peut être affectée par plusieurs facteurs ⁴²⁸. Par exemple, la présence d'un système de tampon (*'buffering system'*) au niveau post-transcriptionnel ou post-traductionnel (comme la dégradation de protéines non assemblées) peut diminuer la corrélation entre les niveaux de transcription et les niveaux de protéines. De plus, l'inexactitude dans les mesures de l'accumulation globale des transcrits peut réduire la capacité à détecter la relation entre les niveaux d'ARNm et les phénotypes en aval.

4.1.4 La découverte d'un nouveau rôle pour YPEL5

4.1.4.1 L'interaction de YPEL5 avec TBK1 / IKBKE

Les études effectuées dans cette thèse nous ont permis de découvrir un nouveau rôle pour YPEL5. En particulier, la régulation négative de YPEL5 a provoqué une expression accrue de l'*IFNBI* et de plusieurs ISGs dans les cellules de souris et humaines. Ces effets proviennent du fait que YPEL5 interagit physiquement et fonctionnellement avec les kinases TBK1/IKBKE. La délétion de *YPEL5*, dans les HEK293T, augmentait également les effets de la surexpression de TBK1/IKBKE sur l'activation du p NF- κ B -LUC et l'induction du facteur de nécrose tumorale (*TNF α*) endogène.

Les kinases TBK1/IKBKE

TBK1/IKBKE sont des I κ B kinases non canoniques (IKK) impliquées dans la régulation de l'activation des voies de signalisation d'IRF3 et NF- κ B. En tant qu'IKK non canonique, TBK1 est structurellement similaire à IKBKE (61% d'identité) et aux IKKs canoniques (IKK α et IKK β) ^{464,466}. Bien que TBK1 et IKBKE aient des activités semblables sur l'expression d'IFN de type I et les réponses antivirales subséquentes, elles ne semblent pas être redondantes et présentent des modèles d'expression différentielle et une spécificité de substrats. Contrairement au TBK1, qui est exprimé de manière constitutive dans pratiquement tous les types cellulaires, IKBKE est principalement exprimé dans des tissus spécifiques tels que le pancréas, le thymus, la rate et les leucocytes du sang périphérique ⁴³². Il est également exprimé à des niveaux très bas dans des lignées cellulaires spécifiques ^{229,433,434}. Il est important de noter que l'ARNm d'IKBKE

peut être induit en réponse à une exposition au LPS, à une infection virale ou à un certain nombre de cytokines inductrices de NF- κ B^{432,433,435,436}.

La régulation de l'activation de TBK1 / IKBKE

En tant que kinases critiques impliquées dans l'immunité antivirale, l'activité de TBK1/IKBKE doit être strictement régulée pour maintenir l'homéostasie immunitaire. L'activité aberrante des deux kinases pourrait être associée à une variété de maladies auto-immunes et de cancers. L'activation des deux kinases peut être régulée par diverses manières, telles que:

1) l'autophosphorylation à la Ser172^{429,437} (qui peut être facilitée par divers molécules comme la GSK3 β ⁴³⁸, la DDX3^{228,320,327,439}, et IKK α / β ⁴⁴⁰.

2) la déphosphorylation, initiée par plusieurs phosphatases telle que SHIP-1 (*inositol 5' phosphatase*)⁴⁴¹, PP2C β (*phosphatase Mg²⁺/Mn²⁺ dependent 1B*)⁴⁴², et PTP-2C (*protein-tyrosine phosphatase 2C*)⁴⁴³.

3) l'ubiquitination de la lysine 63 (K63) par TRAF3⁴⁴⁴, MIB1 et MIB2 (*E3 ubiquitin ligases mind bomb 1/2*)⁴⁴⁵ et Nrdp1⁴⁴⁶.

4) la deubiquitination médiée par CYLD (*deubiquitinating enzyme cylindromatosis*)^{447,448}, et ABIN1 (*A20 binding inhibitor of NF- κ B 1*)^{449,450}.

5) la prévention de la formation de complexes fonctionnels peut aussi inhiber l'activité de TBK1. Ceci implique des molécules comme MIP-T3⁴⁵¹, ISG56 (*IFN-stimulated gene 56*)⁴⁵², et SIKE (*suppressor of IKBKE*)⁴⁵³.

Dans cette étude, nous avons montré que YPEL5 interagit physiquement avec la kinase IKBKE, et régule négativement l'auto- et la cross- phosphorylation (à la Ser172) de TBK1 et IKBKE.

Le rôle de TBK1 et IKBKE dans l'activation d'IFNB et de NF- κ B

TBK1/IKBKE sont des kinases essentielles pour induire la phosphorylation d'IRF3, qui conduit à la production d'IFNB lorsque divers agents infectieux sont reconnus par des PRRs (voir chapitre 1, page 37). Elles sont aussi essentielles pour l'activation de NF- κ B. Il a été montré que la surexpression des deux kinases induit l'activation de NF- κ B^{234,454,455}.

Ila a été montré également que TBK1 ainsi que IKKε peuvent phosphoryler le résidu sérine 36 d'IκBα⁴⁵⁶. Il est possible aussi que TBK1 et IKKε phosphorylent les IKKs canoniques qui, à leur tour, mènent à l'activation de IκBα^{457,458}. De façon intéressante, TBK1 et IKKε ciblent aussi p65 / RelA à la Sérine 536 à un niveau basal, indépendamment des stimuli extracellulaires ; cette modification est requise dans la seconde phase d'activation de NF-κB^{493,496}. Dans les lymphocytes T stimulés, IKKε augmentait l'activité de p65 après sa phosphorylation à la Sérine 468⁴⁶³. La phosphorylation de TANK par IKKε pourrait également induire la libération de TRAF2 suivie de l'activation ultérieure du complexe IKK⁴⁶⁴. Pour limiter l'activation de la voie NF-κB, les IKKs canoniques sont régulées négativement par TBK1 et IKKε⁴⁴⁰. Une fois activés, TBK1 et IKKε phosphorylent IKKα / β, diminuant l'activité du complexe canonique IKK^{440,465}. De plus, TBK1 phosphoryle et conduit à la dégradation du NIK (*NF-κB-inducing kinase*), une protéine kinase qui active IKKα dans la voie non-canonique NF-κB dans les cellules B⁴⁶⁶.

TBK1/IKKε peuvent avoir d'autres cibles de phosphorylation. Par exemple, TBK1/IKKε phosphorylent directement Akt-Thr308 / Ser473. De manière unique, TBK1/IKKε peuvent phosphoryler à la fois la boucle d'activation et le motif hydrophobe à un degré suffisant pour activer l'AKT^{467,468}. TBK1 induit la signalisation de pro-survie par phosphorylation directe d'AKT⁴⁶⁹. La voie PI3K/Akt peut favoriser la survie cellulaire à travers la régulation d'expression ou la phosphorylation de la p21^{502,506}. Indépendamment de NF-κB et AKT, TBK1 peut réguler la mitose et initier la survie des cellules de cancer du poumon par l'activation de PLK1 (*Polo-like kinase 1*) et la 'metadherin'⁴⁷⁵. Les kinases TBK1/IKKε phosphorylent également les récepteurs d'œstrogène alpha (ERα) (sur la Ser-305 et la sérine 167 respectivement). L'inhibition des deux kinases sensibilise les cellules cancéreuses du sein à la mort cellulaire induite par le tamoxifène^{476,477}. De plus, TBK1 peut réguler la dynamique des microtubules et la progression mitotique via la phosphorylation des protéines centrosomiques⁴⁷⁸.

Le rôle de TBK1 / IKKε dans la régulation du cycle cellulaire et le cancer

Dans le cancer, les kinases TBK1/IKKε sont décrites comme des oncogènes possibles qui exercent leurs effets à travers l'activation de plusieurs voies, incluant notamment l'activation constitutive de NF-κB^{318,320,492,511}. Plusieurs études précédentes ont montré que l'expression de

la forme inactive de la kinase IKBKE dans les cellules cancéreuses du sein réduit les niveaux d'expression de deux gènes cibles de NF- κ B (la Cyclin D1 et RelB), ainsi que la croissance et l'invasion ⁴⁵⁸. Aussi, l'inhibition de TBK1/IKBKE réduit la phosphorylation/activation de la p65-NF κ B et supprime la croissance des cellules du cancer du sein humaines (HER2 +) ³⁴¹. De plus, IKBKE phosphoryle CYLD et TRAF2 dans les cellules cancéreuses du sein, ce qui induit l'activation NF- κ B et contribue à la transformation cellulaire ^{480,481}. Des observations récentes ont montré que IKBKE fonctionne également pour protéger les cellules contre la mort cellulaire induite par les dommages d'ADN ; le mécanisme sous-jacent de cette réponse est médié par la sumoylation de IKBKE, suivie par la phosphorylation de la p65 et l'activation de la réponse anti-apoptotique médiée par NF- κ B ⁴⁸².

L'activation constitutive de la voie NF- κ B dans le cancer du sein est très bien documentée ^{515,519}. Cependant son rôle dans la régulation de la progression tumorale est controversé. La plupart des études appuient l'effet de NF- κ B comme une voie promotrice des mécanismes contribuant à la conversion maligne, comme la prolifération, l'angiogenèse, et l'induction de la transition épithéliale-mésenchymateuse (un événement précoce dans la métastase) ^{520,524}. L'activation de NF- κ B favorise également la prolifération et/ou la progression du cycle cellulaire ^{493,494} à travers l'augmentation de l'expression des gènes de croissance/survie cellulaire ^{495,496}. Des études précédentes ont montré que NF- κ B induit l'expression de la *cycline D1* et favorise la progression du cycle cellulaire (de la phase G1- à- S) dans les fibroblastes de souris et dans les cellules de carcinome mammaire T47D ^{528,530}. Dans les CEM humaine T leucémique et les lignes cellulaires de cancer du sein humain (MDA-MB-231), l'induction de la p21 par NF- κ B provoque un arrêt cellulaire en phase G2-M pendant un temps prolongé, suivi d'une augmentation de la rentrée du cycle cellulaire et de la survie ⁴⁹⁹.

Le gène *TNF- α* est également activé par NF- κ B; cependant, son rôle dans la progression des tumeurs est controversé. Il a été d'abord identifié comme une cytokine anti-tumorale accompagnée d'une toxicité importante. De nouvelles preuves ont montré que le TNF- α était l'un des principaux médiateurs de l'inflammation liée au cancer et agit comme un facteur favorisant les tumeurs. Cet effet apparemment paradoxal du TNF- α sur l'initiation de tumeurs peut dépendre de la dose et de la durée d'exposition à cette cytokine. Des études précédentes ont montré que le traitement avec de faibles doses de TNF- α peut améliorer la capacité invasive des cellules MCF-7 ⁵⁰⁰. De manière contradictoire, il a été montré que le traitement de la lignée

cellulaire de carcinome du col utérin humain (ME-180S) avec TNF- α entraîne une induction de la p53, la p21 et de l'apoptose⁵⁰¹. L'effet pro-apoptotique de cette cytokine a été confirmé dans des études montrant que la p21 induite par le TNF- α , dans les lignées cellulaires de gliome humain (LN-18 et LN229), initie l'inhibition du cycle cellulaire⁵⁰². D'autres ont trouvé également que l'incubation des lymphocytes avec du TNF- α recombinant augmente les niveaux de NF- κ B et de la p21⁵⁰³. Une étude récente a montré que TNF- α augmente l'expression des régulateurs participant à la transition du cycle cellulaire G1 \rightarrow S par une voie dépendante de NF- κ B, ce qui rend les cellules plus vulnérables et provoque plus de dommages à l'ADN après les traitements de radiothérapie et de chimiothérapie⁵⁰⁴. D'autres études ont proposé que le TNF- α augmente l'effet anti-cancéreux de la doxorubicine en supprimant l'activité anti-apoptotique de p21⁵⁰⁵. Dans les cellules MCF-7, l'action antiproliférative du TNF- α a été associée à plusieurs facteurs^{506,507}: 1) l'expression accrue de *p21*, l'activation de NF- κ B et l'accumulation de p53 (accompagné d'un arrêt G1)⁵⁰⁸, 2) la délocalisation cytoplasmique de p21 et p27⁵⁴⁷, 3) l'augmentation de l'accumulation de IGFBP-3 (*insulin-like growth factor binding protein-3*)⁵¹⁰, et 4) la régulation négative de l'expression de l'ER alpha⁵¹¹.

4.1.4.2 L'interaction de YPEL5 avec les protéines centrosomiques

La protéine YPEL5 a été détectée à différentes localisations subcellulaires lors du cycle cellulaire. À l'interphase, elle semble être localisée dans le noyau et dans le centrosome. Lors de la division cellulaire, elle change séquentiellement d'emplacement aux pôles de fuseau, au fuseau mitotique, à la midzone de fuseau, et finalement à la zone midbody. Le knockdown de *YPEL5* inhibe la croissance de cellules COS-7 cultivées et le développement précoce d'embryons de poissons médaka, ce qui indique son implication dans la progression du cycle cellulaire. De façon intéressante, YPEL5 montre des interactions physiques avec les deux protéines centrosomiques RanBPM (*Ran Binding Protein in the Microtubule organizing center*), codé par RanBP9, et RanBP10³²⁵.

Le rôle du centrosome dans la régulation du cycle cellulaire et le cancer

Il a été montré par plusieurs études que l'intégrité du centrosome est importante pour la progression du cycle cellulaire. L'arrêt du cycle cellulaire a été observé dans les cellules cultivées suite à une perturbation des centrosomes par ablation laser⁵¹² ou microchirurgie⁵¹³ et

dans les ovocytes de souris suite à la microinjection d'anticorps de centrosomes ⁵¹⁴. Aussi, la fonction du centrosome peut être perturbée par plusieurs facteurs pathogènes ^{547,549} ou de stress (comme les dommages d'ADN et la chaleur) ⁵¹⁸, et ces interventions empêchent la progression du cycle cellulaire. Il a été montré également que le dérèglement de l'expression de presque tous les gènes du centrosome induisaient l'arrêt du cycle cellulaire (au G1)⁵¹⁹. Ces études suggèrent que la perturbation de la structure et / ou de la fonction du centrosome active un 'cell cycle checkpoint' qui conduit à l'arrêt G1 (et empêche les cellules d'entrer en phase S). Ce mécanisme d'arrêt du cycle cellulaire impliquerait le recrutement de p53 au centrosome, suivi par l'activation de la p21 ^{545,551,554}. La délétion de p53 dans de nombreuses tumeurs humaines peut abroger le checkpoint des centrosomes et contribuer aux défauts du centrosome, au dysfonctionnement de fuseau et à l'aneuploïdie ^{523,524}. En fait, la plupart des carcinomes humains sont caractérisés par des centrosomes aberrants, qui mènent à la formation de fuseaux dysfonctionnels et contribuent à l'instabilité génétique ^{523,525}.

L'inactivation des centrosomes est liée aussi à l'inhibition de la réponse immunitaire. L'idée émergente selon laquelle le centrosome joue un rôle dans la réponse immunitaire est également appuyée par l'observation que des patients atteints de mutations dans le gène centrosomique la pericentrine / Pcnt possèdent une réponse immunitaire défectueuse pendant l'enfance ⁵²⁶. Dans ce contexte, l'observation que l'inhibition de la destruction des cellules cibles par les cellules cytotoxique T par le stress thermique est accompagnée par le dysfonctionnement des centrosomes (qui se manifeste par la perturbation de l'organisation des microtubules) ⁵²⁷, implique que le centrosome peut être un organe-clé pour l'activation la réponse immunitaire. Aussi, des observations récentes sur le comportement du centrosome à la synapse immunologique suggèrent un rôle critique pour la polarisation des centrosomes dans le contrôle de la communication entre les cellules immunitaires nécessaires pour générer une réponse immunitaire efficace ^{560,562}.

Le rôle de RanBPM dans la régulation du cycle cellulaire et le cancer

RanBPM a été initialement identifié comme une protéine centrosomique impliquée dans la nucléation des microtubules au centrosome ⁵³¹. Récemment, plusieurs études ont suggéré la contribution de RanBPM dans la croissance et le développement du cerveau ⁵³². RanBPM accélère également la progression de la mitose dans le développement de neuro-épithélium ⁵³³.

La participation de RanBPM dans la régulation du cycle cellulaire et les voies proapoptotiques a été suggérée en fonction de la capacité de RanBPM à interagir avec de multiples protéines impliquées dans des cascades de signalisation apoptotiques, telles que 1) CDK11(p46) (*cyclin-dependent kinase 11*)⁵³⁴ ; 2) le ‘*death domain*’ du récepteur de la neurotrophine p75⁵³⁵ ; 3) HIPK2 (*homeodomain-interacting protein kinase 2*) et la p73^{536,537}. Il existe également des preuves de l'implication de RanBPM dans les voies de signalisation suscitées par les signaux environnementaux. RanBPM est une phosphoprotéine dont la phosphorylation est modulée par des stimuli de stress comme le choc osmotique, le rayonnement UV, et le rayonnement ionisant (IR)^{538,539}. En revanche, il a été montré que les taux d'expression de RanBPM sont nettement réduits dans les lignées cellulaires de cancer du sein humain⁵⁴⁰ et dans les tumeurs gastriques⁵⁴¹. Le knockdown de *RanBPM* dans les cellules cancéreuses gastriques réduit l'adhésion et favorise la survie des cellules cancéreuses gastriques⁵⁴¹, mais entrave les mécanismes moléculaires conduisant à une réparation efficace des dommages de l'ADN dans des cellules du cancer du poumon⁵⁴². De plus, la surexpression de RanBPM dans une lignée cellulaire de carcinome rénal humain (A704) augmente la capacité de migration⁵⁴³.

RanBPM est impliqué également dans les réponses immunitaires. Une étude a montré que RanBPM peut réprimer la voie NF- κ B en inhibant l'auto-ubiquitination de TRAF6 (via l'interaction protéine-protéine avec TRAF6) dans les HEK 293T⁵⁴⁴. Cependant, une autre étude suggère que RANBP9 peut améliorer l'efficacité de l'efferocytose des cellules dendritiques, et la présentation d'antigène in vivo (ce qui provoque une diminution de charge virale et une amélioration de la survie)⁵⁴⁵.

Le rôle de YPEL5 dans la régulation du cycle cellulaire et le cancer

Dans ce projet, on a trouvé que le knockdown de *YPEL5* dans les MCF7 induit l'expression des gènes impliqués dans la régulation du cycle cellulaire : la *p21*, *YPEL3*, et *TNF α* (voir chapitre 3) Ce résultat peut être compatible avec d'autres études montrant que le knockdown de *YPEL5* : 1) retardait la progression du cycle cellulaire en prolongeant la durée des phases G1 et G2 + M dans les cellules COS-7³²⁵, et 2) induit l'expression d'un ensemble de gènes régulateurs de sénescence (incluant la *p21* et *YPEL3*) dans les cellules de macrophages de souris (RAW 264.7) (voir tableau 3.I). Considérant sa localisation dans le centrosome et son interaction avec plusieurs protéines impliquées dans le fonctionnement du centrosome^{478,531}, il

est possible de supposer que le knockdown de *YPEL5* cause une perturbation de la fonction du centrosome suivie par l'induction des médiateurs de l'arrêt du cycle cellulaire. L'arrêt du cycle cellulaire empêche l'accumulation de centrosomes aberrants et l'initiation du cancer.

De plus, nous avons observé que l'exposition à H₂O₂ (un stress génotoxique) induit l'expression de *YPEL5*, mais pas des autres membres de la famille *YPEL*. Le knockdown de *YPEL5* a amplifié l'induction de la *p21* par H₂O₂. Ces résultats suggèrent que *YPEL5* est un régulateur négatif (par rétroaction) de l'induction de sénescence déclenchée par H₂O₂. Étant donné que de nombreux médicaments chimio thérapeutiques sont connus pour induire la sénescence dans les cellules cancéreuses par une voie dépendante de l'induction des ROS et de la *p21*, l'inhibition de *YPEL5* pourrait augmenter la sensibilité des cellules tumorales aux agents chimio thérapeutiques. Cependant, pour confirmer l'implication de *YPEL5* dans la régulation du cycle cellulaire et/ou la sénescence dans les cellules du cancer du sein, d'autres événements doivent être étudiés plus en détail. Donc, dans les cellules en division, il faut tester l'impact de la régulation de *YPEL5* sur le temps de doublement cellulaire, la proportion de cellules à chaque étape du cycle cellulaire (G1, S et G2 / M) et la durée du temps passé par les cellules à chaque stade de cycle cellulaire.

Le rôle de *YPEL5* dans le développement embryonnaire

Le rôle joué par *YPEL5* dans la régulation du cycle cellulaire et la sénescence, pourrait être également lié à l'implication de *YPEL5* (et ou d'autres *YPEL*) dans le développement embryonnaire. Les gènes *YPEL* ont été conservés tout au long de l'évolution. Une étude récente suggère que le gène *Ypel2* peut jouer un rôle essentiel dans le développement embryonnaire précoce, car *YPEL2* est régulé par miR-29b⁵⁴⁶, une miRNA impliqué au cours du développement préimplantatoire et dont l'inhibition conduit à l'arrêt embryonnaire⁵⁴⁷. La surexpression de *YPEL1* entraîne également une morphogenèse mandibulaire anormale associée à une augmentation de l'apoptose⁵⁴⁸. De plus, il a été suggéré que *YPEL3* joue un rôle nécessaire pour le développement précoce du cerveau et du corps⁵⁴⁹.

Nous avons obtenu des cellules souches embryonnaires *Ypel5* - / - et les avons utilisées pour générer des souris knockout *Ypel5*. Les données obtenues jusqu'à présent indiquent que l'inactivation de *Ypel5* est embryonnaire létale, avec la létalité se produisant avant le jour E11.5.

Ces résultats sont compatibles avec les données antérieures obtenues dans les embryons de poissons, où le knockdown de *Ypel5* arrête le développement en augmentant l'apoptose au stade 17 (un stade neurulaire précoce) ³²⁴. De façon intéressante, les mutations qui perturbent les centrosomes activent fortement p53 et la mort cellulaire chez les embryons de souris au même stade de gestation où l'ablation *Ypel5* déclenche également la létalité embryonnaire.

Au total, ces données suggèrent que les gènes YPEL pourrait participer aux voies de sénescence, à la progression du cycle cellulaire et au développement embryonnaire, éventuellement via des interactions avec les protéines centrosomiques (RanBPM et RanBP10) et /ou les kinases (TBK1 et IKBKE).

4.1.4.3 L'interaction entre YPEL3 et YPEL5

Des études précédentes ont montré que YPEL3 et YPEL5 sont exprimés à des niveaux beaucoup plus élevés que n'importe quel autre gène YPEL dans des tissus humains et de souris ³²⁵, et nous avons fait la même observation dans les cellules humaines HEK293T et MCF7 (Figure 3.1). Ceci suscite la question de savoir si ces deux gènes interagissent, dans quelle mesure leurs fonctions peuvent se chevaucher et / ou si elles se régulent mutuellement.

Pour tester si YPEL3 et YPEL5 pourrait avoir des effets en partie similaires, nous avons comparé les effets du knockdown de ces deux gènes sur l'activation de la voie NF- κ B en présence de surexpression de TBK1 et IKBKE sur 1) l'expression de *TNF α* endogène (une cible bien connue de NF- κ B), et 2) l'auto- et la cross-phosphorylation de TBK1 et IKBKE. Ces points expérimentaux semblent être affectés de façon similaire par le knockdown de ces deux gènes (voir Figure 3.3). Par ailleurs, YPEL3 et YPEL5 semblent avoir des effets opposés sur des marqueurs de sénescence cellulaire. Dans les MCF7, on a trouvé que *YPEL3* ne régule ni l'expression de *p21* ni celle de *TNF α* . De même, le knockdown de *YPEL3* n'affectait pas l'induction de *p21* par H₂O₂. Conformément à ces résultats, d'autres études ont montré que l'induction de *YPEL3* initie la sénescence, mais que le knockdown de *YPEL3* était incapable de bloquer la sénescence induite par la doxorubicine dans les cellules MCF-7 ³⁶⁹. Cela suggère que seulement l'induction d'expression de YPEL3 pourrait avoir un effet significatif sur la sénescence. D'autres observations ont suggéré que YPEL5 et YPEL3 exercent des effets opposés sur la régulation de la sénescence ^{362,363,369,370}.

Enfin, on sait peu de choses sur la régulation de ces gènes, et en particulier 1) quels stimuli régulent l'expression de ces gènes; 2) si les gènes sont régulés de manière coordonnée; 3) s'il existe des effets de régulation croisée entre plusieurs membres de la famille YPEL. Tous ces différents facteurs doivent être pris en compte pour comprendre les actions des gènes YPEL. Il a été montré que l'expression de *YPEL3* peut être induite par : 1) la doxorubicin dans la lignée tumorale humaine de colon Hct116+/+p53 ; 2) la bléomycine dans les cellules épithéliales mammaires 76NTert³⁶² ; 3) le tamoxifène dans les MCF7³⁷⁰, mais pas par 5-azadéoxycytidine (un agent qui induit des dommages à l'ADN) dans les lignées cellulaires du cancer du côlon³⁷⁰. Nos résultats suggèrent que l'exposition des MCF7 à H₂O₂ induisait l'expression de *YPEL5*, mais n'affectait pas l'expression de *YPEL3*. Cela suggère que les niveaux d'expression de *YPEL3* et *YPEL5* peuvent être modulés de façons différentes, dépendamment de la nature, de la durée, de la concentration du stress genotoxique, et du type de lignée cellulaire utilisée.

4.2 Conclusion

Une question fondamentale en génétique quantitative est de savoir comment le génotype détermine un phénotype de trait quantitatif complexe⁵⁵⁰. Les études d'association représentent une avancée considérable dans l'identification des locus associés à des traits complexes, mais jusqu'à présent, très peu d'études ont permis de clarifier l'architecture génétique de ces traits. Les études des eQTLs pourraient représenter une approche alternative et ou complémentaire. Particulièrement, les études de '*trans*-eQTL hotspots' peuvent nous aider à identifier les variants génétiques affectant des réseaux de gènes co-régulés qui peuvent être associé à un ou plusieurs traits complexes. Nous avons étudié le rôle de *YPEL5* dans la régulation d'un '*trans*-eQTL hotspot' au chr17. *YPEL5* est une protéine dont les fonctions biologiques sont encore peu connues. Dans cette étude, nous rapportons l'identification de *YPEL5* comme régulateur négatif des voies : 1) IFN β et des ISGs (appartenant au '*trans*-eQTL hotspots'), et 2) NF κ B et TNF α , dans des conditions de bases et après stimulation dans les macrophages et les HEK293T. En plus de son rôle dans la réponse immunitaire, une publication antérieure a impliqué *YPEL5* dans la régulation de la progression du cycle cellulaire, et nos données préliminaires indiquent des interactions possibles de *YPEL5* avec le processus de sénescence. Au total, l'effet pléiotropique de *YPEL5* dans la régulation de la réponse immunitaire et du cycle cellulaire pourrait être dû à :

1) sa localisation dans le centrosome durant l'interphase; et/ou 2) à son interaction avec les RANBP et les kinases non canoniques (IKK), qui sont impliquées dans le fonctionnement du centrosome^{478,531}, les voies de survie cellulaire, et l'activation des voies de signalisation d'IFN β et NF κ B. Finalement, le rôle de YPEL5 dans la régulation négative de la production de cytokines pourrait être pertinent dans le cadre des troubles immunitaires liés à la sur-réactivité du système immunitaire, tels que l'auto-immunité [comme la polyarthrite rhumatoïde, la maladie de Crohn, le lupus érythémateux systémique (SLE) et la sclérose en plaques] et le cancer^{551,552}.

Bibliographie

1. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. *J. Exp. Med.* **79**, 137–158 (1944).
2. Watson, J. D. & Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967 (1953).
3. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
4. Structure et fonction des protéines. *De Boeck Supérieur* (2016). Available at: <http://www.deboecksuperieur.com/ouvrage/9782804158880-structure-et-fonction-des-protéines>. (Accessed: 6th November 2016)
5. Lee, P. J. & Lachmann, R. H. Acute presentations of inherited metabolic disease in adulthood. *Clin. Med. Lond. Engl.* **8**, 621–624 (2008).
6. Charles Darwin. *Wikipédia* (2016).
7. Morgan, T. H. RANDOM SEGREGATION VERSUS COUPLING IN MENDELIAN INHERITANCE. *Science* **34**, 384 (1911).
8. Albert, C. E.-H. Sturtevant, Labial Terminations III and IV, 1913. *Rev. Études Anciennes* **16**, 459–461 (1914).
9. Sturtevant, A. H. 1913. The linear arrangement of six sex-linked factors in *Drosophila*. (first genetic map) - HUM-MOLGEN literature. Available at: <http://hum-molgen.org/literature/10-1999/000002.html>. (Accessed: 23rd November 2016)
10. Hermann Joseph Muller. *Wikipedia* (2016).
11. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Wikipedia* (2016).

12. Hermann Joseph Muller. *Wikipédia* (2016).
13. Consortium, T. 1000 G. P. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
14. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
15. Gusella, J. F. *et al.* A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234–238 (1983).
16. Rommens, J. M. *et al.* Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059–1065 (1989).
17. Antonarakis, S. E. & Beckmann, J. S. Mendelian disorders deserve more attention. *Nat. Rev. Genet.* **7**, 277–282 (2006).
18. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–D517 (2005).
19. Lander, E. S. & Schork, N. J. Genetic dissection of complex traits. *Science* **265**, 2037–2048 (1994).
20. Abiola, O. *et al.* The nature and identification of quantitative trait loci: a community's view. *Nat. Rev. Genet.* **4**, 911–916 (2003).
21. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five Years of GWAS Discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
22. Idaghdour, Y., Storey, J. D., Jadallah, S. J. & Gibson, G. A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. *PLoS Genet.* **4**, e1000052 (2008).
23. De Jager, P. L. *et al.* Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat. Genet.* **41**, 776–782 (2009).

24. Perola, M. *et al.* Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genet.* **3**, e97 (2007).
25. Silventoinen, K. *et al.* Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res. Off. J. Int. Soc. Twin Stud.* **6**, 399–408 (2003).
26. Haley, C. S. & Knott, S. A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324 (1992).
27. Jansen, R. C. Interval Mapping of Multiple Quantitative Trait Loci. *Genetics* **135**, 205–211 (1993).
28. Jayakar, S. D. On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus. *Biometrics* **26**, 451–464 (1970).
29. Lander, E. S. & Botstein, D. Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* **121**, 185–199 (1989).
30. Thoday, J. M. Location of Polygenes. *Nature* **191**, 368–370 (1961).
31. Projet génome humain. *Wikipédia* (2016).
32. Broman, K. W. Review of statistical methods for QTL mapping in experimental crosses. *Lab Anim.* **30**, 44–52 (2001).
33. Martínez, O. & Curnow, R. N. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **85**, 480–488 (1992).
34. Zeng, Z. B. Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468 (1994).
35. Kao, C. H., Zeng, Z. B. & Teasdale, R. D. Multiple interval mapping for quantitative trait loci. *Genetics* **152**, 1203–1216 (1999).
36. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet. TIG* **24**, 408–415 (2008).

37. Bertina, R. M. *et al.* Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* **369**, 64–67 (1994).
38. Altshuler, D. *et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
39. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
40. Corder, E. H. *et al.* Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MAGene dose of apolipoprotein E type 4 allele and the risk of Alzheimer’s disease in late onset families. *Science* 261:921-923. *ResearchGate* **261**, 921–3 (1993).
41. Cudworth, A. G. & Woodrow, J. C. Genetic susceptibility in diabetes mellitus: analysis of the HLA association. *Br. Med. J.* **2**, 846–848 (1976).
42. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
43. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
44. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
45. Klein, R. J. *et al.* Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science* **308**, 385–389 (2005).
46. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
47. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).

48. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
49. Eeles, R. A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
50. Antoniou, A. C. *et al.* The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. *Br. J. Cancer* **98**, 1457–1466 (2008).
51. Hung, R. J. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633–637 (2008).
52. Loos, R. J. F. *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.* **40**, 768–775 (2008).
53. Gudbjartsson, D. F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615 (2008).
54. Stranger, B. E. & Dermitzakis, E. T. From DNA to RNA to disease and back: The ‘central dogma’ of regulatory disease variation. *Hum. Genomics* **2**, 383–390 (2006).
55. Wang, S. S. *et al.* Identification of Pathways for Atherosclerosis in Mice. *Circ. Res.* **101**, e11–e30 (2007).
56. Yang, X. *et al.* Identification and validation of genes affecting aortic lesions in mice. *J. Clin. Invest.* **120**, 2414–2422 (2010).
57. Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* **37**, 243–253 (2005).
58. McDermott-Roe, C. *et al.* Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478**, 114–118 (2011).
59. Hodgin, J. B. *et al.* Identification of Cross-Species Shared Transcriptional Networks of Diabetic Nephropathy in Human and Mouse Glomeruli. *Diabetes* **62**, 299–308 (2013).

60. Keller, M. P. & Attie, A. D. Physiological Insights Gained from Gene Expression Analysis in Obesity and Diabetes. *Annu. Rev. Nutr.* **30**, 341–364 (2010).
61. Wang, S. *et al.* Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet.* **2**, e15 (2006).
62. Yang, X. *et al.* Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nat. Genet.* **41**, 415–423 (2009).
63. Calabrese, G. *et al.* Systems Genetic Analysis of Osteoblast-Lineage Cells. *PLOS Genet.* **8**, e1003150 (2012).
64. Farber, C. R. *et al.* Mouse Genome-Wide Association and Systems Genetics Identify *Asxl2* As a Regulator of Bone Mineral Density and Osteoclastogenesis. *PLOS Genet.* **7**, e1002038 (2011).
65. Park, C. C. *et al.* Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Syst. Biol.* **5**, 43 (2011).
66. Davis, R. C. *et al.* Genome-wide association mapping of blood cell traits in mice. *Mamm. Genome* **24**, 105–118 (2013).
67. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–237 (2013).
68. Huang, J., Roberts, A. J., Leschziner, A. E. & Reck-Peterson, S. L. Lis1 Acts as a “Clutch” between the ATPase and Microtubule-Binding Domains of the Dynein Motor. *Cell* **150**, 975–986 (2012).
69. Ingvarsson, P. K. & Street, N. R. Association genetics of complex traits in plants. *New Phytol.* **189**, 909–922 (2011).
70. Stranger, B. E., Stahl, E. A. & Raj, T. Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics* **187**, 367–383 (2011).
71. Mural, R. J. *et al.* A Comparison of Whole-Genome Shotgun-Derived Mouse Chromosome 16 and the Human Genome. *Science* **296**, 1661–1671 (2002).

72. Erickson, Z. T., Falkenberg, E. A. & Metz, G. A. S. Lifespan Psychomotor Behaviour Profiles of Multigenerational Prenatal Stress and Artificial Food Dye Effects in Rats. *PLoS ONE* **9**, (2014).
73. Peters, L. L. *et al.* The mouse as a model for human biology: a resource guide for complex trait analysis. *Nat. Rev. Genet.* **8**, 58–69 (2007).
74. Alonso-Blanco, C. & Méndez-Vigo, B. Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. *Curr. Opin. Plant Biol.* **18**, 37–43 (2014).
75. Buchner, D. A. & Nadeau, J. H. Contrasting genetic architectures in different mouse reference populations used for studying complex traits. *Genome Res.* (2015). doi:10.1101/gr.187450.114
76. Marshall, J. R. *et al.* Smoking, alcohol, dentition and diet in the epidemiology of oral cancer. *Eur. J. Cancer. B. Oral Oncol.* **28B**, 9–15 (1992).
77. Rapp, J. T., Dozier, C. L., Carr, J. E., Patel, M. R. & Enloe, K. A. Functional analysis of hair manipulation: a replication and extension. *Behav. Interv.* **15**, 121–133 (2000).
78. Threadgill, D. W., Miller, D. R., Churchill, G. A. & de Villena, F. P.-M. The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *ILAR J.* **52**, 24–31 (2011).
79. Marshall, J. D. *et al.* The AXB and BXA set of recombinant inbred mouse strains. *Mamm. Genome Off. J. Int. Mamm. Genome Soc.* **3**, 669–680 (1992).
80. Williams, R. W., Gu, J., Qi, S. & Lu, L. The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol.* **2**, research0046.1-research0046.18 (2001).
81. Rapp, J. P. Genetic Analysis of Inherited Hypertension in the Rat. *Physiol. Rev.* **80**, 135–172 (2000).
82. Jacob, H. J. & Kwitek, A. E. Rat genetics: attaching physiology and pharmacology to the genome. *Nat. Rev. Genet.* **3**, 33–42 (2002).
83. Darvasi, A. & Soller, M. Advanced Intercross Lines, an Experimental Population for Fine Genetic Mapping. *Genetics* **141**, 1199–1207 (1995).

84. Threadgill, D. W. & Churchill, G. A. Ten years of the Collaborative Cross. *Genetics* **190**, 291–294 (2012).
85. Chia, R., Achilli, F., Festing, M. F. W. & Fisher, E. M. C. The origins and uses of mouse outbred stocks. *Nat. Genet.* **37**, 1181–1186 (2005).
86. Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. & Flint, J. A Method for Fine Mapping Quantitative Trait Loci in Outbred Animal Stocks. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 12649–12654 (2000).
87. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
88. Flint, J., Valdar, W., Shifman, S. & Mott, R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat. Rev. Genet.* **6**, 271–286 (2005).
89. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
90. Jansen, R. C. & Nap, J. P. Genetical genomics: the added value from segregation. *Trends Genet. TIG* **17**, 388–391 (2001).
91. Broman, K. W. The Genomes of Recombinant Inbred Lines. *Genetics* **169**, 1133–1146 (2005).
92. Albert, F. W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.* **16**, 197–212 (2015).
93. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
94. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**, 57–64 (2003).
95. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).

96. Gibson, G. & Weir, B. The quantitative genetics of transcription. *Trends Genet. TIG* **21**, 616–623 (2005).
97. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
98. Storey, J. D. *et al.* Gene-Expression Variation Within and Among Human Populations. *Am. J. Hum. Genet.* **80**, 502–509 (2007).
99. Kleinjan, D. A. & van Heyningen, V. Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *Am. J. Hum. Genet.* **76**, 8–32 (2005).
100. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
101. Ueda, H. *et al.* Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* **423**, 506–511 (2003).
102. Tournamille, C., Le Van Kim, C., Gane, P., Cartron, J. P. & Colin, Y. Molecular basis and PCR-DNA typing of the Fya/fyb blood group polymorphism. *Hum. Genet.* **95**, 407–410 (1995).
103. Tsuge, M. *et al.* A variable number of tandem repeats polymorphism in an E2F-1 binding element in the 5' flanking region of SMYD3 is a risk factor for human cancers. *Nat. Genet.* **37**, 1104–1107 (2005).
104. Ghazalpour, A. *et al.* Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* **6**, R59 (2005).
105. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
106. Yaguchi, H., Togawa, K., Moritani, M. & Itakura, M. Identification of candidate genes in the type 2 diabetes modifier locus using expression QTL. *Genomics* **85**, 591–599 (2005).
107. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinforma. Oxf. Engl.* **20**, 307–315 (2004).

108. Schurmann, C. *et al.* Analyzing Illumina Gene Expression Microarray Data from Different Tissues: Methodological Aspects of Data Analysis in the MetaXpress Consortium. *PLOS ONE* **7**, e50938 (2012).
109. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
110. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
111. Monks, S. A. *et al.* Genetic Inheritance of Gene Expression in Human Cell Lines. *Am. J. Hum. Genet.* **75**, 1094–1105 (2004).
112. Spielman, D., Brook, B. W. & Frankham, R. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15261–15264 (2004).
113. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
114. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**, 1378–1385 (2006).
115. Petronis, A. Epigenetics and twins: three variations on the theme. *Trends Genet. TIG* **22**, 347–350 (2006).
116. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701–703 (2005).
117. Dimas, A. S. *et al.* Sex-biased genetic effects on gene regulation in humans. *Genome Res.* **22**, 2368–2375 (2012).
118. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
119. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

120. Knezetic, J. A. & Luse, D. S. The presence of nucleosomes on a DNA template prevents initiation by RNA polymerase II in vitro. *Cell* **45**, 95–104 (1986).
121. Montgomery, S. B. & Dermitzakis, E. T. From expression QTLs to personalized transcriptomics. *Nat. Rev. Genet.* **12**, 277–282 (2011).
122. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic Variation in Human Gene Expression. *Science* **297**, 1143–1143 (2002).
123. Ronald, J., Brem, R. B., Whittle, J. & Kruglyak, L. Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet.* **1**, e25 (2005).
124. Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
125. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207 (2007).
126. Stranger, B. E. *et al.* Patterns of Cis Regulatory Variation in Diverse Human Populations. *PLoS Genet.* **8**, (2012).
127. Veyrieras, J.-B. *et al.* High-Resolution Mapping of Expression-QTLs Yields Insight into Human Gene Regulation. *PLoS Genet.* **4**, (2008).
128. Fehrmann, R. S. N. *et al.* Trans-eQTLs Reveal That Independent Genetic Variants Associated with a Complex Phenotype Converge on Intermediate Genes, with a Major Role for the HLA. *PLoS Genet* **7**, e1002197 (2011).
129. Fairfax, B. P. *et al.* GENETICS OF GENE EXPRESSION IN PRIMARY IMMUNE CELLS IDENTIFIES CELL-SPECIFIC MASTER REGULATORS AND ROLES OF HLA ALLELES. *Nat. Genet.* **44**, 502–510 (2012).
130. Zeller, T. *et al.* Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. *PLoS ONE* **5**, (2010).

131. Innocenti, F. *et al.* Identification, Replication, and Functional Fine-Mapping of Expression Quantitative Trait Loci in Primary Human Liver Tissue. *PLoS Genet.* **7**, (2011).
132. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
133. Li, Q. *et al.* A novel eQTL-based analysis reveals the biology of breast cancer risk loci. *Cell* **152**, 633–641 (2013).
134. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nat. Rev. Genet.* **7**, 862–872 (2006).
135. Lettice, L. A. *et al.* Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7548–7553 (2002).
136. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
137. Göring, H. H. H. *et al.* Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* **39**, 1208–1216 (2007).
138. Price, A. L. *et al.* Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals. *PLOS Genet.* **7**, e1001317 (2011).
139. Keurentjes, J. J. B. *et al.* Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci. *Proc. Natl. Acad. Sci.* **104**, 1708–1713 (2007).
140. Petretto, E. *et al.* Heritability and Tissue Specificity of Expression Quantitative Trait Loci. *PLOS Genet* **2**, e172 (2006).
141. Chesler, E. J. *et al.* Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* **37**, 233–242 (2005).
142. van Nas, A. *et al.* Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics* **185**, 1059–1068 (2010).

143. Kasowski, M. *et al.* Extensive Variation in Chromatin States Across Humans. *Science* **342**, 750–752 (2013).
144. Degner, J. F. *et al.* DNaseI sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).
145. Kasowski, M. *et al.* Variation in Transcription Factor Binding Among Humans. *Science* **328**, 232–235 (2010).
146. Schmidt, D. *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010).
147. Stefflova, K. *et al.* Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* **154**, 530–540 (2013).
148. Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
149. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
150. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* **338**, 1593–1599 (2012).
151. Pai, A. A. *et al.* The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet.* **8**, e1003000 (2012).
152. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
153. Yoon, O. K., Hsu, T. Y., Im, J. H. & Brem, R. B. Genetics and regulatory impact of alternative polyadenylation in human B-lymphoblastoid cells. *PLoS Genet.* **8**, e1002882 (2012).
154. Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* **13**, R7 (2012).

155. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L. M. & Snyder, M. Genetic Analysis of Variation in Transcription Factor Binding in Yeast. *Nature* **464**, 1187–1191 (2010).
156. Kilpinen, H. *et al.* Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription. *Science* **342**, 744–747 (2013).
157. McDaniell, R. *et al.* Heritable Individual-Specific and Allele-Specific Chromatin Signatures in Humans. *Science* **328**, 235–239 (2010).
158. McVicker, G. *et al.* Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749 (2013).
159. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* **24**, 14–24 (2014).
160. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
161. Lu, J. & Clark, A. G. Impact of microRNA regulation on variation in human gene expression. *Genome Res.* **22**, 1243–1254 (2012).
162. Wang, L. *et al.* Genome-wide transcriptional profiling reveals microRNA-correlated genes and biological processes in human lymphoblastoid cell lines. *PloS One* **4**, e5878 (2009).
163. Kumar, V. *et al.* Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. *PLOS Genet.* **9**, e1003201 (2013).
164. Popadin, K., Gutierrez-Arcelus, M., Dermitzakis, E. T. & Antonarakis, S. E. Genetic and epigenetic regulation of human lincRNA gene expression. *Am. J. Hum. Genet.* **93**, 1015–1026 (2013).
165. Breitling, R. *et al.* Genetical Genomics: Spotlight on QTL Hotspots. *PLoS Genet* **4**, e1000232 (2008).
166. Ghazalpour, A. *et al.* High-Resolution Mapping of Gene Expression Using Association in an Outbred Mouse Stock. *PLoS Genet.* **4**, (2008).

167. Mehrabian, M. *et al.* Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat. Genet.* **37**, 1224–1233 (2005).
168. Orozco, L. D. *et al.* Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell* **151**, 658–670 (2012).
169. Rockman, M. V., Skrovanek, S. S. & Kruglyak, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330**, 372–376 (2010).
170. West, M. A. L. *et al.* Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in *Arabidopsis*. *Genetics* **175**, 1441–1450 (2007).
171. Cheung, V. G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369 (2005).
172. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
173. Grieve, I. C. *et al.* Genome-wide co-expression analysis in multiple tissues. *PLoS One* **3**, e4033 (2008).
174. Kang, H. *et al.* Kcnn4 Is a Regulator of Macrophage Multinucleation in Bone Homeostasis and Inflammatory Disease. *Cell Rep.* **8**, 1210–1224 (2014).
175. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6**, e83 (2008).
176. Zhu, J. *et al.* Integrating Large-Scale Functional Genomic Data to Dissect the Complexity of Yeast Regulatory Networks. *Nat. Genet.* **40**, 854–861 (2008).
177. Fairfax, B. P. *et al.* Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. *Science* **343**, 1246949 (2014).
178. Pierce, B. L. *et al.* Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians. *PLoS Genet.* **10**, (2014).

179. Small, K. S. *et al.* Identification of an imprinted master trans-regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564 (2011).
180. Lee, M. N. *et al.* Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science* **343**, 1246980 (2014).
181. Kawasaki, A. *et al.* Association of IRF5 polymorphisms with systemic lupus erythematosus in a Japanese population: support for a crucial role of intron 1 polymorphisms. *Arthritis Rheum.* **58**, 826–834 (2008).
182. Rotival, M. *et al.* Integrating Genome-Wide Genetic Variations and Monocyte Expression Data Reveals Trans-Regulated Gene Modules in Humans. *PLoS Genet.* **7**, (2011).
183. Bystrykh, L. *et al.* Uncovering regulatory pathways that affect hematopoietic stem cell function using ‘genetical genomics’. *Nat. Genet.* **37**, 225–232 (2005).
184. Li, Y. *et al.* Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* **2**, e222 (2006).
185. Lan, H. *et al.* Combined expression trait correlations and expression quantitative trait locus mapping. *PLoS Genet.* **2**, e6 (2006).
186. DeCook, R., Lall, S., Nettleton, D. & Howell, S. H. Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* **172**, 1155–1164 (2006).
187. McClurg, P. *et al.* Genomewide association analysis in diverse inbred mice: power and population structure. *Genetics* **176**, 675–683 (2007).
188. Grieve, I. C. *et al.* Genome-wide co-expression analysis in multiple tissues. *PloS One* **3**, e4033 (2008).
189. Wu, C. *et al.* Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. *PLoS Genet.* **4**, e1000070 (2008).
190. Heinig, M. *et al.* A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* **467**, 460–464 (2010).

191. Wang, J. *et al.* An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population. *J. Exp. Bot.* ert464 (2014). doi:10.1093/jxb/ert464
192. Tian, J. *et al.* Identification of the Bile Acid Transporter Slco1a6 as a Candidate Gene That Broadly Affects Gene Expression in Mouse Pancreatic Islets. *Genetics* **201**, 1253–1262 (2015).
193. Quach, H. *et al.* Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **167**, 643–656.e17 (2016).
194. Hasin-Brumshtein, Y. *et al.* Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. *eLife* **5**,
195. Llamas, B., Bélanger, S., Picard, S. & Deschepper, C. F. Cardiac mass and cardiomyocyte size are governed by different genetic loci on either autosomes or chromosome Y in recombinant inbred mice. *Physiol. Genomics* **31**, 176–182 (2007).
196. Petretto, E. *et al.* New Insights into the Genetic Control of Gene Expression using a Bayesian Multi-tissue Approach. *PLoS Comput Biol* **6**, e1000737 (2010).
197. Scott-Boyer, M. P. *et al.* An Integrated Hierarchical Bayesian Model for Multivariate eQTL Mapping. *Stat. Appl. Genet. Mol. Biol.* **11**, (2012).
198. Imholte, G. C., Scott-Boyer, M.-P., Labbe, A., Deschepper, C. F. & Gottardo, R. iBMQ: a R/Bioconductor package for integrated Bayesian modeling of eQTL data. *Bioinformatics* **29**, 2797–2798 (2013).
199. Scott-Boyer, M.-P. & Deschepper, C. F. Genome-Wide Detection of Gene Coexpression Domains Showing Linkage to Regions Enriched with Polymorphic Retrotransposons in Recombinant Inbred Mouse Strains. *G3 GenesGenomesGenetics* **3**, 597–605 (2013).
200. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**, 241–247 (1995).

201. Stetson, D. B. & Medzhitov, R. Antiviral defense: interferons and beyond. *J. Exp. Med.* **203**, 1837–1841 (2006).
202. Theofilopoulos, A. N., Baccala, R., Beutler, B. & Kono, D. H. Type I interferons (α/β) in immunity and autoimmunity. *Annu. Rev. Immunol.* **23**, 307–336 (2005).
203. Kawai, T. & Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.* **11**, 373–384 (2010).
204. Barber, G. N. Cytoplasmic DNA innate immune pathways. *Immunol. Rev.* **243**, 99–108 (2011).
205. Iwasaki, A. A Virological View of Innate Immune Recognition. *Annu. Rev. Microbiol.* **66**, 177–196 (2012).
206. Hornung, V. *et al.* 5'-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994–997 (2006).
207. Pichlmair, A. *et al.* RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* **314**, 997–1001 (2006).
208. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.* **5**, 730–737 (2004).
209. Matsumiya, T. & Stafforini, D. M. Function and Regulation of Retinoic Acid-Inducible Gene-I. *Crit. Rev. Immunol.* **30**, 489–513 (2010).
210. O'Neill, L. A. J. & Bowie, A. G. Sensing and signaling in antiviral innate immunity. *Curr. Biol. CB* **20**, R328-333 (2010).
211. Saha, S. K. *et al.* Regulation of antiviral responses by a direct and specific interaction between TRAF3 and Cardif. *EMBO J.* **25**, 3257–3263 (2006).
212. Shi, Z. *et al.* Structural Insights into mitochondrial antiviral signaling protein (MAVS)-tumor necrosis factor receptor-associated factor 6 (TRAF6) signaling. *J. Biol. Chem.* **290**, 26811–26820 (2015).
213. Liu, S. *et al.* MAVS recruits multiple ubiquitin E3 ligases to activate antiviral signaling cascades. *eLife* **2**, e00785 (2013).

214. Shi, Y. *et al.* An autoinhibitory mechanism modulates MAVS activity in antiviral innate immune response. *Nat. Commun.* **6**, 7811 (2015).
215. Seth, R. B., Sun, L., Ea, C.-K. & Chen, Z. J. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell* **122**, 669–682 (2005).
216. Sun, L., Wu, J., Du, F., Chen, X. & Chen, Z. J. Cyclic GMP-AMP Synthase is a Cytosolic DNA Sensor that Activates the Type-I Interferon Pathway. *Science* **339**, (2013).
217. Ishikawa, H. & Barber, G. N. STING an Endoplasmic Reticulum Adaptor that Facilitates Innate Immune Signaling. *Nature* **455**, 674–678 (2008).
218. Wallach, D. & Kovalenko, A. How do cells sense foreign DNA? A new outlook on the function of STING. *Mol. Cell* **50**, 1–2 (2013).
219. Zemirli, N. & Arnoult, D. Mitochondrial anti-viral immunity. *Int. J. Biochem. Cell Biol.* **44**, 1473–1476 (2012).
220. Buskiewicz, I. A. *et al.* Reactive oxygen species induce virus-independent MAVS oligomerization in systemic lupus erythematosus. *Sci Signal* **9**, ra115-ra115 (2016).
221. Soucy-Faulkner, A. *et al.* Requirement of NOX2 and Reactive Oxygen Species for Efficient RIG-I-Mediated Antiviral Response through Regulation of MAVS Expression. *PLOS Pathog.* **6**, e1000930 (2010).
222. Tal, M. C. *et al.* Absence of autophagy results in reactive oxygen species-dependent amplification of RLR signaling. *Proc. Natl. Acad. Sci.* **106**, 2770–2775 (2009).
223. Kawai, T. & Akira, S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34**, 637–650 (2011).
224. Häcker, H. *et al.* Specificity in Toll-like receptor signalling through distinct effector functions of TRAF3 and TRAF6. *Nature* **439**, 204–207 (2006).

225. Oganessian, G. *et al.* Critical role of TRAF3 in the Toll-like receptor-dependent and -independent antiviral response. *Nature* **439**, 208–211 (2006).
226. Fitzgerald, K. A. *et al.* IKK ϵ and TBK1 are essential components of the IRF3 signaling pathway. *Nat. Immunol.* **4**, 491–496 (2003).
227. Sharma, S. *et al.* Triggering the Interferon Antiviral Response Through an IKK-Related Pathway. *Science* **300**, 1148–1151 (2003).
228. Schröder, M., Baran, M. & Bowie, A. G. Viral targeting of DEAD box protein 3 reveals its role in TBK1/IKK ϵ -mediated IRF activation. *EMBO J.* **27**, 2147–2157 (2008).
229. Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).
230. Sato, M. *et al.* Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- α / β gene induction. *Immunity* **13**, 539–548 (2000).
231. Tamura, T., Yanai, H., Savitsky, D. & Taniguchi, T. The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.* **26**, 535–584 (2008).
232. McWhirter, S. M. *et al.* IFN-regulatory factor 3-dependent gene expression is defective in Tbk1-deficient mouse embryonic fibroblasts. *Proc. Natl. Acad. Sci.* **101**, 233–238 (2004).
233. Perry, A. K., Chow, E. K., Goodnough, J. B., Yeh, W.-C. & Cheng, G. Differential Requirement for TANK-binding Kinase-1 in Type I Interferon Responses to Toll-like Receptor Activation and Viral Infection. *J. Exp. Med.* **199**, 1651–1658 (2004).
234. Hemmi, H. *et al.* The roles of two IkappaB kinase-related kinases in lipopolysaccharide and double stranded RNA signaling and viral infection. *J. Exp. Med.* **199**, 1641–1650 (2004).
235. Bonnard, M. *et al.* Deficiency of T2K leads to apoptotic liver degeneration and impaired NF- κ B-dependent gene transcription. *EMBO J.* **19**, 4976–4985 (2000).

236. Matsui, K. *et al.* Cutting edge: Role of TANK-binding kinase 1 and inducible I κ B kinase in IFN responses against viruses in innate immune cells. *J. Immunol. Baltim. Md 1950* **177**, 5785–5789 (2006).
237. Plataniias, L. C. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.* **5**, 375–386 (2005).
238. Schoggins, J. W. *et al.* A diverse array of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
239. Levy, D. E. & Darnell, J. E. Stats: transcriptional control and biological impact. *Nat. Rev. Mol. Cell Biol.* **3**, 651–662 (2002).
240. Stark, G. R. & Darnell, J. E. The JAK-STAT Pathway at Twenty. *Immunity* **36**, 503–514 (2012).
241. MacMicking, J. D. Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat. Rev. Immunol.* **12**, 367–382 (2012).
242. Rusinova, I. *et al.* INTERFEROME v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* **41**, D1040–D1046 (2013).
243. Murray, P. J. The JAK-STAT signaling pathway: input and output integration. *J. Immunol. Baltim. Md 1950* **178**, 2623–2629 (2007).
244. Gough, D. J., Messina, N. L., Clarke, C. J. P., Johnstone, R. W. & Levy, D. E. Constitutive type I interferon modulates homeostatic balance through tonic signaling. *Immunity* **36**, 166–174 (2012).
245. Ng, S.-L. *et al.* I κ B kinase epsilon (IKK(epsilon)) regulates the balance between type I and type II interferon responses. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 21170–21175 (2011).
246. Tenover, B. R. *et al.* Multiple functions of the IKK-related kinase IKKepsilon in interferon-mediated antiviral immunity. *Science* **315**, 1274–1278 (2007).
247. Yarilina, A., Park-Min, K.-H., Antoniv, T., Hu, X. & Ivashkiv, L. B. TNF activates an IRF1-dependent autocrine loop leading to sustained expression of chemokines and STAT1-dependent type I interferon-response genes. *Nat. Immunol.* **9**, 378–387 (2008).

248. van Boxel-Dezaire, A. H. H., Rani, M. R. S. & Stark, G. R. Complex modulation of cell type-specific signaling in response to type I interferons. *Immunity* **25**, 361–372 (2006).
249. Mancl, M. E. *et al.* Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. *J. Biol. Chem.* **280**, 21078–21090 (2005).
250. Bhattacharya, S. *et al.* Anti-tumorigenic effects of Type 1 interferon are subdued by integrated stress responses. *Oncogene* **32**, 4214–4221 (2013).
251. Fuchs, S. Y. Hope and Fear for Interferon: The Receptor-Centric Outlook on the Future of Interferon Therapy. *J. Interferon Cytokine Res.* **33**, 211–225 (2013).
252. HuangFu, W.-C. *et al.* Inflammatory signaling compromises cell responses to interferon alpha. *Oncogene* **31**, 161–172 (2012).
253. Huynh, L., Wang, L., Shi, C., Park-Min, K.-H. & Ivashkiv, L. B. ITAM-coupled receptors inhibit IFNAR signaling and alter macrophage responses to TLR4 and *Listeria monocytogenes*. *J. Immunol. Baltim. Md 1950* **188**, 3447–3457 (2012).
254. Yoshimura, A., Naka, T. & Kubo, M. SOCS proteins, cytokine signalling and immune regulation. *Nat. Rev. Immunol.* **7**, 454–465 (2007).
255. Sarasin-Filipowicz, M. *et al.* Alpha Interferon Induces Long-Lasting Refractoriness of JAK-STAT Signaling in the Mouse Liver through Induction of USP18/UBP43. *Mol. Cell. Biol.* **29**, 4841–4851 (2009).
256. Colina, R. *et al.* Translational control of the innate immune response through IRF-7. *Nature* **452**, 323–328 (2008).
257. Fensterl, V. & Sen, G. C. The ISG56/IFIT1 Gene Family. *J. Interferon Cytokine Res.* **31**, 71–78 (2011).

258. Terenzi, F., Hui, D. J., Merrick, W. C. & Sen, G. C. Distinct induction patterns and functions of two closely related interferon-inducible human genes, ISG54 and ISG56. *J. Biol. Chem.* **281**, 34064–34071 (2006).
259. Walsh, D., Mathews, M. B. & Mohr, I. Tinkering with Translation: Protein Synthesis in Virus-Infected Cells. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
260. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat. Rev. Immunol.* **13**, 46–57 (2013).
261. Goff, S. P. Retrovirus restriction factors. *Mol. Cell* **16**, 849–859 (2004).
262. Kok, K.-H. & Jin, D.-Y. Balance of power in host-virus arms races. *Cell Host Microbe* **14**, 5–6 (2013).
263. Munir, M. & Berg, M. The multiple faces of protein kinase R in antiviral defense. *Virulence* **4**, 85–89 (2013).
264. Zhao, C., Collins, M. N., Hsiang, T.-Y. & Krug, R. M. Interferon-induced ISG15 pathway: an ongoing virus-host battle. *Trends Microbiol.* **21**, 181–186 (2013).
265. Sgorbissa, A. & Brancolini, C. IFNs, ISGylation and cancer: Cui prodest? *Cytokine Growth Factor Rev.* **23**, 307–314 (2012).
266. Van Damme, N. *et al.* The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. *Cell Host Microbe* **3**, 245–252 (2008).
267. Neil, S. J. D., Zang, T. & Bieniasz, P. D. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* **451**, 425–430 (2008).
268. Kupzig, S. *et al.* Bst-2/HM1.24 is a raft-associated apical membrane protein with an unusual topology. *Traffic Cph. Den.* **4**, 694–709 (2003).

269. Jones, P. H., Mehta, H. V., Maric, M., Roller, R. J. & Okeoma, C. M. Bone marrow stromal cell antigen 2 (BST-2) restricts mouse mammary tumor virus (MMTV) replication in vivo. *Retrovirology* **9**, 10 (2012).
270. Sarojini, S., Theofanis, T. & Reiss, C. S. Interferon-induced tetherin restricts vesicular stomatitis virus release in neurons. *DNA Cell Biol.* **30**, 965–974 (2011).
271. Ooi, Y. S., Dubé, M. & Kielian, M. BST2/tetherin inhibition of alphavirus exit. *Viruses* **7**, 2147–2167 (2015).
272. Radoshitzky, S. R. *et al.* Infectious Lassa virus, but not filoviruses, is restricted by BST-2/tetherin. *J. Virol.* **84**, 10569–10580 (2010).
273. Galão, R. P., Le Tortorec, A., Pickering, S., Kueck, T. & Neil, S. J. D. Innate sensing of HIV-1 assembly by Tetherin induces NFκB-dependent proinflammatory responses. *Cell Host Microbe* **12**, 633–644 (2012).
274. Mahauad-Fernandez, W. D., Jones, P. H. & Okeoma, C. M. Critical role for bone marrow stromal antigen 2 in acute Chikungunya virus infection. *J. Gen. Virol.* **95**, 2450–2461 (2014).
275. Nasr, N. *et al.* HIV-1 infection of human macrophages directly induces viperin which inhibits viral production. *Blood* **120**, 778–788 (2012).
276. Helbig, K. J. *et al.* The antiviral protein viperin inhibits hepatitis C virus replication via interaction with nonstructural protein 5A. *Hepatology* **54**, 1506–1517 (2011).
277. Seo, J.-Y., Yaneva, R., Hinson, E. R. & Cresswell, P. Human cytomegalovirus directly induces the antiviral protein viperin to enhance infectivity. *Science* **332**, 1093–1097 (2011).
278. Helbig, K. J. *et al.* Viperin Is Induced following Dengue Virus Type-2 (DENV-2) Infection and Has Anti-viral Actions Requiring the C-terminal End of Viperin. *PLoS Negl. Trop. Dis.* **7**, e2178 (2013).
279. Tan, K. S. *et al.* In vivo and in vitro studies on the antiviral activities of viperin against influenza H1N1 virus infection. *J. Gen. Virol.* **93**, 1269–1277 (2012).

280. Szretter, K. J. *et al.* The interferon-inducible gene viperin restricts West Nile virus pathogenesis. *J. Virol.* **85**, 11557–11566 (2011).
281. Qiu, L.-Q., Cresswell, P. & Chin, K.-C. Viperin is required for optimal Th2 responses and T-cell receptor-mediated activation of NF- κ B and AP-1. *Blood* **113**, 3520–3529 (2009).
282. Saitoh, T. *et al.* Antiviral protein viperin promotes toll-like receptor 7- and toll-like receptor 9-mediated type I interferon production in plasmacytoid dendritic cells. *Immunity* **34**, 352–363 (2011).
283. Buss, C. *et al.* Essential role of mitochondrial antiviral signaling, IFN regulatory factor (IRF)3, and IRF7 in *Chlamydia pneumoniae*-mediated IFN- β response and control of bacterial replication in human endothelial cells. *J. Immunol. Baltim. Md 1950* **184**, 3072–3078 (2010).
284. Ishihara, T. *et al.* Inhibition of *Chlamydia trachomatis* growth by human interferon- α : mechanisms and synergistic effect with interferon- γ and tumor necrosis factor- α . *Biomed. Res. Tokyo Jpn.* **26**, 179–185 (2005).
285. de la Maza, L. M., Peterson, E. M., Goebel, J. M., Fennie, C. W. & Czarniecki, C. W. Interferon-induced inhibition of *Chlamydia trachomatis*: dissociation from antiviral and antiproliferative effects. *Infect. Immun.* **47**, 719–722 (1985).
286. Plumlee, C. R. *et al.* Interferons direct an effective innate response to *Legionella pneumophila* infection. *J. Biol. Chem.* **284**, 30058–30066 (2009).
287. Decker, T., Müller, M. & Stockinger, S. The yin and yang of type I interferon activity in bacterial infection. *Nat. Rev. Immunol.* **5**, 675–687 (2005).
288. Bekisz, J., Baron, S., Balinsky, C., Morrow, A. & Zoon, K. C. Antiproliferative Properties of Type I and Type II Interferon. *Pharmaceuticals* **3**, 994–1015 (2010).
289. Mattei, F., Schiavoni, G. & Tough, D. F. Regulation of immune cell homeostasis by type I interferons. *Cytokine Growth Factor Rev.* **21**, 227–236 (2010).
290. Trinchieri, G. Type I interferon: friend or foe? *J. Exp. Med.* **207**, 2053–2063 (2010).

291. Brinkworth, J. F. & Barreiro, L. B. The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. *Curr. Opin. Immunol.* **31**, 66–78 (2014).
292. Barrett, J. C. *et al.* Genome-wide association study and meta-analysis finds over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009).
293. Plagnol, V. *et al.* Genome-Wide Association Analysis of Autoantibody Positivity in Type 1 Diabetes Cases. *PLoS Genet.* **7**, (2011).
294. Zhernakova, A. *et al.* Meta-Analysis of Genome-Wide Association Studies in Celiac Disease and Rheumatoid Arthritis Identifies Fourteen Non-HLA Shared Loci. *PLoS Genet.* **7**, (2011).
295. Eriksson, N. *et al.* Novel Associations for Hypothyroidism Include Known Autoimmune Risk Loci. *PLoS ONE* **7**, (2012).
296. Jin, Y. *et al.* Genome-wide association analyses identify 13 new susceptibility loci for generalized vitiligo. *Nat. Genet.* **44**, 676–680 (2012).
297. Newton-Cheh, C. *et al.* Eight blood pressure loci identified by genome-wide association study of 34,433 people of European ancestry. *Nat. Genet.* **41**, 666–676 (2009).
298. Wain, L. V. *et al.* Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.* **43**, 1005–1011 (2011).
299. Köttgen, A. *et al.* Multiple New Loci Associated with Kidney Function and Chronic Kidney Disease: The CKDGen consortium. *Nat. Genet.* **42**, 376–384 (2010).
300. Rojo, L., Sjöberg, M. K., Hernández, P., Zambrano, C. & Maccioni, R. B. Roles of cholesterol and lipids in the etiopathogenesis of Alzheimer’s disease. *J. Biomed. Biotechnol.* **2006**, 73976 (2006).
301. Berliner, J. A., Leitinger, N. & Tsimikas, S. The role of oxidized phospholipids in atherosclerosis. *J. Lipid Res.* **50 Suppl**, S207-212 (2009).

302. Frangogiannis, N. G. Regulation of the inflammatory response in cardiac repair. *Circ. Res.* **110**, 159–173 (2012).
303. Brenner, M. & Gulko, P. S. The arthritis severity locus *Cia5a* regulates the expression of inflammatory mediators including Syk pathway genes and proteases in pristane-induced arthritis. *BMC Genomics* **13**, 710 (2012).
304. Visan, I. Sensing cholesterol dynamics. *Nat. Immunol.* **17**, 121–121 (2016).
305. O’Neill, L. A. J. How Low Cholesterol Is Good for Anti-viral Immunity. *Cell* **163**, 1572–1574 (2015).
306. Wong, W. Reprogramming lipid metabolism to boost antiviral responses. *Sci Signal* **9**, ec3–ec3 (2016).
307. Parkes, M., Cortes, A., van Heel, D. A. & Brown, M. A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–673 (2013).
308. Hosono, K., Sasaki, T., Minoshima, S. & Shimizu, N. Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species. *Gene* **340**, 31–43 (2004).
309. Roxström-Lindquist, K. & Faye, I. The *Drosophila* gene Yippee reveals a novel family of putative zinc binding proteins highly conserved among eukaryotes. *Insect Mol. Biol.* **10**, 77–86 (2001).
310. Jeidane, S. *et al.* Association of a Network of Interferon-Stimulated Genes with a Locus Encoding a Negative Regulator of Non-conventional IKK Kinases and IFNB1. *Cell Rep.* **17**, 425–435 (2016).
311. Nica, A. C. & Dermitzakis, E. T. Using gene expression to investigate the genetic basis of complex disorders. *Hum. Mol. Genet.* **17**, R129–134 (2008).
312. Pai, A. A., Pritchard, J. K. & Gilad, Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLoS Genet* **11**, e1004857 (2015).

313. Hertzog, P., Forster, S. & Samarajiwa, S. Systems biology of interferon responses. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **31**, 5–11 (2011).
314. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).
315. Bruns, A. M. & Horvath, C. M. Antiviral RNA recognition and assembly by RLR family innate immune sensors. *Cytokine Growth Factor Rev.* **25**, 507–512 (2014).
316. Cui, S. *et al.* The C-Terminal Regulatory Domain Is the RNA 5'-Triphosphate Sensor of RIG-I. *Mol. Cell* **29**, 169–179 (2008).
317. Baril, M. *et al.* Genome-wide RNAi Screen Reveals a New Role of a WNT/CTNNB1 Signaling Pathway as Negative Regulator of Virus-induced Innate Immune Responses. *PLoS Pathog* **9**, e1003416 (2013).
318. Rehwinkel, J. *et al.* RIG-I Detects Viral Genomic RNA during Negative-Strand RNA Virus Infection. *Cell* **140**, 397–408 (2010).
319. Yoo, J.-S., Kato, H. & Fujita, T. Sensing viral invasion by RIG-I like receptors. *Curr. Opin. Microbiol.* **20**, 131–138 (2014).
320. Wang, Y. *et al.* Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nat. Struct. Mol. Biol.* **17**, 781–787 (2010).
321. Park, S.-J., Lee, A.-N. & Youn, H.-S. TBK1-targeted suppression of TRIF-dependent signaling pathway of toll-like receptor 3 by auranofin. *Arch. Pharm. Res.* **33**, 939–945 (2010).
322. Shen, R. R. & Hahn, W. C. Emerging roles for the non-canonical IKKs in cancer. *Oncogene* **30**, 631–641 (2011).
323. Kang, H. M., Ye, C. & Eskin, E. Accurate Discovery of Expression Quantitative Trait Loci Under Confounding From Spurious and Genuine Regulatory Hotspots. *Genetics* **180**, 1909–1925 (2008).

324. Hosono, K., Sasaki, T., Minoshima, S. & Shimizu, N. Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species. *Gene* **340**, 31–43 (2004).
325. Hosono, K. *et al.* YPEL5 protein of the YPEL gene family is involved in the cell cycle progression by interacting with two distinct proteins RanBPM and RanBP10. *Genomics* **96**, 102–111 (2010).
326. Komuro, A. & Horvath, C. M. RNA- and Virus-Independent Inhibition of Antiviral Signaling by RNA Helicase LGP2. *J. Virol.* **80**, 12332–12342 (2006).
327. Gu, L., Fullam, A., Brennan, R. & Schröder, M. Human DEAD Box Helicase 3 Couples I κ B Kinase ϵ to Interferon Regulatory Factor 3 Activation. *Mol. Cell. Biol.* **33**, 2004–2015 (2013).
328. Tremblay, N. *et al.* Spliceosome SNRNP200 Promotes Viral RNA Sensing and IRF3 Activation of Antiviral Response. *PLoS Pathog.* **12**, e1005772 (2016).
329. Hutti, J. E. *et al.* Development of a High-Throughput Assay for Identifying Inhibitors of TBK1 and IKK ϵ . *PLoS ONE* **7**, e41494 (2012).
330. Ma, X. *et al.* Molecular basis of Tank-binding kinase 1 activation by transautophosphorylation. *Proc. Natl. Acad. Sci.* **109**, 9378–9383 (2012).
331. Ng, S.-L. *et al.* I κ B kinase ϵ (IKK ϵ) regulates the balance between type I and type II interferon responses. *Proc. Natl. Acad. Sci.* **108**, 21170–21175 (2011).
332. Liu, S. *et al.* Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* **347**, aaa2630 (2015).
333. Jiang, D.-S. *et al.* Role of Interferon Regulatory Factor 4 in the Regulation of Pathological Cardiac Hypertrophy. *Hypertension* **61**, 1193–1202 (2013).
334. Jiang, D.-S. *et al.* Interferon Regulatory Factor 1 Is Required for Cardiac Remodeling in Response to Pressure Overload. *Hypertension* **64**, 77–86 (2014).
335. Tsushima, K. *et al.* IRF3 regulates cardiac fibrosis but not hypertrophy in mice during angiotensin II-induced hypertension. *FASEB J.* **25**, 1531–1543 (2011).

336. Dai, J. *et al.* IKKi Deficiency Promotes Pressure Overload-Induced Cardiac Hypertrophy and Fibrosis. *PLoS ONE* **8**, e53412 (2013).
337. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
338. Boehm, J. S. *et al.* Integrative Genomic Approaches Identify IKBKE as a Breast Cancer Oncogene. *Cell* **129**, 1065–1079 (2007).
339. Guo, J.-P. *et al.* Deregulation of IKBKE Is Associated with Tumor Progression, Poor Prognosis, and Cisplatin Resistance in Ovarian Cancer. *Am. J. Pathol.* **175**, 324–333 (2009).
340. Verhelst, K., Verstrepen, L., Carpentier, I. & Beyaert, R. IκB kinase ε (IKKε): a therapeutic target in inflammation and cancer. *Biochem. Pharmacol.* **85**, 873–880 (2013).
341. Jiang, Z., Liu, J. C., Chung, P. E. D., Egan, S. E. & Zacksenhaus, E. Targeting HER2+ breast cancer: the TBK1/IKKε axis. *Oncoscience* **1**, 180–182 (2014).
342. Terenius, O. Hemolin-A lepidopteran anti-viral defense factor? *Dev. Comp. Immunol.* **32**, 311–316 (2008).
343. Scott-Boyer, M.-P. & Deschepper, C. F. Genome-Wide Detection of Gene Coexpression Domains Showing Linkage to Regions Enriched with Polymorphic Retrotransposons in Recombinant Inbred Mouse Strains. *G3 GenesGenomesGenetics* **3**, 597–605 (2013).
344. Baril, M., Racine, M.-E., Penin, F. & Lamarre, D. MAVS Dimer Is a Crucial Signaling Component of Innate Immunity and the Target of Hepatitis C Virus NS3/4A Protease. *J. Virol.* **83**, 1299–1311 (2009).
345. Es-Saad, S., Tremblay, N., Baril, M. & Lamarre, D. Regulators of innate immunity as novel targets for panviral therapeutics. *Curr. Opin. Virol.* **2**, 622–628 (2012).
346. Fox, R. J. & Dimmic, M. W. A two-sample Bayesian t-test for microarray data. *BMC Bioinformatics* **7**, 126 (2006).

347. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
348. Esposito, A., Gugliucci, G. & Del Puente, A. Epidemiology and nutritional risk factors. *Aging Clin. Exp. Res.* **16 Suppl**, 10–14 (2004).
349. Hardy, K. *et al.* Transcriptional Networks and Cellular Senescence in Human Mammary Fibroblasts. *Mol. Biol. Cell* **16**, 943–953 (2005).
350. Brady, C. A. *et al.* Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. *Cell* **145**, 571–583 (2011).
351. Post, S. M. *et al.* p53-dependent senescence delays Emu-myc-induced B-cell lymphomagenesis. *Oncogene* **29**, 1260–1269 (2010).
352. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
353. Gu, Y., Turck, C. W. & Morgan, D. O. Inhibition of CDK2 activity in vivo by an associated 20K regulatory subunit. *Nature* **366**, 707–710 (1993).
354. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
355. Xiong, Y. *et al.* p21 is a universal inhibitor of cyclin kinases. *Nature* **366**, 701–704 (1993).
356. Adams, P. D. Healing and hurting: molecular mechanisms, functions, and pathologies of cellular senescence. *Mol. Cell* **36**, 2–14 (2009).
357. Ben-Porath, I. & Weinberg, R. A. The signals and pathways activating cellular senescence. *Int. J. Biochem. Cell Biol.* **37**, 961–976 (2005).
358. Nevins, J. R. Toward an understanding of the functional complexity of the E2F and retinoblastoma families. *Cell Growth Differ. Mol. Biol. J. Am. Assoc. Cancer Res.* **9**, 585–593 (1998).

359. Niculescu, A. B. *et al.* Effects of p21Cip1/Waf1 at Both the G1/S and the G2/M Cell Cycle Transitions: pRb Is a Critical Determinant in Blocking DNA Replication and in Preventing Endoreduplication. *Mol. Cell. Biol.* **18**, 629–643 (1998).
360. de Toledo, S., Azzam, E., Keng, P., Laffrenier, S. & Little, J. Regulation by ionizing radiation of CDC2, cyclin A, cyclin B, thymidine kinase, topoisomerase IIalpha, and RAD51 expression in normal human diploid fibroblasts is dependent on p53/p21Waf1. *Cell Growth Differ.* **9**, 887–896 (1998).
361. Abbas, T. & Dutta, A. p21 in cancer: intricate networks and multiple activities. *Nat. Rev. Cancer* **9**, 400–414 (2009).
362. Kelley, K. D. *et al.* YPEL3, a p53-Regulated Gene that Induces Cellular Senescence. *Cancer Res.* **70**, 3566–3575 (2010).
363. Berberich, S. J., Todd, A. & Tuttle, R. Why YPEL3 represents a novel tumor suppressor. *Front. Biosci. Landmark Ed.* **16**, 1746–1751 (2011).
364. Berberich, S. RNAi knock of HdmX or Hdm2 leads to new insights into p53 signaling. *Cell Cycle* **9**, 3640–3641 (2010).
365. Heminger, K., Markey, M., Mpagi, M. & Berberich, S. J. Alterations in gene expression and sensitivity to genotoxic stress following HdmX or Hdm2 knockdown in human tumor cells harboring wild-type p53. *Aging* **1**, 89–108 (2009).
366. Berberich, S. J., Todd, A. & Tuttle, R. Why YPEL3 represents a novel tumor suppressor. *Front. Biosci. Landmark Ed.* **16**, 1746–1751 (2011).
367. Harris, C. C. & Hollstein, M. Clinical implications of the p53 tumor-suppressor gene. *N. Engl. J. Med.* **329**, 1318–1327 (1993).
368. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).

369. Tuttle, R. *et al.* Novel senescence associated gene, YPEL3, is repressed by estrogen in ER+ mammary tumor cells and required for tamoxifen-induced cellular senescence. *Int. J. Cancer* **130**, 2291–2299 (2012).
370. Tuttle, R. *et al.* Senescence-associated gene YPEL3 is downregulated in human colon tumors. *Ann. Surg. Oncol.* **18**, 1791–1796 (2011).
371. Li, J. *et al.* Selective TBK1/IKKi dual inhibitors with anticancer potency. *Int. J. Cancer* **134**, 1972–1980 (2014).
372. Jeidane, S. *et al.* Association of a Network of Interferon-Stimulated Genes with a Locus Encoding a Negative Regulator of Non-conventional IKK Kinases and IFNB1. *Cell Rep.* **17**, 425–435 (2016).
373. Srsen, V., Gnadt, N., Dammermann, A. & Merdes, A. Inhibition of centrosome protein assembly leads to p53-dependent exit from the cell cycle. *J. Cell Biol.* **174**, 625–630 (2006).
374. Mikule, K. *et al.* Loss of centrosome integrity induces p38-p53-p21-dependent G1-S arrest. *Nat. Cell Biol.* **9**, 160–170 (2007).
375. Schmidt, S. *et al.* The centrosome and mitotic spindle apparatus in cancer and senescence. *Cell Cycle* **9**, 4469–4473 (2010).
376. Wong, Y. L. *et al.* Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science* **348**, 1155–1160 (2015).
377. Muñoz-Espín, D. & Serrano, M. Cellular senescence: from physiology to pathology. *Nat. Rev. Mol. Cell Biol.* **15**, 482–496 (2014).
378. Waghray, M. *et al.* Hydrogen peroxide is a diffusible paracrine signal for the induction of epithelial cell death by activated myofibroblasts. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **19**, 854–856 (2005).
379. Lerit, D. A. & Poulton, J. S. Centrosomes are multifunctional regulators of genome stability. *Chromosome Res.* **24**, 5–17 (2015).

380. Badano, J. L., Teslovich, T. M. & Katsanis, N. The centrosome in human genetic disease. *Nat. Rev. Genet.* **6**, 194–205 (2005).
381. Olson, J. E. *et al.* Centrosome-related genes, genetic variation, and risk of breast cancer. *Breast Cancer Res. Treat.* **125**, 221–228 (2011).
382. Farlie, P. *et al.* Ypel1: a novel nuclear protein that induces an epithelial-like morphology in fibroblasts. *Genes Cells Devoted Mol. Cell. Mech.* **6**, 619–629 (2001).
383. Abiatari, I. *et al.* Consensus transcriptome signature of perineural invasion in pancreatic carcinoma. *Mol. Cancer Ther.* **8**, 1494–1504 (2009).
384. Abiatari, I., Kiladze, M., Kerkadze, V., Friess, H. & Kleeff, J. Expression of YPEL1 in pancreatic cancer cell lines and tissues. *Georgian Med. News* 60–62 (2009).
385. Wu, G. *et al.* Structural analysis of the 17q22-23 amplicon identifies several independent targets of amplification in breast cancer cell lines and tumors. *Cancer Res.* **61**, 4951–4955 (2001).
386. Kelemen, L. E. *et al.* Genetic variation in the chromosome 17q23 amplicon and breast cancer risk. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **18**, 1864–1868 (2009).
387. Oki, K., Plonczynski, M. W., Gomez-Sanchez, E. P. & Gomez-Sanchez, C. E. YPEL4 modulates HAC15 adrenal cell proliferation and is associated with tumor diameter. *Mol. Cell. Endocrinol.* **434**, 93–98 (2016).
388. Alimonti, A. *et al.* A novel type of cellular senescence that can be enhanced in mouse models and human tumor xenografts to suppress prostate tumorigenesis. *J. Clin. Invest.* **120**, 681–693 (2010).
389. Sun, P. *et al.* PRAK is essential for ras-induced senescence and tumor suppression. *Cell* **128**, 295–308 (2007).
390. Ewald, J. A., Desotelle, J. A., Wilding, G. & Jarrard, D. F. Therapy-induced senescence in cancer. *J. Natl. Cancer Inst.* **102**, 1536–1546 (2010).
391. Roninson, I. B. Tumor cell senescence in cancer treatment. *Cancer Res.* **63**, 2705–2715 (2003).

392. Ben-Porath, I. & Weinberg, R. A. When cells get stressed: an integrative view of cellular senescence. *J. Clin. Invest.* **113**, 8–13 (2004).
393. Dasari, A., Bartholomew, J. N., Volonte, D. & Galbiati, F. Oxidative Stress Induces Premature Senescence by Stimulating Caveolin-1 Gene Transcription through p38 Mitogen-Activated Protein Kinase/Sp1–Mediated Activation of Two GC-Rich Promoter Elements. *Cancer Res.* **66**, 10805–10814 (2006).
394. Suzuki, M. & Boothman, D. A. Stress-induced premature senescence (SIPS)--influence of SIPS on radiotherapy. *J. Radiat. Res. (Tokyo)* **49**, 105–112 (2008).
395. Ganie, S. A., Haq, E., Hamid, A., Masood, A. & Zargar, M. A. Long dose exposure of hydrogen peroxide (H₂O₂) in albino rats and effect of Podophyllum hexandrum on oxidative stress. *Eur. Rev. Med. Pharmacol. Sci.* **15**, 906–915 (2011).
396. Hwang, S.-L. & Yen, G.-C. Neuroprotective effects of the citrus flavanones against H₂O₂-induced cytotoxicity in PC12 cells. *J. Agric. Food Chem.* **56**, 859–864 (2008).
397. Storz, G. & Imlay, J. A. Oxidative stress. *Curr. Opin. Microbiol.* **2**, 188–194 (1999).
398. Davies, K. J. Oxidative stress, antioxidant defenses, and damage removal, repair, and replacement systems. *IUBMB Life* **50**, 279–289 (2000).
399. Davies, K. J. A. The Broad Spectrum of Responses to Oxidants in Proliferating Cells: A New Paradigm for Oxidative Stress. *IUBMB Life* **48**, 41–47 (1999).
400. Mahalingaiah, P. K. S. & Singh, K. P. Chronic Oxidative Stress Increases Growth and Tumorigenic Potential of MCF-7 Breast Cancer Cells. *PLOS ONE* **9**, e87371 (2014).
401. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
402. Servick, K. Breast Cancer: A World of Differences. *Science* **343**, 1452–1453 (2014).
403. Bhat, H. K., Calaf, G., Hei, T. K., Loya, T. & Vadgama, J. V. Critical role of oxidative stress in estrogen-induced carcinogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3913–3918 (2003).

404. Patel, M. M. & Bhat, H. K. Differential oxidant potential of carcinogenic and weakly carcinogenic estrogens: Involvement of metabolic activation and cytochrome P450. *J. Biochem. Mol. Toxicol.* **18**, 37–42 (2004).
405. Brown, N. S. & Bicknell, R. Hypoxia and oxidative stress in breast cancer: Oxidative stress: its effects on the growth, metastatic potential and response to therapy of breast cancer. *Breast Cancer Res. BCR* **3**, 323 (2001).
406. Matsui, A. *et al.* Increased formation of oxidative DNA damage, 8-hydroxy-2'-deoxyguanosine, in human breast cancer tissue and its relationship to GSTP1 and COMT genotypes. *Cancer Lett.* **151**, 87–95 (2000).
407. Zhao, M., Wi, H., Oh, T. I. & Woo, E. J. Evaluation of anomaly detection algorithm using trans-admittance mammography with 60×60 electrode array. *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.* **2013**, 6433–6436 (2013).
408. Babior, B. M. NADPH oxidase: an update. *Blood* **93**, 1464–1476 (1999).
409. Szatrowski, T. P. & Nathan, C. F. Production of large amounts of hydrogen peroxide by human tumor cells. *Cancer Res.* **51**, 794–798 (1991).
410. Dolado, I. *et al.* p38alpha MAP kinase as a sensor of reactive oxygen species in tumorigenesis. *Cancer Cell* **11**, 191–205 (2007).
411. Storz, P. Reactive oxygen species in tumor progression. *Front. Biosci. J. Virtual Libr.* **10**, 1881–1896 (2005).
412. Burdon, R. H., Gill, V. & Rice-Evans, C. Oxidative stress and tumour cell proliferation. *Free Radic. Res. Commun.* **11**, 65–76 (1990).
413. Valko, M., Rhodes, C. J., Moncol, J., Izakovic, M. & Mazur, M. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem. Biol. Interact.* **160**, 1–40 (2006).
414. Alexandre, J. *et al.* Accumulation of hydrogen peroxide is an early and crucial step for paclitaxel-induced cancer cell death both in vitro and in vivo. *Int. J. Cancer* **119**, 41–48 (2006).

415. Llobet, D. *et al.* Antioxidants block proteasome inhibitor function in endometrial carcinoma cells. *Anticancer. Drugs* **19**, 115–124 (2008).
416. Gouazé, V. *et al.* Glutathione Peroxidase-1 Overexpression Prevents Ceramide Production and Partially Inhibits Apoptosis in Doxorubicin-Treated Human Breast Carcinoma Cells. *Mol. Pharmacol.* **60**, 488–496 (2001).
417. Sinha, B. K., Mimnaugh, E. G., Rajagopalan, S. & Myers, C. E. Adriamycin Activation and Oxygen Free Radical Formation in Human Breast Tumor Cells: Protective Role of Glutathione Peroxidase in Adriamycin Resistance. *Cancer Res.* **49**, 3844–3848 (1989).
418. Lee, Y.-H., Kang, B. S. & Bae, Y.-S. Premature senescence in human breast cancer and colon cancer cells by tamoxifen-mediated reactive oxygen species generation. *Life Sci.* **97**, 116–122 (2014).
419. Nica, A. C. & Dermitzakis, E. T. Using gene expression to investigate the genetic basis of complex disorders. *Hum. Mol. Genet.* **17**, R129–R134 (2008).
420. Yang, X. Use of functional genomics to identify candidate genes underlying human genetic association studies of vascular diseases. *Arterioscler. Thromb. Vasc. Biol.* **32**, 216–222 (2012).
421. Nicolae, D. L. *et al.* Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLOS Genet* **6**, e1000888 (2010).
422. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
423. Pravenec, M. *et al.* Identification of renal Cd36 as a determinant of blood pressure and risk for hypertension. *Nat. Genet.* **40**, 952–954 (2008).
424. Deschepper, C. F. *et al.* Functional alterations of the Nppa promoter are linked to cardiac ventricular hypertrophy in WKY/WKHA rat crosses. *Circ. Res.* **88**, 223–228 (2001).
425. Petretto, E. *et al.* Integrated genomic approaches implicate osteoglycin (Ogn) in the regulation of left ventricular mass. *Nat. Genet.* **40**, 546–552 (2008).

426. Meng, H. *et al.* Identification of Abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4530–4535 (2007).
427. Rockman, M. V., Skrovaneck, S. S. & Kruglyak, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330**, 372–376 (2010).
428. Jansen, R. C., Nap, J.-P. & Mlynárová, L. Errors in genomics and proteomics. *Nat. Biotechnol.* **20**, 19 (2002).
429. Kishore, N. *et al.* IKK-i and TBK-1 are Enzymatically Distinct from the Homologous Enzyme IKK-2 COMPARATIVE ANALYSIS OF RECOMBINANT HUMAN IKK-i, TBK-1, AND IKK-2. *J. Biol. Chem.* **277**, 13840–13847 (2002).
430. Pomerantz, J. L. & Baltimore, D. NF- κ B activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* **18**, 6694–6704 (1999).
431. Tojima, Y. *et al.* NAK is an IkappaB kinase-activating kinase. *Nature* **404**, 778–782 (2000).
432. Shimada, T. *et al.* IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases. *Int. Immunol.* **11**, 1357–1362 (1999).
433. Bibeau-Poirier, A. *et al.* Involvement of the IkappaB kinase (IKK)-related kinases tank-binding kinase 1/IKKi and cullin-based ubiquitin ligases in IFN regulatory factor-3 degradation. *J. Immunol. Baltim. Md 1950* **177**, 5059–5067 (2006).
434. Gravel, S.-P. & Servant, M. J. Roles of an IkappaB kinase-related pathway in human cytomegalovirus-infected vascular smooth muscle cells: a molecular link in pathogen-induced proatherosclerotic conditions. *J. Biol. Chem.* **280**, 7477–7486 (2005).
435. Kravchenko, V. V., Mathison, J. C., Schwamborn, K., Mercurio, F. & Ulevitch, R. J. IKKi/IKKepsilon plays a key role in integrating signals induced by pro-inflammatory stimuli. *J. Biol. Chem.* **278**, 26612–26619 (2003).

436. Wang, N., Ahmed, S. & Haqqi, T. M. Genomic structure and functional characterization of the promoter region of human IkappaB kinase-related kinase IKKi/IKKvarepsilon gene. *Gene* **353**, 118–133 (2005).
437. Ma, X. *et al.* Molecular basis of Tank-binding kinase 1 activation by transautophosphorylation. *Proc. Natl. Acad. Sci.* **109**, 9378–9383 (2012).
438. Lei, C.-Q. *et al.* Glycogen Synthase Kinase 3 β Regulates IRF3 Transcription Factor-Mediated Antiviral Response via Activation of the Kinase TBK1. *Immunity* **33**, 878–889 (2010).
439. Soulat, D. *et al.* The DEAD-box helicase DDX3X is a critical component of the TANK-binding kinase 1-dependent innate immune response. *EMBO J.* **27**, 2135–2146 (2008).
440. Clark, K. *et al.* Novel cross-talk within the IKK family controls innate immunity. *Biochem. J.* **434**, 93–104 (2011).
441. Gabhann, J. N. *et al.* Absence of SHIP-1 Results in Constitutive Phosphorylation of Tank-Binding Kinase 1 and Enhanced TLR3-Dependent IFN- β Production. *J. Immunol.* **184**, 2314–2320 (2010).
442. Zhao, Y. *et al.* PPM1B negatively regulates antiviral response via dephosphorylating TBK1. *Cell. Signal.* **24**, 2197–2204 (2012).
443. An, H. *et al.* SHP-2 Phosphatase Negatively Regulates the TRIF Adaptor Protein-Dependent Type I Interferon and Proinflammatory Cytokine Production. *Immunity* **25**, 919–928 (2006).
444. Li, S., Wang, L., Berman, M., Kong, Y.-Y. & Dorf, M. E. Mapping a Dynamic Innate Immunity Protein Interaction Network Regulating Type I Interferon Production. *Immunity* **35**, 426–440 (2011).
445. Mizushima, N., Yoshimori, T. & Ohsumi, Y. The Role of Atg Proteins in Autophagosome Formation. *Annu. Rev. Cell Dev. Biol.* **27**, 107–132 (2011).
446. Wang, C. *et al.* The E3 ubiquitin ligase Nrdp1 ‘preferentially’ promotes TLR-mediated production of type I interferon. *Nat. Immunol.* **10**, 744–752 (2009).

447. Friedman, C. S. *et al.* The tumour suppressor CYLD is a negative regulator of RIG-I-mediated antiviral response. *EMBO Rep.* **9**, 930–936 (2008).
448. Zhang, M. *et al.* Regulation of I κ B Kinase-related Kinases and Antiviral Responses by Tumor Suppressor CYLD. *J. Biol. Chem.* **283**, 18621–18626 (2008).
449. Gao, L. *et al.* ABIN1 Protein Cooperates with TAX1BP1 and A20 Proteins to Inhibit Antiviral Signaling. *J. Biol. Chem.* **286**, 36592–36602 (2011).
450. Parvatiyar, K., Barber, G. N. & Harhaj, E. W. TAX1BP1 and A20 Inhibit Antiviral Signaling by Targeting TBK1-IKKi Kinases. *J. Biol. Chem.* **285**, 14999–15009 (2010).
451. Ng, M.-H. J. *et al.* MIP-T3 is a negative regulator of innate type I IFN response. *J. Immunol.* **187**, 6473–6482 (2011).
452. Li, Y. *et al.* ISG56 is a negative-feedback regulator of virus-triggered signaling and cellular antiviral response. *Proc. Natl. Acad. Sci.* **106**, 7945–7950 (2009).
453. Huang, J. *et al.* SIKE is an IKK ϵ /TBK1-associated suppressor of TLR3- and virus-triggered IRF-3 activation pathways. *EMBO J.* **24**, 4018–4028 (2005).
454. Peters, R. T., Liao, S. M. & Maniatis, T. IKKepsilon is part of a novel PMA-inducible I κ appaB kinase complex. *Mol. Cell* **5**, 513–522 (2000).
455. Pomerantz, J. L. & Baltimore, D. NF-kappaB activation by a signaling complex containing TRAF2, TANK and TBK1, a novel IKK-related kinase. *EMBO J.* **18**, 6694–6704 (1999).
456. Clément, J.-F., Meloche, S. & Servant, M. J. The IKK-related kinases: from innate immunity to oncogenesis. *Cell Res.* **18**, 889–899 (2008).
457. Boehm, J. S. *et al.* Integrative genomic approaches identify IKBKE as a breast cancer oncogene. *Cell* **129**, 1065–1079 (2007).
458. Eddy, S. F. *et al.* Inducible I κ appaB kinase/I κ appaB kinase epsilon expression is induced by CK2 and promotes aberrant nuclear factor-kappaB activation in breast cancer cells. *Cancer Res.* **65**, 11375–11383 (2005).

459. Buss, H. *et al.* Constitutive and interleukin-1-inducible phosphorylation of p65 NF- κ B at serine 536 is mediated by multiple protein kinases including I κ B kinase (IKK)- α , IKK β , IKK ϵ , TRAF family member-associated (TANK)-binding kinase 1 (TBK1), and an unknown kinase and couples p65 to TATA-binding protein-associated factor II31-mediated interleukin-8 transcription. *J. Biol. Chem.* **279**, 55633–55643 (2004).
460. Chen, L.-F. *et al.* NF- κ B RelA phosphorylation regulates RelA acetylation. *Mol. Cell. Biol.* **25**, 7966–7975 (2005).
461. Fujita, F. *et al.* Identification of NAP1, a regulatory subunit of I κ B kinase-related kinases that potentiates NF- κ B signaling. *Mol. Cell. Biol.* **23**, 7780–7793 (2003).
462. Lawrence, T., Bebién, M., Liu, G. Y., Nizet, V. & Karin, M. IKK α limits macrophage NF- κ B activation and contributes to the resolution of inflammation. *Nature* **434**, 1138–1143 (2005).
463. Mattioli, I. *et al.* Inducible phosphorylation of NF- κ B p65 at serine 468 by T cell costimulation is mediated by IKK ϵ . *J. Biol. Chem.* **281**, 6175–6183 (2006).
464. Nomura, F., Kawai, T., Nakanishi, K. & Akira, S. NF- κ B activation through IKK-i-dependent I-TRAF/TANK phosphorylation. *Genes Cells Devoted Mol. Cell. Mech.* **5**, 191–202 (2000).
465. Delhase, M., Hayakawa, M., Chen, Y. & Karin, M. Positive and negative regulation of I κ B kinase activity through IKK β subunit phosphorylation. *Science* **284**, 309–313 (1999).
466. Jin, J. *et al.* The kinase TBK1 controls IgA class switching by negatively regulating noncanonical NF- κ B signaling. *Nat. Immunol.* **13**, 1101–1109 (2012).
467. Guo, J.-P. *et al.* I κ BKE phosphorylation and inhibition of FOXO3a: a mechanism of I κ BKE oncogenic function. *PloS One* **8**, e63636 (2013).
468. Xie, X. *et al.* I κ B kinase ϵ and TANK-binding kinase 1 activate AKT by direct phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6474–6479 (2011).

469. Ou, Y.-H. *et al.* TBK1 directly engages Akt/PKB survival signaling to support oncogenic transformation. *Mol. Cell* **41**, 458–470 (2011).
470. Jia, T. *et al.* The differential susceptibilities of MCF-7 and MDA-MB-231 cells to the cytotoxic effects of curcumin are associated with the PI3K/Akt-SKP2-Cip/Kips pathway. *Cancer Cell Int.* **14**, 126 (2014).
471. Li, Y., Dowbenko, D. & Lasky, L. A. AKT/PKB phosphorylation of p21Cip/WAF1 enhances protein stability of p21Cip/WAF1 and promotes cell survival. *J. Biol. Chem.* **277**, 11352–11361 (2002).
472. Mitsuuchi, Y. *et al.* The phosphatidylinositol 3-kinase/AKT signal transduction pathway plays a critical role in the expression of p21WAF1/CIP1/SDI1 induced by cisplatin and paclitaxel. *Cancer Res.* **60**, 5390–5394 (2000).
473. Rössig, L. *et al.* Akt-dependent phosphorylation of p21(Cip1) regulates PCNA binding and proliferation of endothelial cells. *Mol. Cell. Biol.* **21**, 5644–5657 (2001).
474. Zhou, B. P. *et al.* Cytoplasmic localization of p21Cip1/WAF1 by Akt-induced phosphorylation in HER-2/neu-overexpressing cells. *Nat. Cell Biol.* **3**, 245–252 (2001).
475. Kim, J.-Y. *et al.* Dissection of TBK1 signaling via phosphoproteomics in lung cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12414–12419 (2013).
476. Guo, J.-P. *et al.* IKKepsilon phosphorylation of estrogen receptor alpha Ser-167 and contribution to tamoxifen resistance in breast cancer. *J. Biol. Chem.* **285**, 3676–3684 (2010).
477. Wei, C. *et al.* Elevated expression of TANK-binding kinase 1 enhances tamoxifen resistance in breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E601–E610 (2014).
478. Pillai, S. *et al.* Tank binding kinase 1 is a centrosome-associated kinase necessary for microtubule dynamics and mitosis. *Nat. Commun.* **6**, 10072 (2015).

479. Adli, M. & Baldwin, A. S. IKK-i/IKKepsilon controls constitutive, cancer cell-associated NF-kappaB activity via regulation of Ser-536 p65/RelA phosphorylation. *J. Biol. Chem.* **281**, 26976–26984 (2006).
480. Hutti, J. E. *et al.* Phosphorylation of the Tumor Suppressor CYLD by the Breast Cancer Oncogene IKKε Promotes Cell Transformation. *Mol. Cell* **34**, 461–472 (2009).
481. Shen, R. R. *et al.* IκB kinase ε phosphorylates TRAF2 to promote mammary epithelial cell transformation. *Mol. Cell. Biol.* **32**, 4756–4768 (2012).
482. Renner, F., Moreno, R. & Schmitz, M. L. SUMOylation-dependent localization of IKKepsilon in PML nuclear bodies is essential for protection against DNA-damage-triggered cell death. *Mol. Cell* **37**, 503–515 (2010).
483. Biswas, D. K. *et al.* NF-kappa B activation in human breast cancer specimens and its role in cell proliferation and apoptosis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10137–10142 (2004).
484. Cogswell, P. C., Guttridge, D. C., Funkhouser, W. K. & Baldwin, A. S. Selective activation of NF-kappa B subunits in human breast cancer: potential roles for NF-kappa B2/p52 and for Bcl-3. *Oncogene* **19**, 1123–1131 (2000).
485. Nakshatri, H. & Goulet, R. J. NF-kappaB and breast cancer. *Curr. Probl. Cancer* **26**, 282–309 (2002).
486. Nakshatri, H., Bhat-Nakshatri, P., Martin, D. A., Goulet, R. J. & Sledge, G. W. Constitutive activation of NF-kappaB during progression of breast cancer to hormone-independent growth. *Mol. Cell. Biol.* **17**, 3629–3639 (1997).
487. Sovak, M. A. *et al.* Aberrant nuclear factor-kappaB/Rel expression and the pathogenesis of breast cancer. *J. Clin. Invest.* **100**, 2952–2960 (1997).
488. Arkan, M. C. & Greten, F. R. IKK- and NF-κB-mediated functions in carcinogenesis. *Curr. Top. Microbiol. Immunol.* **349**, 159–169 (2011).

489. Bassères, D. S. & Baldwin, A. S. Nuclear factor-kappaB and inhibitor of kappaB kinase pathways in oncogenic initiation and progression. *Oncogene* **25**, 6817–6830 (2006).
490. Chua, H. L. *et al.* NF-kappaB represses E-cadherin expression and enhances epithelial to mesenchymal transition of mammary epithelial cells: potential involvement of ZEB-1 and ZEB-2. *Oncogene* **26**, 711–724 (2007).
491. Huber, M. A. *et al.* NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J. Clin. Invest.* **114**, 569–581 (2004).
492. Prasad, S., Ravindran, J. & Aggarwal, B. B. NF-kappaB and cancer: how intimate is this relationship. *Mol. Cell. Biochem.* **336**, 25–37 (2010).
493. Smith, S. M., Lyu, Y. L. & Cai, L. NF-κB Affects Proliferation and Invasiveness of Breast Cancer Cells by Regulating CD44 Expression. *PLOS ONE* **9**, e106966 (2014).
494. Yamaguchi, N. *et al.* Constitutive activation of nuclear factor-kappaB is preferentially involved in the proliferation of basal-like subtype breast cancer cell lines. *Cancer Sci.* **100**, 1668–1674 (2009).
495. Heckman, C. A., Mehew, J. W. & Boxer, L. M. NF-kappaB activates Bcl-2 expression in t(14;18) lymphoma cells. *Oncogene* **21**, 3898–3908 (2002).
496. Hinz, M. *et al.* NF-kappaB function in growth control: regulation of cyclin D1 expression and G0/G1-to-S-phase transition. *Mol. Cell. Biol.* **19**, 2690–2698 (1999).
497. Guttridge, D. C., Albanese, C., Reuther, J. Y., Pestell, R. G. & Baldwin, A. S. NF-kappaB controls cell growth and differentiation through transcriptional regulation of cyclin D1. *Mol. Cell. Biol.* **19**, 5785–5799 (1999).
498. Joyce, D. *et al.* NF-kappaB and cell-cycle regulation: the cyclin connection. *Cytokine Growth Factor Rev.* **12**, 73–90 (2001).

499. Wuerzberger-Davis, S. M., Chang, P.-Y., Berchtold, C. & Miyamoto, S. Enhanced G2-M arrest by nuclear factor- κ B-dependent p21waf1/cip1 induction. *Mol. Cancer Res. MCR* **3**, 345–353 (2005).
500. Chen, X., Shu, Y., Li, W. & Yin, Y. TNF-alpha-induced metastasis gene changes in MCF-7 cells. *J. Nanjing Med. Univ.* **22**, 366–371 (2008).
501. Akca, H., Yenisoy, S., Yanikoglu, A. & Ozes, O. N. Tumor necrosis factor-alpha-induced accumulation of tumor suppressor protein p53 and cyclin-dependent protein kinase inhibitory protein p21 is inhibited by insulin in ME-180S cells. *Clin. Chem. Lab. Med.* **40**, 764–768 (2002).
502. Kumar, P. S. *et al.* Differential expression and role of p21cip/waf1 and p27kip1 in TNF- α -induced inhibition of proliferation in human glioma cells. *Mol. Cancer* **6**, 42 (2007).
503. Ryazantseva, N. V. *et al.* Role of NF- κ B, p53, and p21 in the regulation of TNF- α mediated apoptosis of lymphocytes. *Bull. Exp. Biol. Med.* **149**, 50–53 (2010).
504. Wu, X. *et al.* TNF- α sensitizes chemotherapy and radiotherapy against breast cancer cells. *Cancer Cell Int.* **17**, (2017).
505. Cao, W., Chi, W.-H., Wang, J., Tang, J.-J. & Lu, Y.-J. TNF-alpha promotes Doxorubicin-induced cell apoptosis and anti-cancer effect through downregulation of p21 in p53-deficient tumor cells. *Biochem. Biophys. Res. Commun.* **330**, 1034–1040 (2005).
506. Ameyar, M. *et al.* Analysis of Human Breast Adenocarcinoma MCF7 Resistance to Tumor Necrosis Factor-induced Cell Death LACK OF CORRELATION BETWEEN JNK ACTIVATION AND CERAMIDE PATHWAY. *J. Biol. Chem.* **273**, 29002–29008 (1998).
507. Burow, M. E. *et al.* Differences in susceptibility to tumor necrosis factor alpha-induced apoptosis among MCF-7 breast cancer cell variants. *Cancer Res.* **58**, 4940–4946 (1998).
508. Drané, P. *et al.* Accumulation of an inactive form of p53 protein in cells treated with TNF alpha. *Cell Death Differ.* **9**, 527–537 (2002).

509. Wang, Z., Kishimoto, H., Bhat-Nakshatri, P., Crean, C. & Nakshatri, H. TNF α resistance in MCF-7 breast cancer cells is associated with altered subcellular localization of p21CIP1 and p27KIP1. *Cell Death Differ.* **12**, 98–100 (2005).
510. Rozen, F., Zhang, J. & Pollak, M. Antiproliferative action of tumor necrosis factor-alpha on MCF-7 breastcancer cells is associated with increased insulin-like growth factor binding protein-3 accumulation. *Int. J. Oncol.* **13**, 865–869 (1998).
511. Lee, S.-H. & Nam, H.-S. TNF alpha-induced down-regulation of estrogen receptor alpha in MCF-7 breast cancer cells. *Mol. Cells* **26**, 285–290 (2008).
512. Khodjakov, A., Cole, R. W., Oakley, B. R. & Rieder, C. L. Centrosome-independent mitotic spindle formation in vertebrates. *Curr. Biol. CB* **10**, 59–67 (2000).
513. Hinchcliffe, E. H., Miller, F. J., Cham, M., Khodjakov, A. & Sluder, G. Requirement of a centrosomal activity for cell cycle progression through G1 into S phase. *Science* **291**, 1547–1550 (2001).
514. Balczon, R., Simerly, C., Takahashi, D. & Schatten, G. Arrest of cell cycle progression during first interphase in murine zygotes microinjected with anti-PCM-1 antibodies. *Cell Motil. Cytoskeleton* **52**, 183–192 (2002).
515. Grieshaber, S. S., Grieshaber, N. A., Miller, N. & Hackstadt, T. Chlamydia trachomatis causes centrosomal defects resulting in chromosomal segregation abnormalities. *Traffic Cph. Den.* **7**, 940–949 (2006).
516. Jouvenet, N. & Wileman, T. African swine fever virus infection disrupts centrosome assembly and function. *J. Gen. Virol.* **86**, 589–594 (2005).
517. Ploubidou, A. *et al.* Vaccinia virus infection disrupts microtubule organization and centrosome function. *EMBO J.* **19**, 3932–3944 (2000).

518. Vidair, C. A., Doxsey, S. J. & Dewey, W. C. Thermotolerant cells possess an enhanced capacity to repair heat-induced alterations to centrosome structure and function. *J. Cell. Physiol.* **163**, 194–203 (1995).
519. Mikule, K. *et al.* Loss of centrosome integrity induces p38—p53—p21-dependent G1—S arrest. *Nat. Cell Biol.* **9**, 160–170 (2007).
520. Bazzi, H. & Anderson, K. V. Acentriolar mitosis activates a p53-dependent apoptosis pathway in the mouse embryo. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1491-1500 (2014).
521. Khodjakov, A. & Rieder, C. L. Centrosomes enhance the fidelity of cytokinesis in vertebrates and are required for cell cycle progression. *J. Cell Biol.* **153**, 237–242 (2001).
522. Srsen, V., Gnadt, N., Dammermann, A. & Merdes, A. Inhibition of centrosome protein assembly leads to p53-dependent exit from the cell cycle. *J. Cell Biol.* **174**, 625–630 (2006).
523. Nigg, E. A. Centrosome aberrations: cause or consequence of cancer progression? *Nat. Rev. Cancer* **2**, 815–825 (2002).
524. Pihan, G. A. *et al.* Centrosome defects can account for cellular and genetic changes that characterize prostate cancer progression. *Cancer Res.* **61**, 2212–2219 (2001).
525. Pihan, G. A. *et al.* Centrosome defects and genetic instability in malignant tumors. *Cancer Res.* **58**, 3974–3985 (1998).
526. Hall, J. G., Flora, C., Scott, C. I., Pauli, R. M. & Tanaka, K. I. Majewski osteodysplastic primordial dwarfism type II (MOPD II): natural history and clinical findings. *Am. J. Med. Genet. A.* **130A**, 55–72 (2004).
527. Knox, J. D., Mitchel, R. E. & Brown, D. L. Effects of hyperthermia on microtubule organization and cytolytic activity of murine cytotoxic T lymphocytes. *Exp. Cell Res.* **194**, 275–283 (1991).
528. Kumar, A., Rajendran, V., Sethumadhavan, R. & Purohit, R. Role of centrosome in regulating immune response. *Curr. Drug Targets* **15**, 558–563 (2014).

529. de la Roche, M., Asano, Y. & Griffiths, G. M. Origins of the cytolytic synapse. *Nat. Rev. Immunol.* **16**, 421–432 (2016).
530. Stinchcombe, J. C. & Griffiths, G. M. Communication, the centrosome and the immunological synapse. *Phil Trans R Soc B* **369**, 20130463 (2014).
531. Nakamura, M. *et al.* When overexpressed, a novel centrosomal protein, RanBPM, causes ectopic microtubule nucleation similar to gamma-tubulin. *J. Cell Biol.* **143**, 1041–1052 (1998).
532. Palavicini, J. P. *et al.* RanBP9 Plays a Critical Role in Neonatal Brain Development in Mice. *PLOS ONE* **8**, e66908 (2013).
533. Chang, Y. *et al.* RanBPM Regulates the Progression of Neuronal Precursors through M-Phase at the Surface of the Neocortical Ventricular Zone. *Dev. Neurobiol.* **70**, 1–15 (2010).
534. Mikolajczyk, M., Shi, J., Vaillancourt, R. R., Sachs, N. A. & Nelson, M. The cyclin-dependent kinase 11(p46) isoform interacts with RanBPM. *Biochem. Biophys. Res. Commun.* **310**, 14–18 (2003).
535. Bai, D., Chen, H. & Huang, B.-R. RanBPM is a novel binding protein for p75NTR. *Biochem. Biophys. Res. Commun.* **309**, 552–557 (2003).
536. Kramer, S. *et al.* Protein stability and function of p73 are modulated by a physical interaction with RanBPM in mammalian cultured cells. *Oncogene* **24**, 938–944 (2005).
537. Wang, Y. *et al.* HIPK2 associates with RanBPM. *Biochem. Biophys. Res. Commun.* **297**, 148–153 (2002).
538. Denti, S. *et al.* RanBPM Is a Phosphoprotein That Associates with the Plasma Membrane and Interacts with the Integrin LFA-1. *J. Biol. Chem.* **279**, 13027–13034 (2004).
539. Matsuoka, S. *et al.* ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. *Science* **316**, 1160–1166 (2007).

540. Wei, J.-D., Jang, J.-H. & Kim, J.-H. RanBPM inhibits BLT2-mediated IL-8 production and invasiveness in aggressive breast cancer cells. *Biochem. Biophys. Res. Commun.*
doi:10.1016/j.bbrc.2016.12.147
541. Shao, S. *et al.* Reduced RanBPM Expression Is Associated with Distant Metastasis in Gastric Cancer and Chemoresistance. *Anticancer Res.* **36**, 1295–1303 (2016).
542. Palmieri, D. *et al.* Ran Binding Protein 9 (RanBP9) is a novel mediator of cellular DNA damage response in lung cancer cells. *Oncotarget* **7**, 18371–18383 (2016).
543. Wang, D., Li, Z., Messing, E. M. & Wu, G. Activation of Ras/Erk pathway by a novel MET-interacting protein RanBPM. *J. Biol. Chem.* **277**, 36216–36222 (2002).
544. Wang, L. *et al.* The Ran-binding protein RanBPM can depress the NF- κ B pathway by interacting with TRAF6. *Mol. Cell. Biochem.* **359**, 83–94 (2012).
545. Subramanian, M. *et al.* An AXL/LRP-1/RANBP9 complex mediates DC efferocytosis and antigen cross-presentation in vivo. *J. Clin. Invest.* **124**, 1296–1308 (2014).
546. Wang, Y. *et al.* Comparative transcriptome analysis reveals a regulatory network of microRNA-29b during mouse early embryonic development. *Oncotarget* (2016).
doi:10.18632/oncotarget.10741
547. Zhang, J. *et al.* Expression and potential role of microRNA-29b in mouse early embryo development. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **35**, 1178–1187 (2015).
548. Tan, T. Y., Gordon, C. T., Miller, K. A., Amor, D. J. & Farlie, P. G. YPEL1 overexpression in early avian craniofacial mesenchyme causes mandibular dysmorphogenesis by up-regulating apoptosis. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* **244**, 1022–1030 (2015).
549. Blaker-Lee, A., Gupta, S., McCammon, J. M., De Rienzo, G. & Sive, H. Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain

- development, and include two deletion dosage sensor genes. *Dis. Model. Mech.* **5**, 834–851 (2012).
550. Mackay, T. F. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).
551. Beyaert, R. *et al.* Cancer risk in immune-mediated inflammatory diseases (IMID). *Mol. Cancer* **12**, 98 (2013).
552. Kuek, A., Hazleman, B. L. & Östör, A. J. K. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. *Postgrad. Med. J.* **83**, 251–260 (2007).

