

Université de Montréal

**Les macrophages d'ascendance européenne et africaine répondent
différemment aux infections bactériennes**

par
Ariane Pagé Sabourin

Biochimie
Faculté de Médecine

Mémoire présenté à la Faculté de médecine
en vue de l'obtention du grade de Maîtrise en biochimie

Décembre 2013

© Ariane Pagé Sabourin, 2013

Université de Montréal
Faculté des études supérieures et postdoctorales

ce mémoire intitulé :
**Les macrophages d'ascendance européenne et africaine répondent
différemment aux infections bactériennes**

présenté par :
Ariane Pagé Sabourin

a été évalué par un jury composé des personnes suivantes :

Nikolaus Heveker
Président-rapporteur

Luis B. Barreiro
Directeur de recherche

Jesse Shapiro
Membre du jury

Résumé

Des études antérieures démontrent que les descendants de peuples européens et africains présentent des différences de susceptibilité à certaines maladies infectieuses. Ces différences suggèrent des variations interpopulationnelles de la réponse immunitaire qui résultent probablement de l'adaptation de ces individus aux pathogènes de leur environnement. Nous avons caractérisé la réponse immunitaire chez des descendants de peuples européens et africains à des infections bactériennes. Nous avons infecté des macrophages dérivés de monocytes de 30 Américains d'origine africaine (Africains) et de 31 Américains d'origine européenne (Européens) avec les pathogènes intracellulaires *Listeria monocytogenes* et *Salmonella typhimurium* pendant 4 heures, puis nous avons mesuré le niveau d'expression pangénomique des cellules infectées et non infectées par séquençage de l'ARNm. Nous avons estimé le niveau de contrôle de l'infection par les macrophages à 2, 4 et 24 heures post-infection en évaluant le taux de survie des bactéries. Nous avons observé que les Africains présentent significativement moins de bactéries intracellulaires après 4 et 24 heures que les Européens, suggérant que les Africains contrôlent mieux les infections bactériennes. Nous avons identifié des différences interpopulationnelles dans le niveau de sécrétion des cytokines et dans le niveau d'expression de certains gènes, ce qui suggère que les Africains modulent une réponse inflammatoire plus forte que les Européens. Nous avons démontré que plusieurs de ces gènes ont subi des événements de sélection positive récents seulement chez les Européens. Notre étude a identifié plusieurs gènes candidats susceptibles d'influencer le cours des infections bactériennes chez les humains. Nos résultats indiquent que les différences dans la progression des maladies infectieuses entre les populations européennes et africaines seraient le résultat de la sélection naturelle.

Mots clés : génomique humaine, interaction hôte-pathogène, expression génétique, immunologie, génétique des populations

Abstract

Previous studies demonstrate that people of African and European ancestry differ in their susceptibility to certain infectious diseases. Differences in infection progression between these populations suggest inter-population variation in the immune response, possibly caused by adaptation to the pathogens of their historical environments. Here, we characterize the immune response of people of African and European ancestry to bacterial infections. Monocyte-derived macrophages from 30 African Americans (Africans) and 31 European Americans (Europeans) were infected with the intracellular pathogens *Listeria monocytogenes* and *Salmonella typhimurium* for 4 hours and whole genome gene expression of infected and non-infected cells was measured by RNA-sequencing. Macrophage control of bacterial infection at 2, 4 and 24 hours was assessed by culturing infected cell lysate and counting colony-forming units to approximate bacterial survival rate. We found that macrophages derived from Africans presented fewer intracellular bacteria after 4 and 24 hours than Europeans, suggesting that Africans better control intracellular bacterial infections. Concordant with this observation, we identified inter-population differences in cytokine secretion and gene expression that might explain this pattern of increased infection control in Africans. Interestingly, several of those differences indicate that Africans have a stronger pro-inflammatory response than Europeans. We show that several of these genes appear to have been subject to recent selection in the Europeans population alone. We also identify multiple candidate genes that may affect the course of infection in these populations. Overall, our findings suggest that differences in infectious disease progression observed in Africans and in Europeans may be the outcome of natural selection.

Keywords: human genomics, host-pathogen interactions, gene expression, immunology, population genetics

Table des matières

Résumé	i
Abstract.....	ii
Liste des tableaux	vi
Liste des figures	vii
Liste des sigles et des abréviations.....	ix
Remerciements.....	xiv
1. Introduction	1
1.1 Les maladies infectieuses et l'homme	1
1.1.1 La sélection naturelle.....	1
1.1.2 Méthodes statistiques de détection de la sélection naturelle	3
1.1.3 La sélection naturelle et les maladies infectieuses.....	5
1.2 Biologie des infections bactériennes	9
1.2.1 Salmonella typhimurium.....	9
1.2.2 Listeria monocytogenes.....	12
1.3 L'étude du transcriptome et les variations fonctionnelles chez l'humain.....	15
1.3.1 Régulation de l'expression génétique	16
1.3.2 Génomique des populations de l'expression génétique	18
2. Objectifs et hypothèse.....	20
2.1 Problématique	20
2.2 Hypothèse.....	20
2.3 Objectifs.....	21
3. Matériels et méthodes	21
3.1 Culture cellulaire.....	24
3.1.1 Extraction et purification	24
3.1.2 Différenciation des monocytes en macrophages	24
3.1.3 Infection.....	24
3.2 Séquençage de l'ARNm.....	26
3.2.1 Préparation des bibliothèques pour le séquençage de l'ARNm	26
3.2.2 Description de la technique de séquençage de l'ARNm.....	26

3.3	Génotypage de l'ADN	27
3.3.1	Préparation de l'ADN	27
3.3.2	Description de la technique de génotypage de l'ADN	27
3.3.3	Analyse des données de génotypage de l'ADN	28
3.4	Quantification des cytokines sécrétées	28
3.4.1	Description de la technique de quantification des cytokines sécrétées	28
3.4.2	Analyse des données quantification des cytokines sécrétées	29
4.	Résultats	30
4.1	Caractérisation de la réponse immunitaire des macrophages aux infections bactériennes.....	30
4.1.1	Analyse en composantes principales.....	30
4.1.2	Caractérisation des différences dans les niveaux d'expression génétique entre la condition non-infectée et les conditions infectées.....	31
4.1.3	Caractérisation des différences dans les niveaux d'expression génétique entre les conditions <i>Listeria monocytogenes</i> et <i>Salmonella typhimurium</i>	35
4.2	Caractérisation interpopulationnelle de la réponse immunitaire des macrophages aux infections bactériennes	38
4.2.1	Caractérisation interpopulationnelle de l'efficacité du contrôle des bactéries par les macrophages.....	38
4.2.2	Caractérisation interpopulationnelle des différences dans les niveaux de cytokines sécrétées	40
4.2.3	Caractérisation interpopulationnelle des niveaux d'expression génétique	44
4.2.4	Identification d'eQTLs	48
4.3	Signatures de sélection naturelle dans les gènes différemment exprimés entre les populations	50
4.3.1	Différenciation des populations	50
4.3.2	Identification de signatures de sélection positive récentes	52
5.	Discussion	55
5.1	<i>Listeria monocytogenes</i> et <i>Salmonella typhimurium</i> ont induit une réponse immunitaire innée spécifique.....	55
5.2	Les Européens éliminent les bactéries intracellulaires plus lentement que les Africains	55
5.3	Les Européens sécrètent une plus grande quantité de cytokines anti-inflammatoires que les Africains	56

5.4 Les Africains et les Européens présentent des différences dans le niveau d'expression de certains gènes	57
5.5 Les Africains présentent une réponse inflammatoire plus forte que les Européens	59
5.6 Des gènes différemment exprimés entre les populations sont marqués par la sélection naturelle chez les Européens	61
5.7 Perspectives	61
6. Conclusion	63
Bibliographie	64

Liste des tableaux

Tableau I : Différences engendrées par les maladies infectieuses entre le taux de mortalité des Afro-Américains et celui des Américains d'origine caucasienne

Tableau II : Tableau des valeurs p des différences interpopulationnelles des niveaux de sécrétion de cytokines par les macrophages

Tableau III : Tableau des 5 groupes ontologiques les plus significatifs pour les gènes différemment exprimés entre les Africains et les Européens dans chacune des conditions

Liste des figures

Figure 1: Sélection négative

Figure 2: Sélection positive

Figure 3: Sélection balancée

Figure 4: Déséquilibre de liaison et la sélection positive

Figure 5: Distribution de l'allèle FY*O dans le monde

Figure 6: Biologie d'une infection à *Salmonella typhimurium*

Figure 7: Le SPI2 et la formation du SCV chez le macrophage

Figure 8: Entrée et propagation de *Listeria monocytogenes* dans les cellules

Figure 9: Activation de la réponse immunitaire par un macrophage infecté par *Listeria monocytogenes*

Figure 10: Effet d'un variant cis et trans sur l'expression d'un gène

Figure 11: Objectifs du projet de recherche

Figure 12: Schéma expérimental

Figure 13: Analyse en composantes principales

Figure 14: Graphique de type Volcano mettant en évidence les gènes différemment exprimés entre les conditions *Listeria monocytogenes* et non infectée

Figure 15: Graphique de type Volcano mettant en évidence les gènes différemment exprimés entre les conditions *Salmonella typhimurium* et non infectée

Figure 16: Graphique à barres des 20 termes GO les plus significatifs pour les gènes *up-régulés* suite aux infections bactériennes

Figure 17: Graphique à barres des 20 termes GO les plus significatifs pour les gènes *down-régulés* suite aux infections bactériennes

Figure 18: Graphique de type Volcano mettant en évidence les gènes différemment exprimés entre les conditions *Salmonella typhimurium* et *Listeria monocytogenes*

Figure 19: Voie de signalisation de la production d'interféron-gamma

Figure 20: Boîte à moustaches des ratios CIP / CI pour les infections avec *Listeria monocytogenes*

Figure 21: Boîte à moustaches des ratios CIP / CI pour les infections avec *Salmonella typhimurium*

Figure 22: Niveaux normalisés de façon quantile des cytokines IL-1RN, IL-10 et IL-13 sécrétées dans le surnageant des macrophages chez les Africains et les Européens

Figure 23: Diagramme de Venn des gènes différemment exprimés entre les Africains et les Européens dans chacune des conditions

Figure 24: Le gène IL1R1 est plus exprimé chez les Africains que chez les Européens dans la condition non infectée

Figure 25: Le gène mir155HG est plus exprimé chez les Africains que chez les Européens dans la condition *Salmonella typhimurium*

Figure 26: Proportion des gènes qui ont un eQTL parmi les gènes qui sont différemment exprimés entre les populations

Figure 27: Le gène RETN est plus exprimé chez les Africains que chez les Européens dans la condition *Salmonella typhimurium*

Figure 28: Proportion des gènes qui ont au moins un SNP avec un coefficient F_{ST} élevé parmi les gènes qui sont différemment exprimés entre les populations

Figure 29: Proportion des gènes qui ont une forte proportion de SNPs avec des valeurs d'iHS élevées chez les Africains et chez les Européens parmi les gènes qui sont différemment exprimés entre les populations

Liste des sigles et des abréviations

°C	degrés Celsius
µL	microlitre
3'UTR	<i>three prime untranslated region</i>
A	adénine
ADN	acide désoxyribonucléique
ADNc	acide désoxyribonucléique complémentaire
AF	Africains
AP-1	<i>activator protein 1</i>
ARN	acide ribonucléique
ARNm	acide ribonucléique messenger
C	cytosine
CCL11	<i>C-C motif chemokine 11</i>
CCL2	<i>C-C motif chemokine 2</i>
CCL3	<i>C-C motif chemokine 3</i>
CCL5	<i>C-C motif chemokine 5</i>
CCR5	<i>C-C chemokine receptor type 5</i>
CD14	<i>cluster of differentiation 14</i>
CD1A	<i>Cluster of Differentiation 1a</i>
CD40	<i>Cluster of Differentiation 40</i>

CD59	<i>Cluster of Differentiation 59</i>
CD83	<i>Cluster of Differentiation 83</i>
CEU	résidents de l'Utah originaires de l'Europe du Nord et de l'Ouest
CFU	<i>colony-forming unit</i>
CHB	Chinois Han de Beijing
CI	valeurs de CFU au moment de l'inoculation
CIP	valeurs de CFU post-infection
CMAH	<i>Putative cytidine monophosphate-N-acetylneuraminic acid hydroxylase-like protein</i>
CSF3	<i>Colony Stimulating Factor 3</i>
CXCL1	<i>Chemokine (C-X-C motif) ligand 1</i>
CXCL10	<i>Chemokine (C-X-C motif) ligand 10</i>
DARC	<i>Duffy antigen/chemokine receptor</i>
DO	densité optique
eQTL	<i>expression quantitative trait loci</i>
EU	Européens
FC	<i>fold change</i>
FDR	<i>false discovery rate</i>
FGF2	<i>Basic fibroblast growth factor</i>
F_{ST}	indice de différenciation
G	gramme
G	guanine

G6PD	<i>Glucose-6-phosphate dehydrogenase</i>
GO	<i>gene ontology</i>
HLA	<i>Human leukocyte antigen</i>
HLA-DR	<i>Human leukocyte antigen receptor</i>
IFNG	<i>Interferon gamma</i>
iHS	<i>integrated haplotype Score</i>
IL-10	<i>Interleukine 10</i>
IL-12	<i>Interleukine 12</i>
IL-12A	<i>Interleukine 12 subunit alpha</i>
IL-13	<i>Interleukine 13</i>
IL-15	<i>Interleukine 15</i>
IL-17A	<i>Interleukine 17 subunit alpha</i>
IL-1R1	<i>Interleukin-1 receptor 1</i>
IL-1RN	<i>Interleukin-1 receptor antagonist</i>
IL-1 α	<i>Interleukin-1 alpha</i>
IL-1 β	<i>Interleukin-1 beta</i>
IL-2	<i>Interleukine 2</i>
IL-4	<i>Interleukine 4</i>
IL-4R α	<i>Interleukine 4 receptor</i>
IL-5	<i>Interleukine 5</i>
IL-6	<i>Interleukine 6</i>

IL-7	<i>Interleukine 7</i>
IL-8	<i>Interleukine 8</i>
IL-9	<i>Interleukine 9</i>
JAK2	<i>Janus kinase 2</i>
JPT	Japonais de Tokyo
Kb	kilobases
Log2	logarithme de base 2
M-CSF	<i>Macrophage colony-stimulating factor</i>
miR-155	micro-ARN 155
MIR155HG	<i>MIR155 host gene</i>
mL	millilitres
MOI	multiplicity of infection
MyD88	<i>Myeloid differentiation primary response gene 88</i>
NF-κB	<i>Nuclear factor kappa-B</i>
ng	nanogrammes
pb	paires de bases
PBS	<i>phosphate buffered saline</i>
PCR	<i>polymerase chain reaction</i>
PDGFB	<i>Platelet-derived growth factor subunit B</i>
pH	potentiel hydrogène
RETN	<i>Resistin</i>

SCV	<i>Salmonella-containing vacuole</i>
Sifs	<i>Salmonella-induced filaments</i>
SNP	<i>single nucleotide polymorphism</i>
SPI1	<i>Salmonella pathogenicity island 1</i>
SPI2	<i>Salmonella pathogenicity island 2</i>
STAT3	<i>Signal transducer and activator of transcription 3</i>
STAT4	<i>Signal transducer and activator of transcription 4</i>
STAT6	<i>Signal transducer and activator of transcription 6</i>
T	thymine
TLR	<i>Toll-like receptor</i>
TLR2	<i>Toll-like receptor 2</i>
TLR4	<i>Toll-like receptor 4</i>
TLR5	<i>Toll-like receptor 5</i>
TNF	<i>Tumor necrosis factor</i>
TNFA	<i>Tumor necrosis factor alpha</i>
TSS	<i>transcription starting site</i>
VEGFA	<i>Vascular endothelial growth factor A</i>
VIH	virus de l'immunodéficience humaine
YRI	Yoruba d'Ibadan, Nigéria

Remerciements

D'abord, je voudrais remercier mon directeur de recherche, Dr Luis Barreiro, qui a su me guider et me conseiller à travers cette maîtrise. Je le remercie tout spécialement pour sa grande disponibilité et tout le temps qu'il m'a accordé pour faire de moi une meilleure scientifique.

Un merci tout particulier à Vania Yotova, ma «maman» du laboratoire, qui m'a offert un soutien et une aide énorme. Sans toi le projet ne serait certainement pas rendu au point où il en est. Merci pour tout, tu es une personne très spéciale.

J'aimerais aussi remercier Jean-Christophe Grenier et Yohann Nédélec pour les conseils, les discussions et leur support en bioinformatique.

Une pensée spéciale pour Anne Dumaine, Natacha Cotta, Jonathan Boulais, Jessica Brinkworth, Alain Pacis ainsi que tous les membres du laboratoire pour votre présence et vos conseils. J'ai eu beaucoup de plaisir à travailler avec vous.

Finalement, j'aimerais remercier Olivier Rolko qui a pris le temps de lire et de corriger mon mémoire. Merci d'être là.

1. Introduction

1.1 Les maladies infectieuses et l'homme

Jusqu'à tout récemment, les maladies infectieuses étaient un problème de santé majeur, restreignant fortement l'espérance de vie et imposant une forte pression sélective sur le génome humain. Depuis la théorie de l'origine microbienne des maladies infectieuses introduite par Pasteur au 19^e siècle, nos connaissances sur l'interaction hôte-pathogènes ont bien évolué, menant à l'utilisation des vaccins et des antibiotiques.¹ Malgré tout, les maladies infectieuses sont encore un problème de santé mondiale, causant la mort de 15 millions de personnes chaque année.² Les variations interindividuelles dans la susceptibilité aux maladies infectieuses peuvent être, en partie, expliquées par des facteurs environnementaux tels que l'alimentation, les conditions d'hygiène et la souche responsable de l'infection mais aussi par les facteurs génétiques de l'hôte, qui jouent un rôle déterminant.^{3,4}

La dynamique évolutive de l'interaction hôte-pathogènes est constamment en changement, entraînant une adaptation bilatérale chez l'humain et les agents infectieux. Au cours de l'évolution, notre système immunitaire a dû s'adapter et développer des mécanismes complexes pour se défendre contre les infections microbiennes. Il a été démontré que les gènes du système immunitaire sont marqués par des événements de sélection naturelle récents.^{5,6} Cette observation suggère que notre système immunitaire a été mis à l'épreuve durant les phases plus récentes de notre évolution, probablement à cause du changement de mode vie associé à l'arrivée de l'agriculture, il y a 10 000 ans.⁶ L'implantation de l'agriculture a eu pour effet de regrouper un plus grand nombre d'individus ensemble et ainsi de faciliter la propagation de certaines infections nécessitant une grande densité de population pour persister. Ceci, combiné à de mauvaises pratiques d'hygiène et à la domestication d'animaux, qui a favorisé les zoonoses, a contribué à la propagation des maladies infectieuses.⁷

1.1.1 La sélection naturelle

L'étendue des variations phénotypiques observées dans les populations humaines est due, en partie, à différents processus de sélection naturelle et d'adaptation aux conditions environnementales. La

sélection naturelle peut prendre différentes formes, entraînant des conséquences évolutives distinctes.

1.1.1.1 Sélection négative

La sélection négative correspond à la suppression d'allèles délétères dans une population (Figure 1). C'est probablement la forme de sélection naturelle la plus répandue dans notre génome, sachant que 38 à 75 % des mutations causant un changement d'acides aminés en sont affectées.^{8,9} En présence de sélection négative, les allèles dommageables vont tendre à disparaître ou à être maintenue à une faible fréquence dans une population. Ceci a pour effet de diminuer la diversité et d'augmenter la présence d'allèles rares (dans le cas d'une faible sélection négative).¹⁰

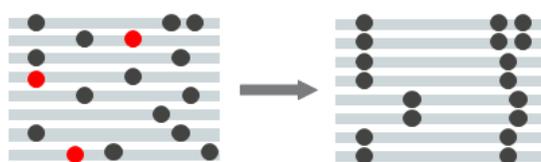


Figure 1: La sélection négative tend à faire disparaître les allèles délétères d'une population. L'image de gauche représente le profil de diversité génétique avant l'effet de la sélection négative et l'image de droite représente le profil de diversité génétique après l'effet de la sélection négative. Les allèles porteurs de variants délétères (points rouges) ont été éliminés.

1.1.1.2 Sélection positive

La sélection positive correspond à l'augmentation de la fréquence d'un allèle dans une population (Figure 2). Lorsqu'une mutation ayant un effet avantageux apparaît dans une population, l'allèle porteur de cette mutation augmente en fréquence, entraînant avec lui les polymorphismes neutres qui lui sont associés. Cela a donc pour effet de diminuer la diversité dans la région du génome positivement sélectionnée.¹⁰



Figure 2: La sélection positive tend à favoriser un allèle au détriment des autres dans une population. L'image de gauche représente le profil de diversité génétique avant l'effet de la sélection positive et

l'image de droite représente le profil de diversité génétique après l'effet de la sélection positive. L'allèle porteur du variant ayant un effet avantageux (point rouge) a augmenté en fréquence.

1.1.1.3 Sélection balancée

La sélection balancée favorise la présence d'allèles à des fréquences intermédiaires (Figure 3). On la retrouve dans deux types de situations : lorsque le stade hétérozygote est avantageux et lorsque l'avantage d'une mutation dépend de sa fréquence relative aux autres phénotypes d'une population. Ce type de sélection favorise la diversité.¹⁰

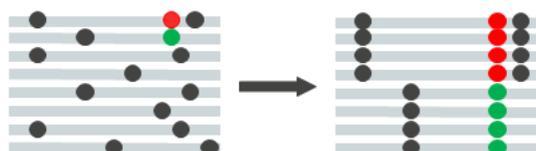


Figure 3: La sélection balancée tend à conserver les allèles avantageux au détriment des autres dans une population. L'image de gauche représente le profil de diversité génétique avant l'effet de la sélection balancée et l'image de droite représente le profil de diversité génétique après l'effet de la sélection balancée. Les allèles porteurs des variants ayant des effets avantageux (point rouge et point vert) ont augmenté en fréquence.

1.1.2 Méthodes statistiques de détection de la sélection naturelle

La sélection naturelle perturbe les patrons de variation génétique des régions qu'elle affecte par rapport aux régions du génome qui ne sont pas sous son influence. Par exemple, la sélection positive réduit la diversité génétique, augmente la présence d'allèles rares et augmente les niveaux de déséquilibre de liaison.¹⁰ Il existe des méthodes statistiques reposant sur l'observation des données polymorphiques pour détecter ces signatures de sélection naturelle. Deux méthodes statistiques pour détecter des événements de sélection naturelle récents à l'intérieur d'une même espèce seront décrites dans cette section.

1.1.2.1 La différenciation des populations

Le niveau de différenciation des populations, qui est mesuré par le coefficient F_{ST} , est calculé à partir des variations dans la fréquence allélique des SNPs entre les populations. Selon un modèle de neutralité, le coefficient F_{ST} est seulement déterminé par la dérive génétique (évolution aléatoire). L'action de la sélection naturelle dans une population, pour un locus précis, aura pour effet

d'augmenter le coefficient F_{ST} pour ce locus. La sélection positive locale a pour effet d'augmenter la différenciation des populations tandis que la sélection négative et la sélection balancée réduisent la différenciation des populations.¹¹

1.1.2.2 Le déséquilibre de liaison

Les tests basés sur les motifs de déséquilibre de liaison pour détecter la sélection naturelle sont très puissants pour identifier des événements sélectifs récents.¹² Ces tests sont : le *long-range haplotype test*¹³, le *integrated haplotype score (iHS) test*⁵ et le *linkage disequilibrium decay test*⁶. Ils partagent tous la même logique : une mutation survient sur un haplotype possédant déjà un patron de déséquilibre de liaison complet, donc cette mutation est pleinement associée aux polymorphismes présents sur cet haplotype. Selon un modèle d'évolution neutre, ce nouveau polymorphisme nécessite une longue période de temps pour atteindre une fréquence élevée dans la population et le patron de déséquilibre de liaison autour de cette région a diminué après les nombreuses recombinaisons. Les allèles anciens et communs ont donc une fréquence élevée et un patron de déséquilibre de liaison court. Les allèles marqués par la sélection positive ont typiquement présents à une haute fréquence, mais ont un long patron de déséquilibre de liaison puisqu'ils n'ont pas subi suffisamment d'événements de recombinaison pour briser l'haplotype sur lequel la mutation initiale est apparue (Figure 4).¹⁴

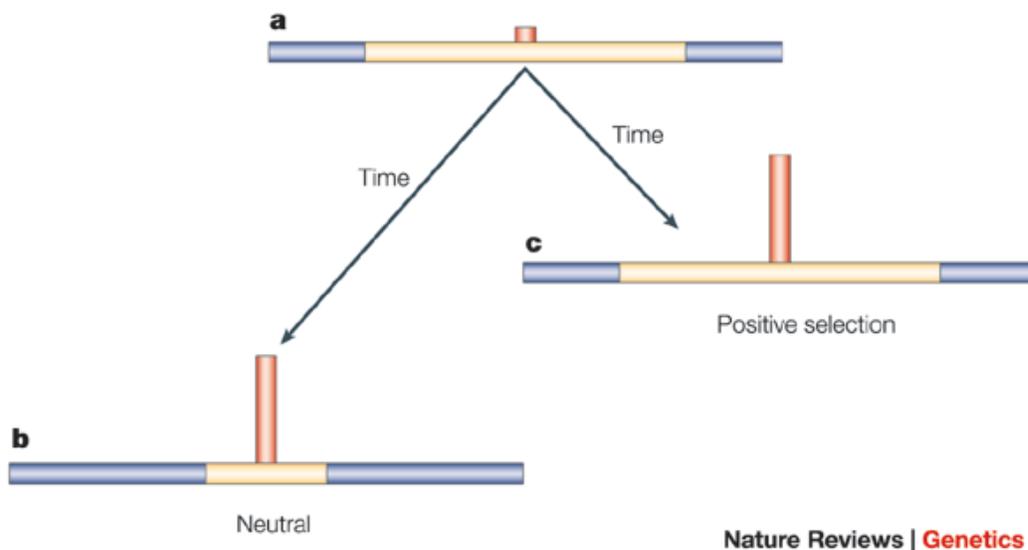


Figure 4: Le déséquilibre de liaison et la sélection positive.¹⁴ A) Le nouvel allèle existe à une faible fréquence (représenté par la barre verticale rouge) sur un haplotype qui présente un long patron de déséquilibre de liaison (en rose). B) Avec le temps, par dérive génétique, la fréquence de cet allèle augmente et les recombinaisons diminuent la longueur du patron de déséquilibre de liaison. C) Un allèle sous l'influence de la sélection positive augmente en fréquence rapidement et la recombinaison n'a pas le temps de briser le déséquilibre de liaison de cet allèle et des marqueurs associés.

1.1.3 La sélection naturelle et les maladies infectieuses

1.1.3.1 La sélection naturelle dans les gènes impliqués dans l'interaction hôte-pathogène

Parmi les populations humaines, les gènes de la réponse immunitaire sont la cible première de la sélection naturelle, résultat de l'interaction dynamique entre ces gènes et les pathogènes au cours de l'évolution.^{5,15,16} Un des exemples les plus marquants de sélection naturelle imposée par les pathogènes est illustré par les *human leukocyte antigens* (HLA). Les protéines du HLA jouent un rôle central dans le contrôle des infections en présentant les antigènes provenant du milieu intracellulaire ou extracellulaire aux lymphocytes. Les gènes du HLA sont hautement polymorphiques. La grande diversité génétique observée dans ce groupe de gènes est le résultat de la sélection balancée et il permet de reconnaître un large éventail de pathogènes.^{17,18} La diversité génétique du HLA a d'ailleurs été positivement corrélée avec la diversité locale des pathogènes.¹⁹

Parmi les gènes impliqués dans la réponse immunitaire et marqués par la sélection naturelle (revus par Vallender & Lahn¹⁵), on retrouve, entre autres les gènes CD59, CCR5, CMAH, IL4 et TNFSSF5. Ces gènes ne sont probablement qu'un petit échantillon de l'ensemble de ceux impliqués dans l'interaction hôte-pathogène et marqués par la sélection naturelle. En 2010, Barreiro & Quintana-Murci ont répertorié 187 gènes impliqués dans la réponse immunitaire, qui ont été associés à des événements de sélection positive dans au moins 2 scans pangénomiques de détection de la sélection naturelle.²⁰

1.1.3.2 L'exemple de la résistance à la malaria

Un exemple clair de sélection naturelle engendrée par les agents infectieux est celui de la malaria, une maladie infectieuse qui a imposé une grande pression sélective sur le génome humain et qui est probablement la force évolutive récente la plus puissante dans l'histoire de notre génome. Plusieurs variants conférant une protection contre la malaria ont été répertoriés à ce jour, notamment dans le gène G6PD²¹, dans le gène de la chaîne β de l'hémoglobine²² et dans le gène DARC²³. Le gène DARC code pour un récepteur de chimiokine exprimé à la surface de plusieurs types cellulaires. Le parasite *Plasmodium vivax* utilise ce récepteur pour pénétrer à l'intérieur des globules rouges. La mutation non-sens FY*O dans une région régulatrice empêche l'expression du gène DARC dans les globules rouges et induit une résistance complète à *Plasmodium vivax* chez les individus homozygotes porteurs de cet allèle.²⁴ Cet allèle est présent à des fréquences très élevées en Afrique, pouvant aller jusqu'à près de 100 % dans certaines populations africaines subsahariennes (Figure 5). La présence de la mutation FY*O en Afrique à des fréquences très élevées se retrouve dans des régions où *Plasmodium vivax* est absent, ce qui suggère que ce parasite a été éradiqué à cause de l'adaptation humaine.²⁵ Des études de génétique des populations ont démontré que le niveau de différenciation des populations pour l'allèle FY*O est parmi les plus élevés chez l'humain.^{26,27} Le cas de la mutation FY*O est donc un exemple où la sélection naturelle a eu effet dans certaines populations seulement en engendrant une différenciation des populations.

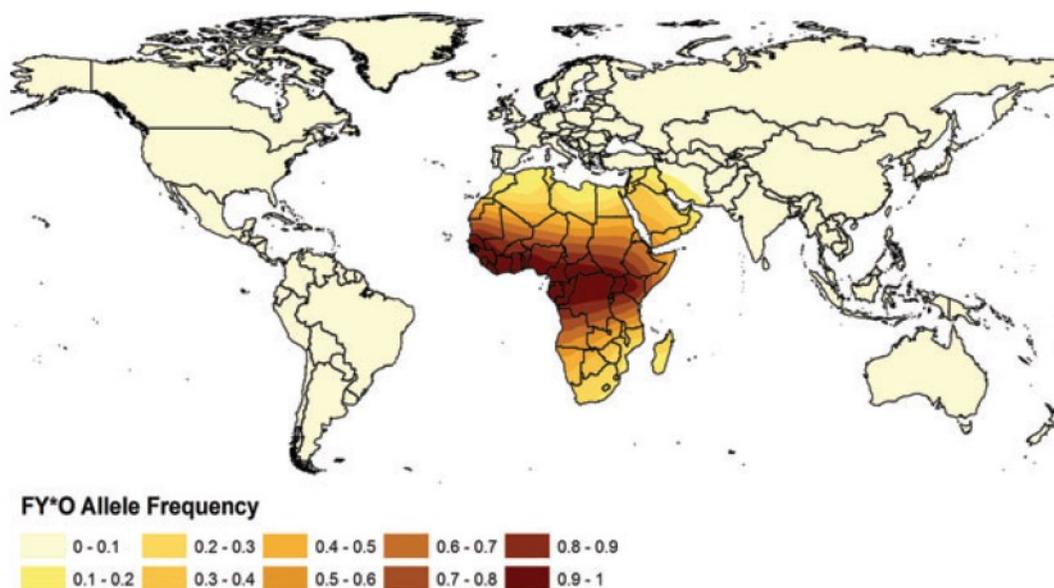


Figure 5: Distribution de l'allèle FY*O dans le monde. ²⁷

1.1.3.3 Les différences interpopulationnelles dans la susceptibilité aux infections

La résistance à la malaria est un exemple marquant et bien documenté où la sélection naturelle a introduit des variations dans la susceptibilité aux maladies infectieuses entre les populations. Néanmoins, sachant que les gènes impliqués dans la réponse immunitaire sont une cible importante de la sélection naturelle, il est raisonnable de penser que plusieurs autres gènes marqués par la sélection naturelle jouent un rôle important dans la réponse immunitaire. ^{4,16,17}

En 2001, une étude a démontré que les Américains d'origine caucasienne vivent en moyenne cinq ans de plus que les Afro-Américains. ²⁸ Cette étude a analysé des données provenant du *National Longitudinal Mortality Study*, qui est une base de données qui a été développée dans le but d'étudier les effets socio-économiques et démographiques sur les taux de mortalité. ²⁹ Cette étude compte un nombre d'individus total de 394 083 au début du suivi (de 1980 à 1989) et comprend 9,8% d'Afro-Américains. ³⁰ La mortalité associée aux maladies infectieuses est plus élevée chez les Afro-Américains que chez les Américains d'origine caucasienne (Tableau I). ³⁰ Par exemple, dans le Tableau I, le ratio des taux de mortalité (Afro-Américains/Américains d'origine caucasienne), pour la tuberculose, indique un excès de mortalité chez les Afro-Américains. Cet excès du taux de mortalité chez les Afro-Américains se traduit par un ratio des taux corrigé pour les facteurs socioéconomiques de 3,25.

		Pas contrôlé pour les facteurs socioéconomiques		Contrôlé pour les facteurs socioéconomiques	
Maladies infectieuses	Nombre total de morts	Ratio des taux	Intervalle de confiance 95%	Ratio des taux	Intervalle de confiance 95%
Tuberculose	22	4.11	1.61-10.50	3.25	1.22-8.61
Méningite bactérienne	15	0.84	0.11-6.29	0.71	0.09-5.48
Septicémie	299	2.15	1.57-2.95	1.87	1.35-2.58
VIH/sida	79	2.01	1.08-3.75	2.32	1.23-4.38
Maladies hépato-biliaires	26	2.16	0.74-6.35	2.05	0.67-6.28
Mycoses	42	3.05	1.39-6.69	1.93	0.86-4.37
Infections du cœur	127	0.88	0.44-1.73	0.73	0.37-1.46
Infections respiratoires	1163	1.16	0.94-1.42	1.03	0.84-1.27
Infection gastro-intestinales	31	2.64	1.08-6.43	2.16	0.86-5.44
Infections des reins et des voies urinaires	181	2.57	1.76-3.77	2.17	1.46-3.20
Toutes les maladies infectieuses	1985	1.53	1.33-1.76	1.34	1.16-1.54

Tableau I : Différences engendrées par les maladies infectieuses entre le taux de mortalité des Afro-Américains et celui des Américains d'origine caucasienne. Dans ce tableau, l'excès du taux de mortalité est calculé à partir du rapport des taux de mortalité (Afro-Américains/Américains d'origine caucasienne) et il est corrigé pour l'âge et le sexe. Dans les colonnes de droite, il est en plus corrigé pour des facteurs socioéconomiques, soit le niveau d'éducation et le revenu. Les maladies associées à un taux de mortalité significativement plus grand chez les Afro-Américains sont la tuberculose, les septicémies, le VIH/sida, les infections des reins et les infections urinaires.³⁰

1.2 Biologie des infections bactériennes

Dans ce projet, nous avons choisi les pathogènes intracellulaires *Salmonella typhimurium*, une bactérie Gram négatif et *Listeria monocytogenes*, une bactérie Gram positif. Ces deux bactéries ont développé des mécanismes leur permettant de pénétrer dans les cellules hôtes, de survivre aux défenses antimicrobiennes et de se répliquer puis de se propager vers d'autres cellules. Puisque ces deux pathogènes sont différents et ont une façon d'interagir avec la cellule hôte qui leur est propre, ils activent des voies de signalisation distinctes de la réponse immunitaire innée. Par exemple, nous savons que *Salmonella typhimurium* active préférentiellement le récepteur *Toll-like 4* (TLR4) tandis que *Listeria monocytogenes* active les récepteurs *Toll-like 2* et *5* (TLR2, TLR5). Nous considérons que ces deux bactéries nous permettent d'activer une grande proportion des voies de signalisation de la réponse immunitaire innée contre les infections bactériennes. Ceci nous permet donc d'obtenir un portrait global des variations de la réponse immunitaire innée entre les populations.

1.2.1 *Salmonella typhimurium*

Salmonella typhimurium est une bactérie Gram négative infectant un large éventail d'animaux et dont la transmission se fait généralement par l'ingestion d'eau ou d'aliments contaminés.³¹ Les manifestations cliniques sont multiples et incluent la gastro-entérite, des douleurs abdominales, des épisodes de diarrhée et des vomissements.³¹ *Salmonella typhimurium* est un pathogène qui survit au faible pH de l'estomac, puis colonise le petit intestin en pénétrant dans les cellules M (microfold) ou dans les entérocytes (Figure 6).^{32,33} La colonisation de l'épithélium intestinal par *Salmonella typhimurium* provoque une perturbation de la bordure en brosse et des jonctions serrées.^{34,35} Ces perturbations, combinées avec la réponse inflammatoire de l'intestin à l'invasion bactérienne, provoquent la sécrétion de différentes interleukines qui recrutent les leucocytes polymorphonucléaires.³⁶ Les leucocytes polymorphonucléaires sécrètent des granules cytotoxiques, stimulant ainsi l'inflammation de la muqueuse intestinale, tandis que l'action de *Salmonella*

typhimurium sur les jonctions serrées de l'épithélium intestinal provoque la diarrhée.³¹

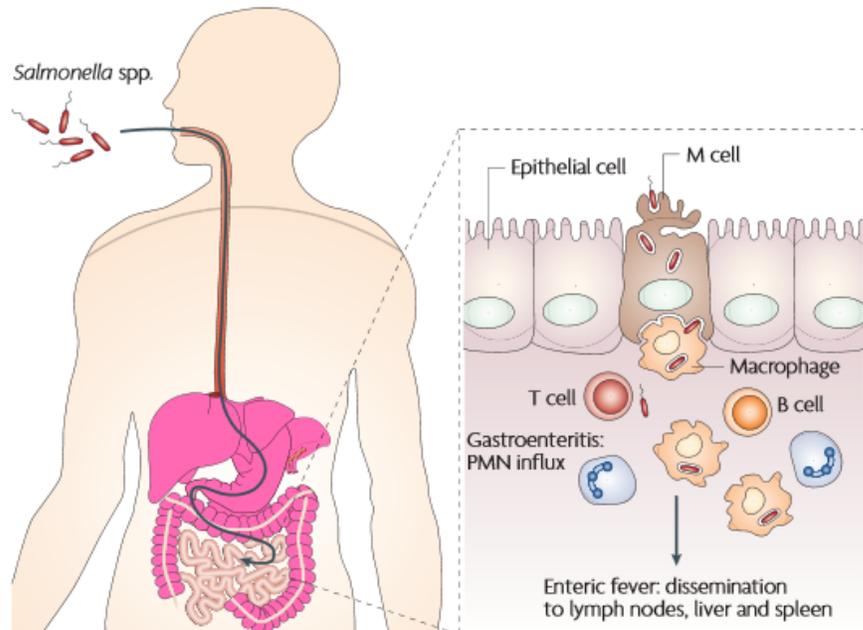


Figure 6: Biologie d'une infection à *Salmonella typhimurium*.³¹

1.2.1.1 Système de sécrétion de type 3

Le système de sécrétion de type 3 est essentiel dans la pathogénicité de *Salmonella*. Il est formé d'une vingtaine de protéines dont la structure rappelle la forme d'une aiguille et permet le transfert de protéines effectrices à travers la membrane cellulaire ou de la membrane du phagosome vers le cytoplasme de l'hôte. Ces protéines effectrices interagissent avec les fonctions cellulaires de l'hôte telles que l'architecture du cytosquelette, le transport membranaire, les signaux de transduction et l'expression de cytokines.³¹ Plusieurs protéines effectrices sont connues pour transiter par le système de sécrétion de type 3, mais leurs fonctions ne sont pas tous connues (revues par Garai et al.³⁷). *Salmonella typhimurium* encode plusieurs systèmes de sécrétion de type 3, dont les deux principaux sont le SPI1 et le SPI2. Le SPI1 est requis pour l'invasion des cellules non-phagocytiques et est actif dès que la bactérie entre en contact avec la cellule hôte. La bactérie se sert du SPI1 pour transférer ses protéines effectrices à travers la membrane cellulaire de l'hôte. Le SPI2 est exprimé lorsque la bactérie est dans le *Salmonella-containing vacuole* (SCV) (Figure 7). Il permet le transfert de protéines effectrices du SCV vers le cytoplasme de la cellule hôte. Il a été démontré que le SPI1 est requis pour l'invasion des cellules intestinales, pour la colonisation de l'intestin et pour le déclenchement d'une réponse inflammatoire de l'intestin. Le SPI2 est plutôt impliqué dans la survie et la réplication de la

bactérie à l'intérieur des macrophages et il confère une protection contre les dérivés réactifs de l'oxygène et contre les dérivés réactifs de l'azote.³⁸⁻⁴⁰

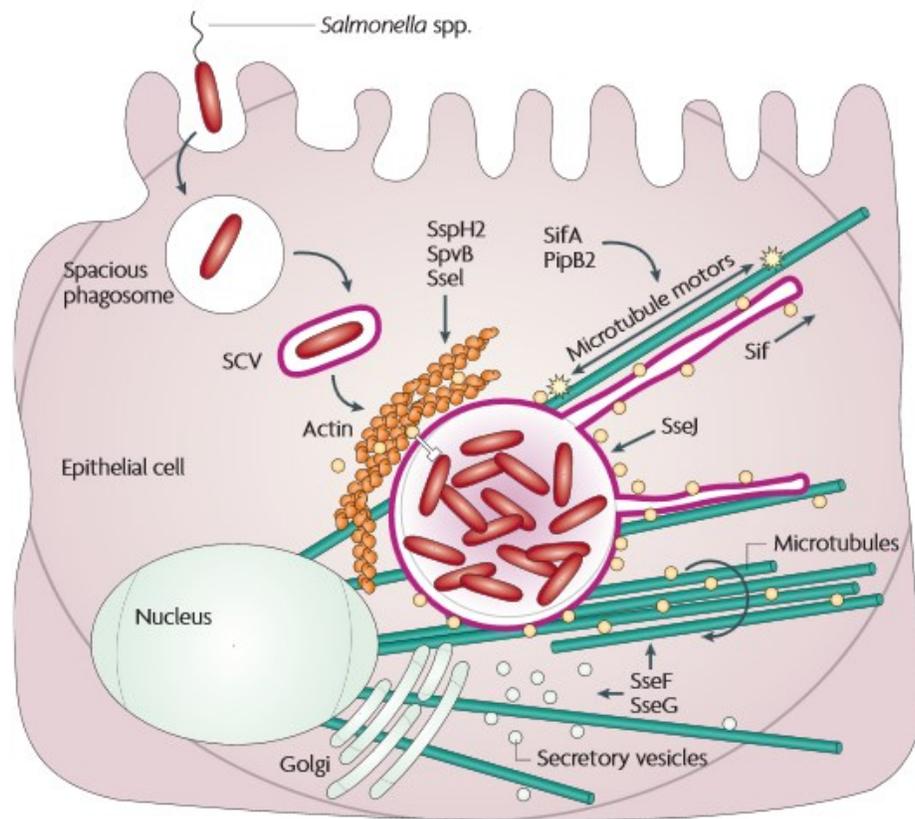


Figure 7: Le SPI2 et la formation du SCV chez le macrophage.³¹ Au centre de cette figure on voit le SCV, c'est-à-dire le phagosome contenant les bactéries *Salmonella typhimurium* et le SPI2 qui permet la translocation des effecteurs bactériens (représentés par les sphères jaunes) à travers la membrane du phagosome.

1.2.1.2 Survie et répllication de *Salmonella* dans les macrophages

Salmonella typhimurium est connue pour se répliquer principalement dans l'environnement hostile des macrophages. Elle a la capacité de s'adapter au contenu du SCV à l'intérieur des macrophages en remodelant sa paroi bactérienne pour résister aux peptides antimicrobiens et au stress oxydatif. Elle restructure aussi sa paroi bactérienne afin de réprimer les molécules que l'hôte reconnaît comme des marqueurs d'une infection, tels les lipopolysaccharides et la flagelline.⁴¹ Le système de sécrétion de type 3 SPI2 induit, dans les macrophages, la formation de longues membranes filamenteuses, les *Salmonella-induced filaments* (Sifs), qui prennent origine du SCV et qui s'étendent dans l'ensemble de

la cellule grâce à l'utilisation des microtubules.⁴² Ces Sifs auraient pour fonction d'augmenter la taille des SCV afin de faciliter la réplication des bactéries.³¹ Il semblerait que les protéines effectrices livrées dans le cytoplasme par le SPI2 auraient pour rôle de détourner le trafic cellulaire afin d'envoyer les métabolites nécessaires à la réplication de *Salmonella typhimurium* vers le SCV.^{43,44} Les protéines effectrices qui interagissent avec les microtubules semblent jouer un rôle dans le positionnement du SCV dans la cellule.⁴⁵ Par exemple, il semblerait qu'à certains moments durant l'infection, il serait préférable que le SCV se positionne près de l'appareil de Golgi afin d'intercepter les vésicules contenant des métabolites tandis qu'à certains autres moments, il serait préférable que le SCV se dirige vers la périphérie de la cellule pour prendre de l'expansion.³¹

1.2.2 *Listeria monocytogenes*

Listeria monocytogenes est une bactérie Gram positif qui infecte un large éventail de mammifères et dont la transmission se fait généralement par l'ingestion d'aliments contaminés.⁴⁶ Les individus immunodéprimés, les femmes enceintes et les personnes âgées sont particulièrement sensibles à cette bactérie, qui se manifeste sous forme de méningite et de septicémie.⁴⁷ Chez les femmes enceintes, une infection à *Listeria monocytogenes* peut se manifester sous forme de chorioamnionite et d'infection du fœtus, et peut causer un avortement spontané.⁴⁸ *Listeria monocytogenes* infecte généralement l'humain par la voie orale, survit au faible pH de l'estomac, puis pénètre les cellules épithéliales de l'intestin pour traverser la paroi intestinale et se propager dans la circulation sanguine et les ganglions lymphatiques. Elle se dirige ensuite vers le foie et la rate, où elle se réplique préférentiellement dans les macrophages spléniques et hépatiques.⁴⁹ En l'absence d'une réponse immunitaire appropriée, *Listeria monocytogenes* peut repénétrer la circulation sanguine pour se diriger vers le cerveau ou le placenta.⁵⁰ Elle a la capacité de traverser la barrière hémato-encéphalique et la barrière placentaire, et peut mener à une infection systémique fatale.⁴⁹

1.2.2.1 Régulation des facteurs de virulence

Les facteurs de virulences de *Listeria monocytogenes* contribuent à sa polyvalence et sont essentiels à sa pathogénicité. La presque totalité des gènes qui contribuent à l'invasion bactérienne, à l'entrée dans le cytosol, à la mobilité intracellulaire et à la propagation de *Listeria monocytogenes* est régulée par le facteur de transcription PrfA.⁵¹ Le cœur du régulon PrfA ne contient que dix gènes tandis qu'on connaît à ce jour 160 gènes directement ou indirectement régulés par PrfA.⁵² Les dix gènes régulés directement par PrfA sont revus par Freitag & al.⁵³ *Listeria monocytogenes* pénètre les cellules

épithéliales de l'intestin en liant leur internaline A membranaire à la l'E-cadhérine membranaire des cellules épithéliales.⁵⁴ Cette interaction est essentielle pour l'invasion cellulaire, car elle active des voies de signalisation menant à une réorganisation du cytosquelette. Elle traverse ensuite la barrière épithéliale vers le flux sanguin pour atteindre la rate et le foie, où elle est internalisée par les macrophages spléniques et hépatiques en exprimant une autre protéine de surface, l'internaline B (Figure 8).⁴⁸ Une fois à l'intérieur des cellules, *Listeria monocytogenes* s'échappe des phagosomes en sécrétant un facteur de virulence, la listériolysine, qui détruit la membrane phagosomale.⁵⁵ Une fois dans le cytosol, *Listeria monocytogenes* exprime actin-assembly-inducing protein, qui entraîne la nucléation de l'actine, créant des polymères que la bactérie peut utiliser pour se déplacer dans la cellule et vers d'autres cellules.^{56,57} Pour infecter une seconde cellule, *Listeria monocytogenes* crée une protrusion formée des deux membranes des cellules, puis force la formation d'une vacuole. Une fois à l'intérieur de la seconde cellule, elle peut la lyser puis recommencer un cycle d'infection.⁵⁸

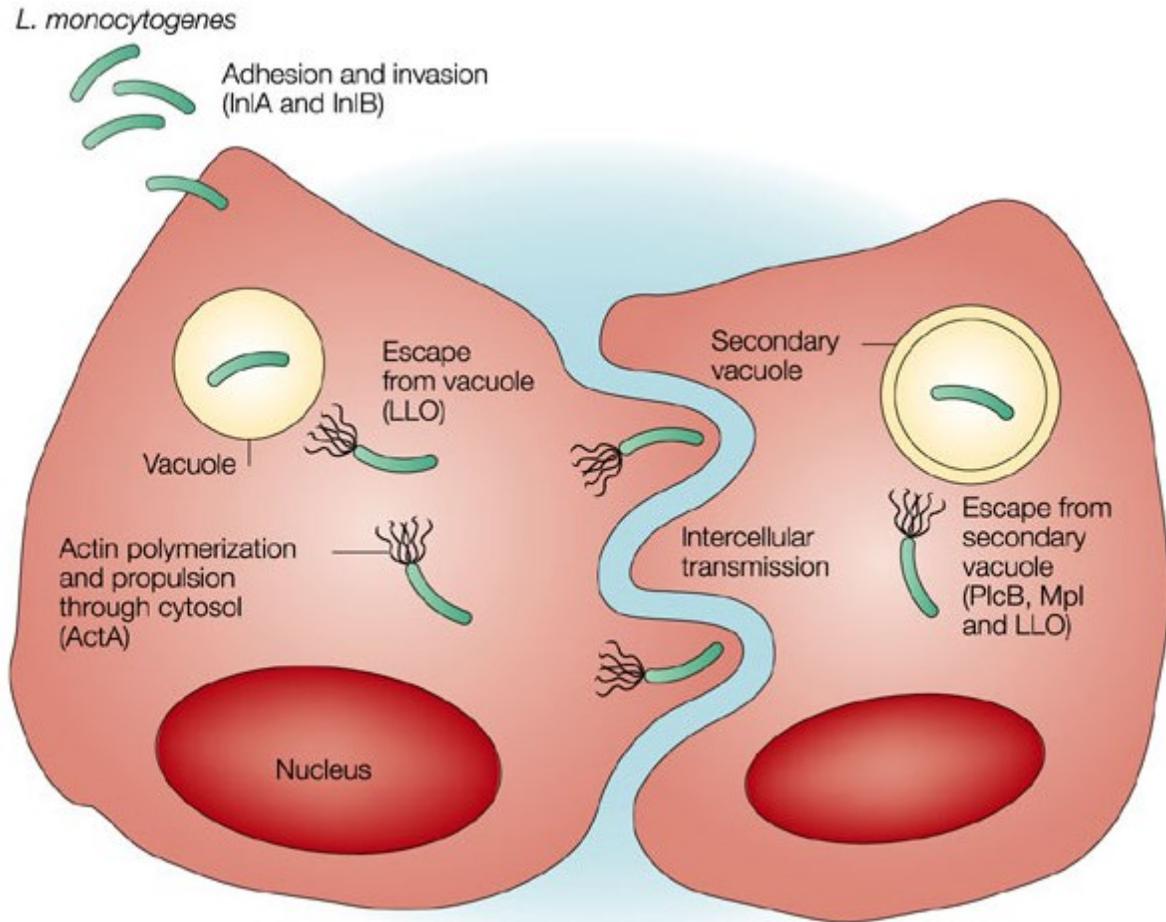


Figure 8: Entrée et propagation de *Listeria monocytogenes* dans les cellules.⁴⁸

Les neutrophiles et les macrophages sont les principaux médiateurs du contrôle d'une infection à *Listeria monocytogenes*. Les bactéries circulant librement dans le cytosol activent le TLR2 et le TLR5, qui transmettent leur signal au noyau par l'entremise de la protéine MyD88 pour activer le facteur de transcription NF- κ B, qui stimule la sécrétion de cytokines (Figure 9). Des produits microbiens sont relargués par les macrophages et, avec l'aide des cytokines sécrétées, ils contribuent au recrutement des monocytes. Ces monocytes se différencient en cellules dendritiques producteurs de TNF et de synthase de l'oxyde nitrique.

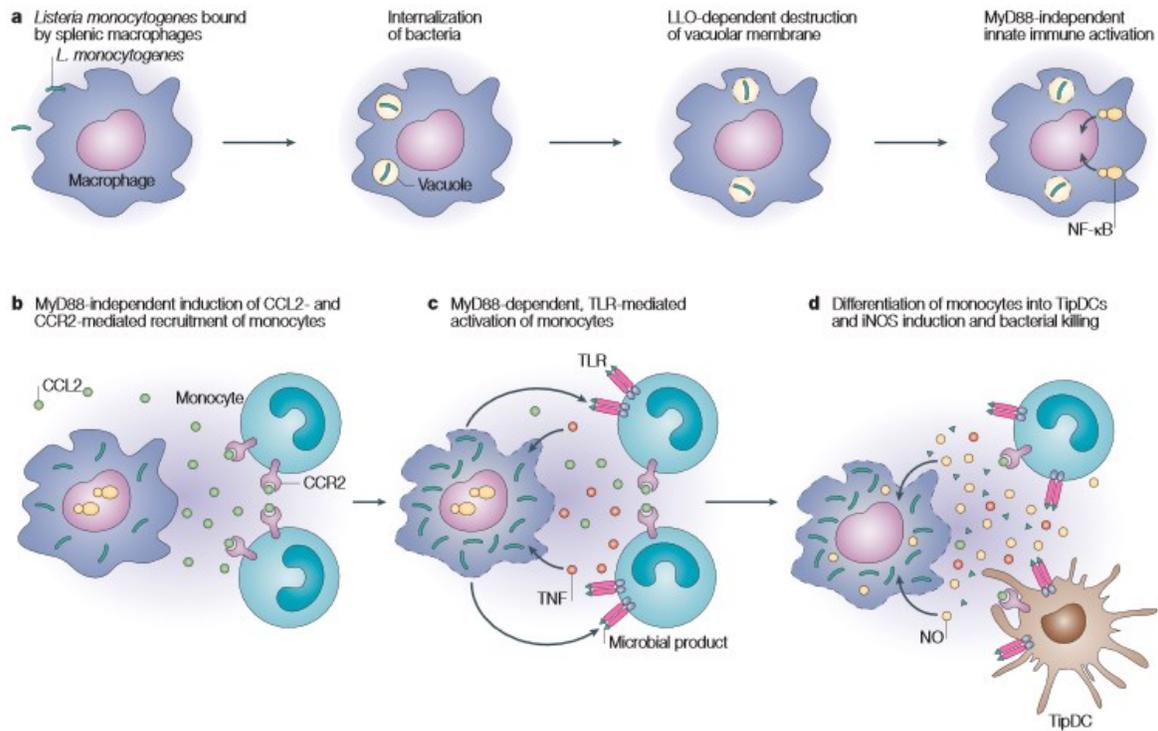


Figure 9: Activation de la réponse immunitaire par un macrophage infecté par *Listeria monocytogenes*.⁴⁸

1.3 L'étude du transcriptome et les variations fonctionnelles chez l'humain

Au cours des dernières années, de nombreux efforts ont été faits pour caractériser les bases génétiques des variations phénotypiques observées entre les individus. Par exemple, plusieurs études ont caractérisé des gènes spécifiques dans le but de mieux comprendre les maladies monogéniques ou complexes. De nombreuses études d'associations pangénomiques ont aussi associé des variants génétiques, souvent dans des régions non codantes, à des traits complexes.⁵⁹ Bien que ce type d'étude ait grandement contribué à améliorer notre compréhension du fonctionnement du génome humain, notre habileté à identifier le variant causal et le mécanisme biologique sous-jacent est encore limitée.

1.3.1 Régulation de l'expression génétique

Le niveau d'expression d'un transcrite est une mesure quantitative en partie liée aux variations de l'ADN qui caractérise l'expression de nombreux phénotypes.⁶⁰ Jusqu'à maintenant, la plupart des études pangénomiques sur l'expression ont été faites en combinant des données de génotypage et des données d'expression obtenues avec des micropuces. Ces études ont permis d'établir qu'une grande partie des variations transcriptionnelles sont contrôlées génétiquement.^{61,62}

Il existe deux types de variants affectant l'expression d'un gène : le variant cis, qui contrôle l'expression d'un seul allèle, et le variant trans, qui contrôle l'expression des deux allèles (Figure 10). Le variant cis est un polymorphisme situé sur le même chromosome que le gène qu'il régule et il est près de celui-ci. Par exemple, ce type de variant peut se trouver dans un promoteur ou une région *enhancer*. Le variant trans, quant à lui, se situe n'importe où dans le génome. Par exemple, il peut être situé dans un facteur de transcription ou dans un micro-ARN.

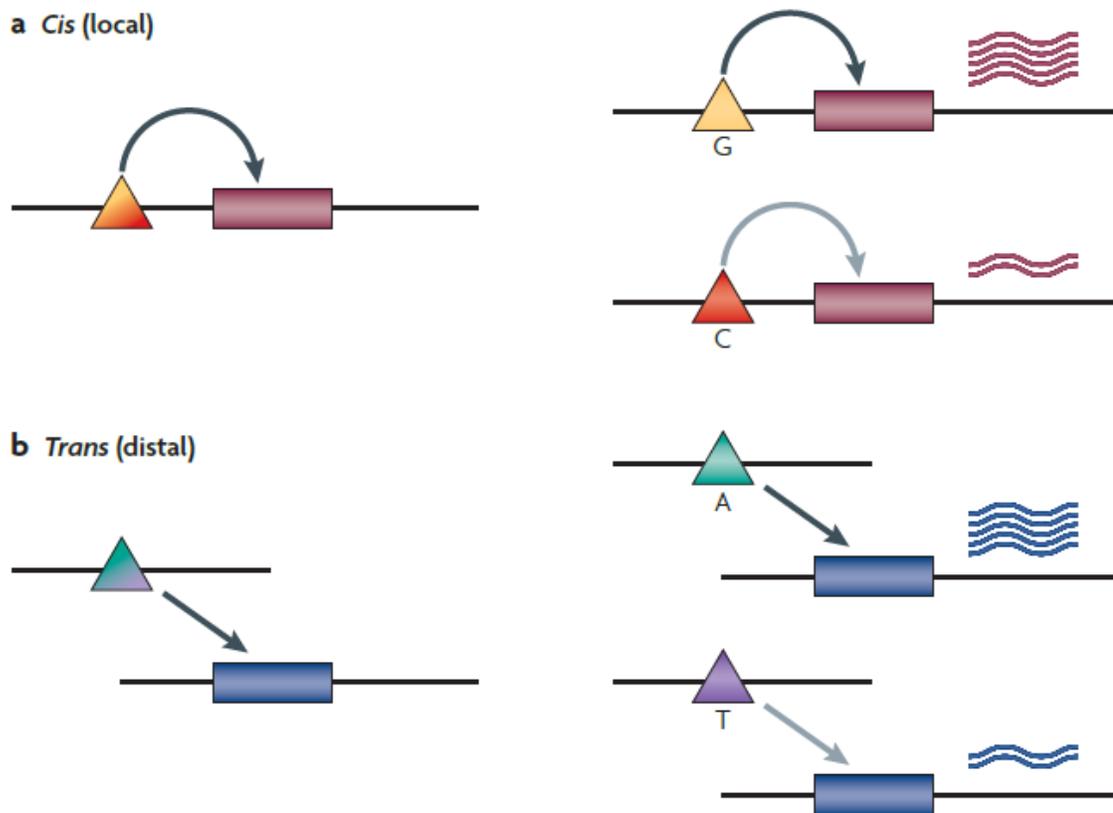


Figure 10: Effet d'un variant cis et trans sur l'expression d'un gène. A) Le polymorphisme se situe près du gène qu'il régule. Les individus porteurs du variant G ont un niveau d'expression du gène cible plus élevé que les individus porteurs du variant C. B) Le polymorphisme se situe loin du gène qu'il régule. Les individus porteurs du variant A ont un niveau d'expression du gène cible plus élevé que les individus porteurs du variant T.

Il est maintenant bien connu que l'expression d'une grande proportion de gènes est régulée par des polymorphismes situés dans des régions régulatrices cis.^{61,63-65} Il semble que le plus fort signal d'association en cis avec le niveau d'expression d'un gène se trouve près du site d'initiation de la transcription. À cause de la diminution du pouvoir statistique engendrée par le nombre élevé de tests, il est plus difficile d'associer les polymorphismes situés en trans avec l'expression d'un gène. Néanmoins, il a été démontré que les variants situés dans des régions régulatrices cis sont en plus susceptibles d'avoir des effets sur l'expression de certains gènes en trans, c'est-à-dire en aval du premier gène qu'ils régulent en affectant la cascade de signalisation.^{61,64} D'autres polymorphismes en

trans sont susceptibles d'être identifiés dans les régions codantes, dans les régions régulatrices de l'épissage alternatif, dans la région 3' des gènes et dans les micro-ARNs.

1.3.2 Génomique des populations de l'expression génétique

Étant donné que le niveau d'expression des gènes est contrôlé en partie génétiquement, il est raisonnable de penser que la composante populationnelle y joue un rôle.⁶² Les différences interpopulationnelles dans le niveau d'expression génétique sont significatives, comme l'ont démontré Stranger & al. en 2007. Ils ont mesuré le niveau d'expression pangénomique par micropuce dans les lignées cellulaires lymphoblastoïdes du *International HapMap Project* de quatre populations (habitants de l'Utah descendants des Européens du Nord et de l'Ouest (CEU), Yorubas du Nigéria en Afrique (YRI), Japonais (JPT), Chinois (CHB)). Ils ont situé la proportion de gènes significativement différemment exprimés entre n'importe quelle paire de populations entre 17 % et 29 %. Néanmoins, ces chiffres doivent être considérés avec prudence étant donné que la population CEU présente étonnement le plus de divergence, probablement parce que cette lignée cellulaire a été créée plus tôt que les trois autres⁶⁶.

Plus récemment, Lappalainen & al. ont mesuré le niveau d'expression pangénomique par séquençage du transcriptome dans les lignées cellulaires lymphoblastoïdes du *Consortium 1000 Genomes Project* de quatre populations européennes et une population africaine. Ils ont estimé que la proportion de gènes avec des niveaux d'expression différents ou une utilisation différente des transcrits d'ARN se situe entre 6 % et 40 % pour les comparaisons de n'importe quelle paire de populations européennes et entre 75 % et 85 % pour les comparaisons de n'importe quelle population européenne à la population africaine. La proportion élevée de gènes différemment exprimés et/ou différemment transcrits entre les populations européennes et la population africaine s'explique par une plus grande contribution de gènes différemment épissés entre les deux continents.

⁶⁷

1.3.2.1 Séquençage du transcriptome

La mesure des variations de l'expression génétique est un objectif clé, autant dans la compréhension des mécanismes biologiques complexes que dans l'estimation des variations phénotypiques interpopulationnelles. La compréhension du transcriptome est essentielle pour interpréter les éléments fonctionnels du génome et caractériser les constituants moléculaires des traits complexes. Le séquençage à haut débit du transcriptome est une technologie intéressante permettant de

mesurer le niveau d'expression d'un transcrit sur toute sa longueur. Plusieurs techniques ont été utilisées dans le passé pour mesurer l'expression des gènes, dont celle de la micropuce d'expression, qui utilise des sondes s'hybridant aux transcrits à mesurer. Bien que cette technique ait permis d'approfondir nos connaissances sur le génome humain, le séquençage à haut débit du transcriptome offre maintenant de nouvelles possibilités. Il offre une résolution permettant de connaître la séquence complète des transcrits tout en quantifiant l'ensemble des transcrits appartenant à un même gène ou à un même exon. Il offre aussi une échelle de détection beaucoup plus grande que l'utilisation des micropuces et permet la détection des différents isoformes et des nouveaux transcrits.⁶⁸ En 2010, Pickrell & al. ont démontré que le séquençage de nouvelle génération est un outil puissant dans la compréhension des mécanismes sous-jacents aux variations observées dans les niveaux d'expression chez les humains. Ils ont séquencé l'ARN des lignées cellulaires lymphoblastoïdes du *International HapMap Project* et ont identifié près de 1000 nouveaux transcrits présentant des évidences d'appartenance à des sites d'épissage alternatifs ainsi que plus de 115 nouveaux exons. Ils ont aussi identifié plus de 7 000 nouveaux sites de clivage polyadénylés putatifs, suggérant que la partie 3' UTR des gènes joue un rôle important dans la régulation. Ils ont confirmé, une fois de plus, que la plus forte association du niveau d'expression d'un gène avec un SNP se trouve près du gène, estimant que 90 % des SNPs qui influencent le niveau d'expression d'un gène se trouve à moins de 15 kb de ce gène. Ils ont validé le fait que les associations cis entre un SNP et l'expression d'un gène ont bien un effet allèle-spécifique dans la majorité des cas. Ils ont finalement mis en évidence des associations cis entre un SNP et le niveau d'expression des exons, indiquant que certains variants sont impliqués dans le mécanisme d'épissage alternatif.⁶⁹ Leurs résultats indiquent donc que le séquençage du transcriptome est un outil robuste pour analyser le niveau d'expression génétique et les variations dans l'épissage alternatif en plus de permettre d'améliorer l'annotation du génome.

2. Objectifs et hypothèse

2.1 Problématique

Dans ce projet, nous suggérons qu'il existe des différences d'ordre génétique entre les populations qui expliquent la différence de susceptibilité aux maladies infectieuses entre les populations d'origine européenne et africaine. Ces différences suggèrent des variabilités phénotypiques de la réponse immunitaire probablement causées par l'histoire évolutive bien différente des Européens et des Africains. Cette histoire évolutive pourrait avoir engendré des réponses immunitaires variables entre ces deux populations à cause d'évènements de sélection naturelle distincts.

Les scans pangénomiques pour détecter les régions du génome marquées par la sélection naturelle ont mis en évidence plusieurs gènes impliqués dans l'interaction hôte-pathogène. Néanmoins, ce genre de test identifie des régions génomiques plutôt que des gènes. Dans certains cas, ces gènes sont en fort déséquilibre de liaison avec d'autres gènes et aucune étude fonctionnelle n'a validé l'influence de la sélection naturelle sur la réponse immunitaire, ce qui implique qu'ils ne sont peut-être pas la cible réelle de la sélection naturelle. Pour mieux comprendre l'effet de la sélection naturelle sur les mécanismes immunologiques, il est maintenant nécessaire d'utiliser une approche permettant d'intégrer l'information phénotypique aux signatures de sélection naturelle. Le séquençage à haut débit du transcriptome combiné à des scans pangénomiques pour détecter la sélection naturelle permet de quantifier le degré de variabilité interindividuelle et interpopulationnelle dans la réponse à une infection, puis d'évaluer la participation de la sélection naturelle dans ce phénomène. Cette méthode permet donc d'identifier précisément des gènes candidats qui sont situés en déséquilibre de liaison sur de longs haplotypes marqués par la sélection naturelle. Pour la première fois, nous avons utilisé cette approche afin d'étudier la variabilité de la réponse immunitaire à des infections in vitro entre les Afro-Américains et les Américains d'origine caucasienne.

2.2 Hypothèse

Les variations interpopulationnelles dans l'expression de gènes après une infection bactérienne peuvent expliquer des différences dans la susceptibilité aux maladies infectieuses.

2.3 Objectifs

- 1- Caractériser les différences de la réponse immunitaire des macrophages à une infection avec *Salmonella typhimurium* et *Listeria monocytogenes*.
- 2- Caractériser la variation de la réponse immunitaire aux infections bactériennes entre les populations d'origine européenne et africaine.
- 3- Identifier des signatures de sélection naturelle parmi les gènes différemment régulés en réponse aux infections bactériennes entre les populations d'origine européenne et africaine.

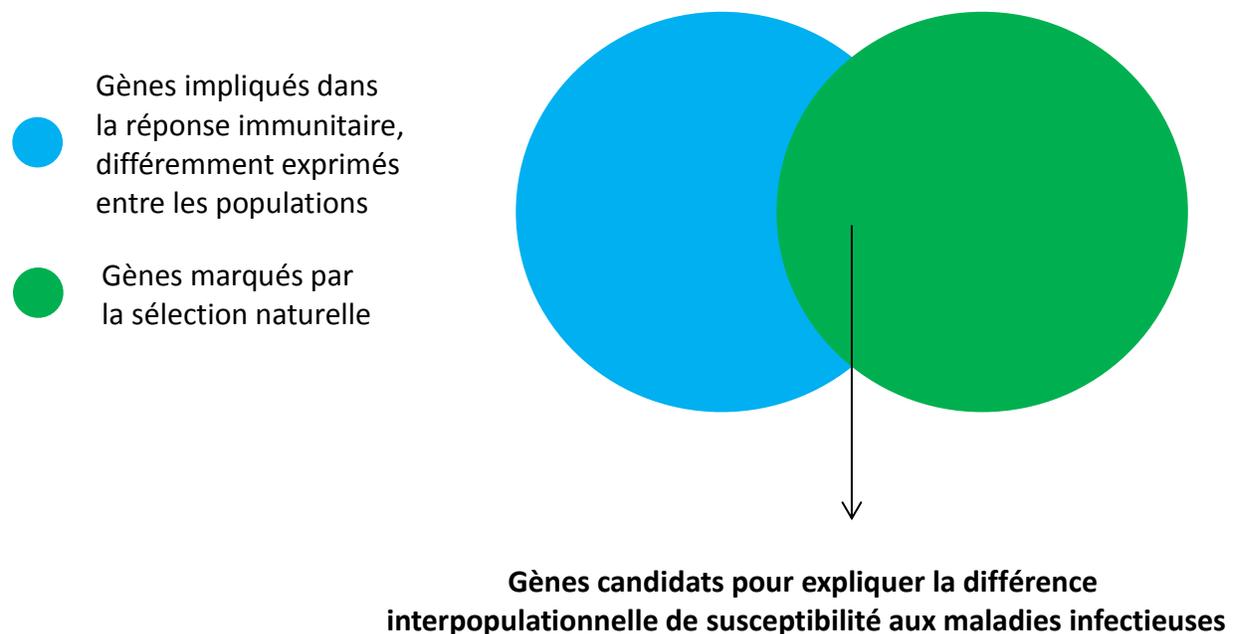


Figure 11: Objectifs du projet de recherche.

3. Matériels et méthodes

Nous avons étudié la variabilité de la réponse immunitaire à des infections bactériennes in vitro entre les Afro-Américains et les Américains d'origine européenne. Nous avons préféré utiliser des échantillons provenant des États-Unis, plutôt que d'Afrique ou d'Europe, afin de faciliter la collecte et

pour éliminer le biais qu'aurait introduit des échantillons provenant de lieux différents. Pour faciliter la lecture, j'utiliserai les termes « Européens » et « Africains », bien qu'en réalité il s'agisse plutôt d'Américains d'origine européenne et africaine. Nous avons utilisé l'approche expérimentale illustrée à la Figure 12. Brièvement, les monocytes ont été purifiés par sélection positive des buffy coat de 31 individus d'origine européenne et de 30 individus d'origine africaine, puis ils ont été dérivés en macrophages. Ils ont été infectés avec deux bactéries, *Salmonella typhimurium* et *Listeria monocytogenes*. Après 4 heures d'infection, l'ARNm des macrophages infectés et non infectés a été extrait et séquencé pour quantifier les niveaux d'expression génétique. Les cytokines sécrétées dans le surnageant ont aussi été quantifiées. De plus, l'efficacité de la lyse des bactéries par les macrophages a été mesurée en comptant le nombre de bactéries intracellulaires à 2, 4 et 24 heures post-infection. En parallèle, chaque individu a été génotypé. Cette méthode nous a permis d'identifier les gènes impliqués dans la réponse immunitaire aux infections bactériennes et d'identifier les gènes différemment exprimés entre les populations à la suite des infections bactériennes (Figure 11). Ultiment, cette méthode nous a permis d'évaluer la contribution de la sélection naturelle dans l'explication de la variabilité de la réponse immunitaire aux infections bactériennes observées entre les populations d'origine européenne et africaine.

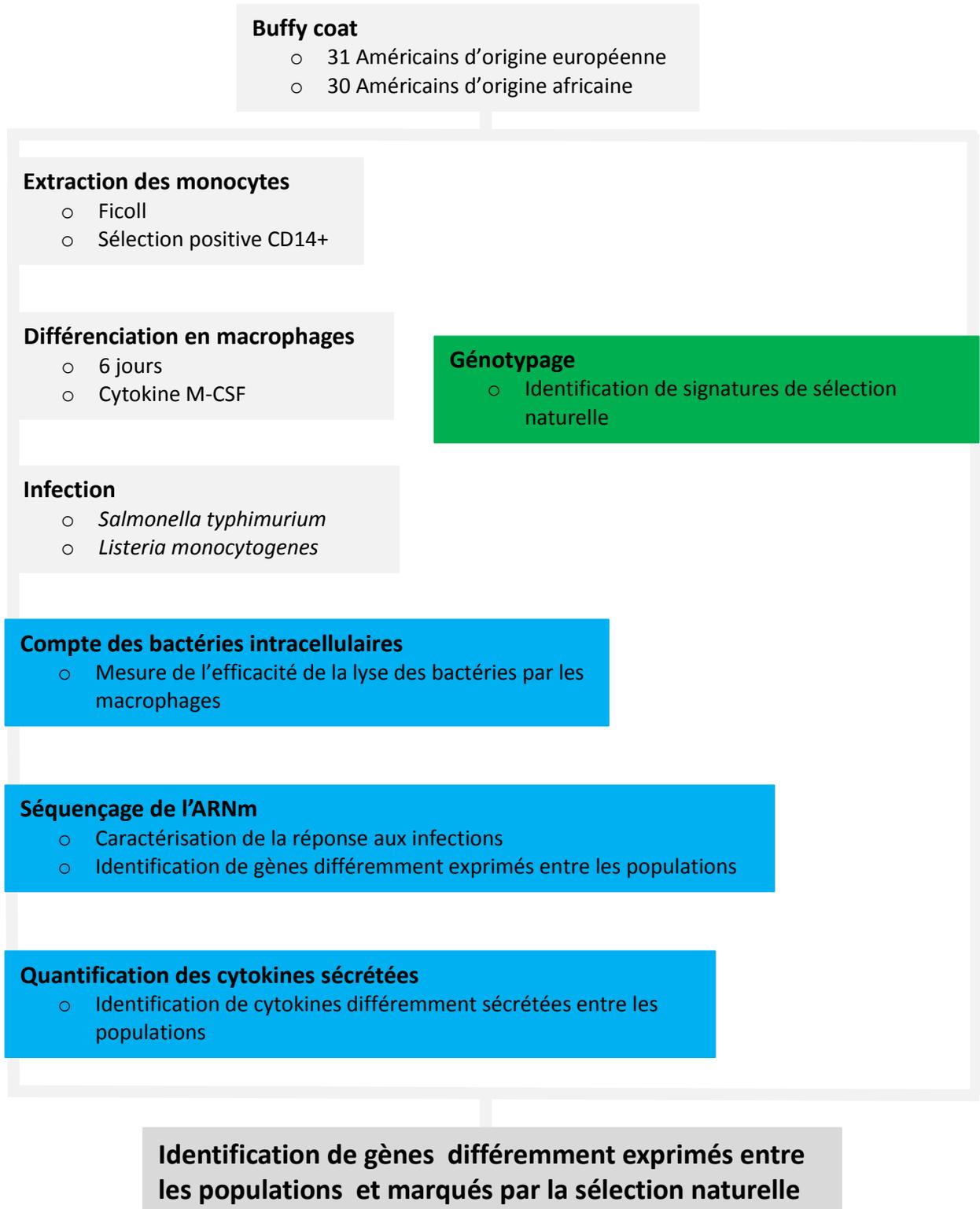


Figure 12: Schéma expérimental.

3.1 Culture cellulaire

3.1.1 Extraction et purification

Les buffy coat nous ont été fournis par la compagnie Indiana Blood Center. Les cellules mononucléées du sang périphérique ont été séparées en utilisant le Ficoll-Paque PREMIUM (GE Healthcare Life Sciences). Les monocytes ont ensuite été purifiés à l'aide de CD14 MicroBeads (Miltenyi Biotec) avec le robot auto MACS Pro Separator (Miltenyi Biotec). Les CD14 MicroBeads sont des microbilles magnétiques conjuguées à un anticorps monoclonal antihumain CD14. Les cellules mononucléées du sang périphérique ont été incubées avec un ratio de 10 μ L de microbilles CD14 par 10^7 cellules pendant 15 minutes à 4°C. Elles ont ensuite été déposées sur une colonne magnétique puis lavées 3 fois avec l'autoMACS Pro Washing Solution (Miltenyi Biotec). La fraction positive CD14 a été éluée dans l'autoMACS Running Buffer (Miltenyi Biotec).

3.1.2 Différenciation des monocytes en macrophages

Les monocytes ont été mis en culture dans des pétris pour être différenciés en macrophages dans un milieu RPMI-1640 1X (Hyclone) complété avec 10 % de sérum bovin foetal FBS Canadian Origin Gold (VWR), de la L-glutamine (Hyclone), 1 μ g/mL de gentamicine (Gibco) et 20 ng/mL de M-CSF (Cederlane) à un ratio de 2 millions de cellules pour 3 mL de milieu de culture. La moitié du milieu de culture a été changé au jour 2 et 4 post-purification.

3.1.3 Infection

La veille de l'infection, au jour 6, les macrophages ont été décollés des pétris avec du Cellstripper (Corning cellgro), lavés 2 fois avec du PBS 1X (MultiCell Technologies), puis mis en culture dans du milieu complété avec 10 % de sérum bovin foetal, de la L-glutamine et 20 ng/mL de M-CSF à un ratio de 1,5 million de cellules par mL de milieu de culture. Les cellules ont été étalées à une densité de 1 million de cellules par puits d'une plaque 12 puits ou 2 millions de cellules par puits d'une plaque 6 puits. L'expression à la membrane cellulaire de différentes protéines a été vérifiée par cytométrie en flux afin de valider la différenciation des monocytes en macrophages. Les cellules ont été incubées à 4 °C avec des anticorps monoclonaux PE Mouse Anti-Human CD14, CD1A, CD83 et HLA-DR conjugués à la phycoérythrine (BD Biosciences). Les cellules ont été lavées dans du PBS 1X, puis resuspendues dans une solution de PBS 1X contenant 2 % de formaldéhyde (Affymetrix). L'expression à la

membrane cellulaire des différentes protéines a été mesurée et analysée par cytométrie en flux sur le FACS Calibur flow cytometer (BD Biosciences).

La veille de l'infection, les bactéries *Salmonella typhimurium* et *Listeria monocytogenes* ont été décongelées et mises en culture dans du bouillon trypticase soja (BD) à 37 °C avec agitation pour la nuit. Le jour de l'infection, la densité optique (DO) des bouillons de bactéries a été mesurée sur l'appareil Epoch (BioTek) jusqu'à l'obtention d'une DO de 0,5. Les bouillons de bactéries ont ensuite été dilués dans du milieu de culture cellulaire complété avec 10 % de sérum bovin fœtal et de la L-glutamine, jusqu'à l'obtention d'une concentration de 30 millions de *Salmonella typhimurium* par mL de milieu de culture et 15 millions de *Listeria monocytogenes* par mL de milieu de culture. Les dilutions nécessaires pour l'obtention de ces concentrations de bactéries ont préalablement été calculées, lors d'essais préliminaires, en plaquant les dilutions sur des géloses trypticase soja (BD). Les géloses ont été incubées à 37 °C pendant 24 heures. Les colonies ayant poussé sur les géloses ont été comptées afin d'estimer la concentration de bactéries. Les macrophages ont été inoculés avec les dilutions de bactéries à un *multiplicity of infection* (MOI) de 10 pour *Salmonella typhimurium* et de 5 pour *Listeria monocytogenes* tandis que les puits correspondant à la condition non infectée ont été supplémentés avec du milieu de culture complété avec 10 % de sérum bovin fœtal et de la L-glutamine. Le MOI correspond au ratio du nombre de bactéries par rapport au nombre de cellules au moment de l'inoculation. Les macrophages ont été centrifugés à 500 g pendant 5 minutes, puis incubés à 37 °C pendant 40 minutes. Les macrophages ont été lavés avec du PBS 1X, puis ont été incubés 1 heure à 37 °C avec du milieu de culture complété avec 10 % de sérum bovin fœtal, de la L-glutamine et 50 µg/mL de gentamicine. L'antibiotique est nécessaire à cette étape de l'infection pour éliminer les bactéries extracellulaires. Après 1 heure, les macrophages ont été lavés deux fois avec du PBS 1X. Les puits destinés aux comptes des bactéries intracellulaires ont été lysés avec une solution d'eau nanopure contenant 1 % de Triton X-100 (Sigma-Aldrich). Les macrophages restants ont été incubés 2 heures dans du milieu de culture complété avec 10% de sérum bovin fœtal, de la L-glutamine, 3 µg/mL de gentamicine et 20 ng/mL de M-CSF. Après 2 heures d'infection, le surnageant a été retiré des puits et conservé. Les macrophages ont été lavés avec du PBS 1X. Les puits destinés aux comptes de bactéries intracellulaires ont été lysés avec une solution d'eau nanopure contenant 1 % de triton et les puits destinés à la récolte de l'ARNm ont été lysés dans 1 mL de Qiazol (QIAGEN).

Chaque lysat de cellules contenant les bactéries a été dilué en série dans du bouillon trypticase soja dans un volume total de 1 mL. Chaque dilution en série a été faite en duplicata. Deux dilutions de chaque dilution en série ont été plaquées sur des géloses trypticase soja (BD), puis incubés à 37 °C pendant 24 heures. Les colonies ayant poussé sur les géloses ont été comptées afin d'estimer le nombre de bactéries persistant à l'intérieur des macrophages.

3.2 Séquençage de l'ARNm

3.2.1 Préparation des bibliothèques pour le séquençage de l'ARNm

L'ARN des macrophages infectés et non infectés a été extrait avec l'ensemble miRNeasy Mini Kit (QIAGEN). Les bibliothèques d'ARNm ont été préparées à l'aide de l'ensemble TruSeq RNA Sample Prep Kit v2 (Illumina). Brièvement, l'ARNm a été purifié par sélection poly-A, chimiquement fragmenté, puis converti en ADNc simple-brin avec des amorces d'hexamères aléatoires. Le second brin a été créé pour générer un ADNc double-brin. Des bouts francs ont été générés à partir de l'ADNc, puis une base « A » a été ajoutée au bout franc de chaque brin afin de les préparer pour la ligation des adaptateurs de séquençage. Chaque adaptateur contient une base « T » pour s'hybrider à la base « A » de l'ADNc. Ces adaptateurs contiennent la séquence complémentaire des amorces d'amplification PCR et une séquence unique permettant de multiplexer plusieurs échantillons. Une étape d'amplification PCR a finalement complété le protocole de préparation des bibliothèques. Les bibliothèques ont été séquencées *single-end*, 100 pb, avec l'appareil HiSeq 2000 (Illumina), au Centre d'innovation Génome Québec et Université McGill.

3.2.2 Description de la technique de séquençage de l'ARNm

Six bibliothèques ont été multiplexées par canal de *flowcell*. Des fragments monocaténaux des bibliothèques ont été liés de façon aléatoire à la surface des canaux de la *flowcell*. Des nucléotides non marqués ont été ajoutés pour amorcer une amplification en forme de pont par l'entremise des amorces présentes à la surface des canaux. Une fois le pont double brin d'ADN formé, une étape de dénaturation a suivi pour créer deux molécules d'ADN simple brin. Le processus d'amplification en pont a été répété plusieurs fois jusqu'à l'obtention de millions de grappes denses, remplissant ainsi la *flowcell*. Les cycles de séquençage ont ensuite commencé avec l'ajout des 4 bases marquées, d'amorces et d'ADN polymérase. Après une excitation au laser, la fluorescence émise par l'addition chaque base marquée a été capturée et identifiée par un ordinateur.^{68,70}

3.2.3 Analyse des données de séquençage de l'ARNm

Les données de séquençage brutes ont été fournies par le Centre d'innovation Génome Québec et Université McGill. La qualité des séquences a été vérifiée avec le programme FastQC.⁷¹ Les séquences ont été alignées avec le programme Tophat2 contre la référence du génome humain hg19 (Human Genome version 19).⁷² Nous avons obtenu en moyenne 34 millions de séquences par échantillon et nous avons aligné en moyenne 31 millions de séquences par échantillon, ce qui correspond à un pourcentage d'alignement moyen de 92 %. Un échantillon a dû être retiré à cause d'un nombre de séquences trop faible. Le niveau d'expression de chaque gène a été quantifié avec le programme HTseq à l'aide du fichier de définition iGenomes du 30-08-2011.⁷³ Les données ont été normalisées pour le taux de bases G et C avec la librairie EDAsq dans le programme R.⁷⁴ Le programme DESeq a finalement normalisé les données pour le nombre total de séquences et a calculé, par gène, le *fold change* (FC) et le *false discovery rate* (FDR) entre deux conditions en modélant les données sur une distribution binomiale négative.⁷⁵

3.3 Génotypage de l'ADN

3.3.1 Préparation de l'ADN

L'ADN de chaque individu a été extrait des cellules de la fraction négative (les cellules mononucléées du sang périphérique sans les monocytes) obtenue à l'étape de purification des monocytes avec l'ensemble Puregene Blood Core Kit B (QIAGEN). L'ADN a été génotypé sur des puces HumanOmni5Exome BeadChip ou HumanOmni2.5 (Illumina) au Centre d'innovation Génome Québec et Université McGill. La puce HumanOmni5Exome comprend 4,5 millions de marqueurs en plus de 250 000 marqueurs du contenu exonique représentant différentes populations (dont Européens et Africains). La puce HumanOmni2.5 comprend 2 015 318 marqueurs représentant différentes populations (dont Européens et Africains).

3.3.2. Description de la technique de génotypage de l'ADN

Les puces de génotypage BeadChip sont recouvertes de centaines de milliers de billes colorées sur lesquelles un oligonucléotide spécifique est lié. Ces oligonucléotides ont joué un rôle de sonde en s'hybridant à une séquence spécifique de l'ADN et en permettant ainsi sa reconnaissance en libérant un signal fluorescent qui a été détecté et identifié par un ordinateur.⁷⁶

3.3.3 Analyse des données de génotypage de l'ADN

Les données de génotypage brutes ont été traitées au Centre d'innovation Génome Québec et Université McGill par le programme GenomeStudio (Illumina). Les données ont ensuite été analysées à l'aide du logiciel Plink, et les SNPs avec une fréquence allélique mineure (MAF) inférieure à 5 % parmi tous les individus ont été éliminés.⁷⁷

3.3.3.1 Calcul de l'ethnicité

Nous savons que les Américains d'origine africaine peuvent aussi avoir des ancêtres d'origine européenne.⁷⁸ Nous avons donc estimé la proportion du génome d'origine africaine chez chaque individu. Nous avons utilisé le programme Admixture en utilisant un nombre de population ancestrale de 2.⁷⁹

3.4 Quantification des cytokines sécrétées

Les niveaux de cytokines sécrétées dans le surnageant des cellules infectées et non infectées ont été mesurés avec le Bio-Plex Pro Human Cytokine 27-plex Assay (Bio-Rad). Brièvement, les microbilles multiplexées, combinées à des anticorps de détection, ont été distribuées dans une plaque 96 puits. Elles ont été lavées deux fois avec une solution tampon de lavage. Les surnageants non dilués ainsi que les solutions standards ont été ajoutés dans les puits, puis incubés à température de la pièce avec agitation pendant 30 minutes. Chaque puits a été lavé trois fois avec une solution tampon de lavage. L'anticorps de détection a été ajouté, puis la plaque a été incubée à température de la pièce avec agitation pendant 30 minutes. Chaque puits a été lavé trois fois avec une solution tampon de lavage. Le conjugué streptavidine-phycoérythrine a été ajouté, puis la plaque a été incubée à température de la pièce avec agitation pendant 10 minutes. Chaque puits a été lavé trois fois avec une solution tampon de lavage. Les microbilles ont finalement été resuspendues dans une solution tampon, puis agitées pendant 30 secondes. Les niveaux de cytokines sécrétées de chaque surnageant ont été mesurés en duplicata sur l'appareil Bio-Plex MAGPIX Multiplex Reader (Bio-Rad).

3.4.1 Description de la technique de quantification des cytokines sécrétées

Le principe de cette technique est très similaire au principe de l'ELISA. Les microbilles sont magnétiques et couplées de façon covalente à un anticorps dirigé contre le biomarqueur d'intérêt. Plusieurs microbilles de couleurs différentes et reconnaissant des biomarqueurs différents ont été multiplexées ensemble. Les microbilles ont été mises en solution dans le surnageant contenant les

cytokines à quantifier afin qu'elles se lient à leur cytokine cible. Un anticorps de détection biotinylé a été ajouté à la solution pour créer un complexe ayant la forme d'un sandwich. Finalement, un conjugué streptavidine-phycoérythrine a été ajouté à la solution pour qu'il se lie à la biotine, formant ainsi le complexe final. La plaque a été ensuite lue sur l'appareil Bio-Plex MAGPIX Multiplex Reader. Un laser a excité chaque microbille afin de les classer par couleur (et donc par biomarqueur) tandis qu'un autre laser a excité la phycoérythrine afin de détecter le signal rapporteur. La quantité de fluorescence émise par chaque type de microbille a ensuite été comparée à la courbe étalon, qui a été calculée à partir des standards, pour quantifier le niveau de chaque cytokine dans les surnageants.⁸⁰

3.4.2 Analyse des données quantification des cytokines sécrétées

Les données brutes générées par le Bio-Plex MAGPIX Multiplex Reader ont subi une normalisation quantile, puis ont été intégrées dans un modèle de régression linéaire. Ce modèle a permis de comparer les niveaux de cytokines sécrétées entre les populations.

4. Résultats

4.1 Caractérisation de la réponse immunitaire des macrophages aux infections bactériennes

4.1.1 Analyse en composantes principales

Nous avons obtenu un total de 181 échantillons dont nous avons mesuré les niveaux d'expression par séquençage de l'ARNm. Chacun de nos individus (31 Européens et 30 Africains) a été séquençé dans les 3 conditions : non infectée, *Listeria monocytogenes* et *Salmonella typhimurium*. Nous avons éliminé 2 échantillons qui ne passaient pas les filtres de qualité à cause d'un nombre de séquences trop faible. Les données de séquençage du transcriptome des macrophages infectés et non infectés démontrent que *Salmonella typhimurium* et *Listeria monocytogenes* ont induit une réponse immunitaire spécifique. Pour chaque gène, le nombre de séquences obtenues par séquençage du transcriptome a été corrigé pour le nombre total de séquences par échantillon avec le programme DESeq1.⁷⁵ Puisque les différents échantillons ont été séquençés sur plusieurs *flowcells*, nous avons régressé l'effet du *batch effect* dans un modèle de régression linéaire. Les valeurs normalisées du nombre de séquences par gène de nos échantillons ont été utilisées pour tracer une analyse des composantes principales (Figure 13). La composante principale 1, qui explique 77 % de la variance, sépare bien les échantillons non infectés des échantillons infectés. Elle permet en plus de distinguer les échantillons infectés avec *Salmonella typhimurium* des échantillons infectés avec *Listeria monocytogenes*.

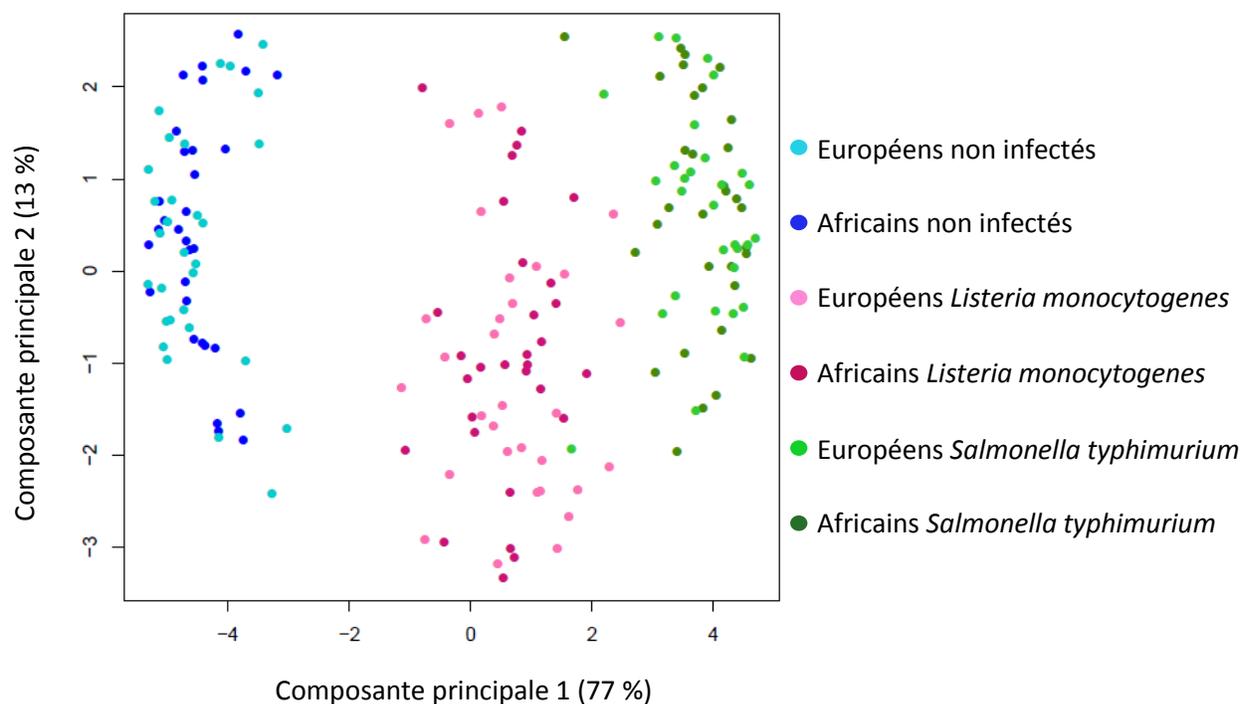


Figure 13: Analyse en composantes principales de l'ensemble des gènes exprimés par les macrophages non infectés, infectés avec *Salmonella typhimurium* et infectés avec *Listeria monocytogenes* de nos 181 échantillons.

4.1.2 Caractérisation des différences dans les niveaux d'expression génétique entre la condition non-infectée et les conditions infectées

Pour mieux caractériser les différences entre les 3 conditions (non infectée, *Salmonella typhimurium* et *Listeria monocytogenes*), nous avons utilisé le programme DESeq2, qui a calculé le *fold change* (FC) et le *false discovery rate* (FDR) de chaque gène entre 2 conditions. Le FDR est une méthode statistique utilisée lors du test de plusieurs hypothèses qui permet de corriger la valeur p pour les tests multiples. Dans notre modèle statistique, nous avons inclus les paramètres suivants : Niveau d'expression d'un gène = Condition + Ethnicité + *Batch effect*.⁷⁵ Nous avons considéré les gènes comme étant différemment exprimés lorsque qu'ils avaient un FDR inférieur à 1 % et une valeur absolue de $\log_2(\text{FC})$ supérieure à 1. Nous avons identifié 1 447 et 2 236 gènes *up-régulés*, respectivement par *Listeria monocytogenes* et *Salmonella typhimurium*, tandis que nous avons

identifié 1025 et 1789 gènes *down-régulés*, respectivement par *Listeria monocytogenes* et *Salmonella typhimurium* (Figures 14-15).

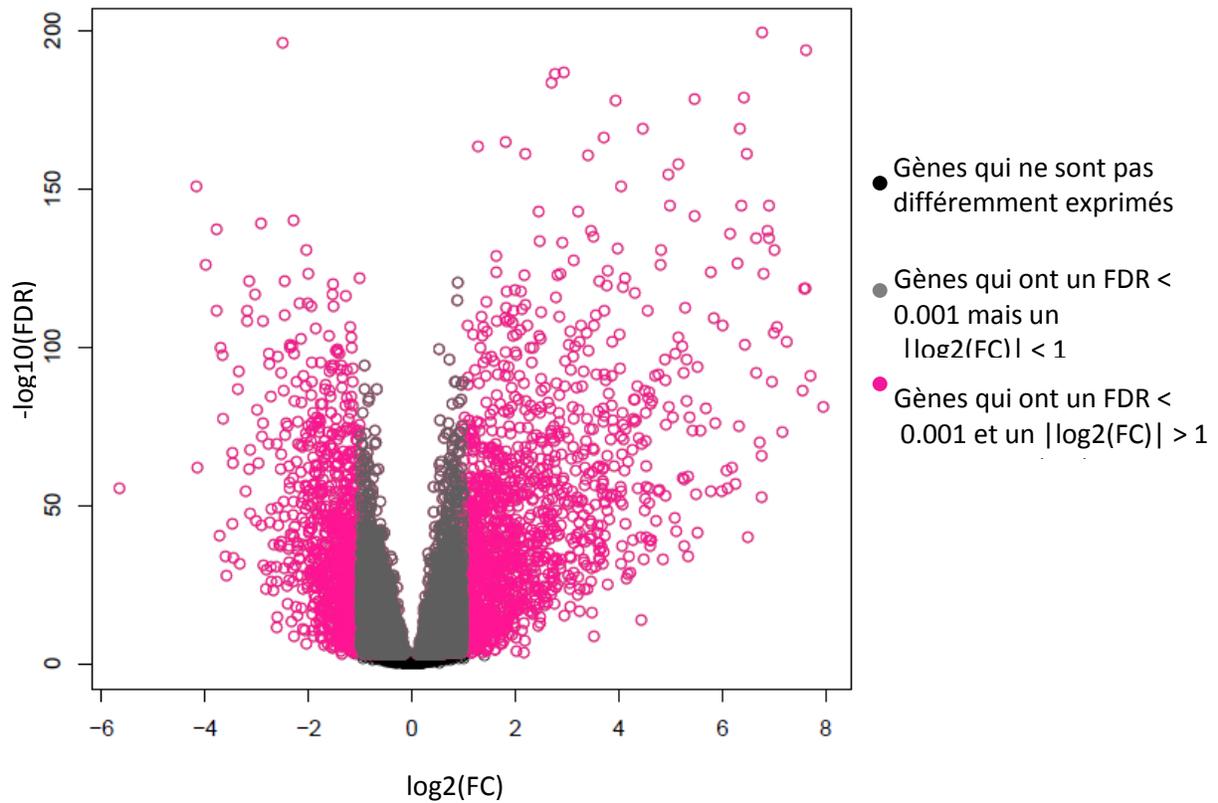


Figure 14: Graphique de type Volcano mettant en évidence les gènes différemment exprimés ($\text{FDR} < 0.001$ et valeur absolue de $\log_2(\text{FC}) > 1$) entre les conditions *Listeria monocytogenes* et non infectée.

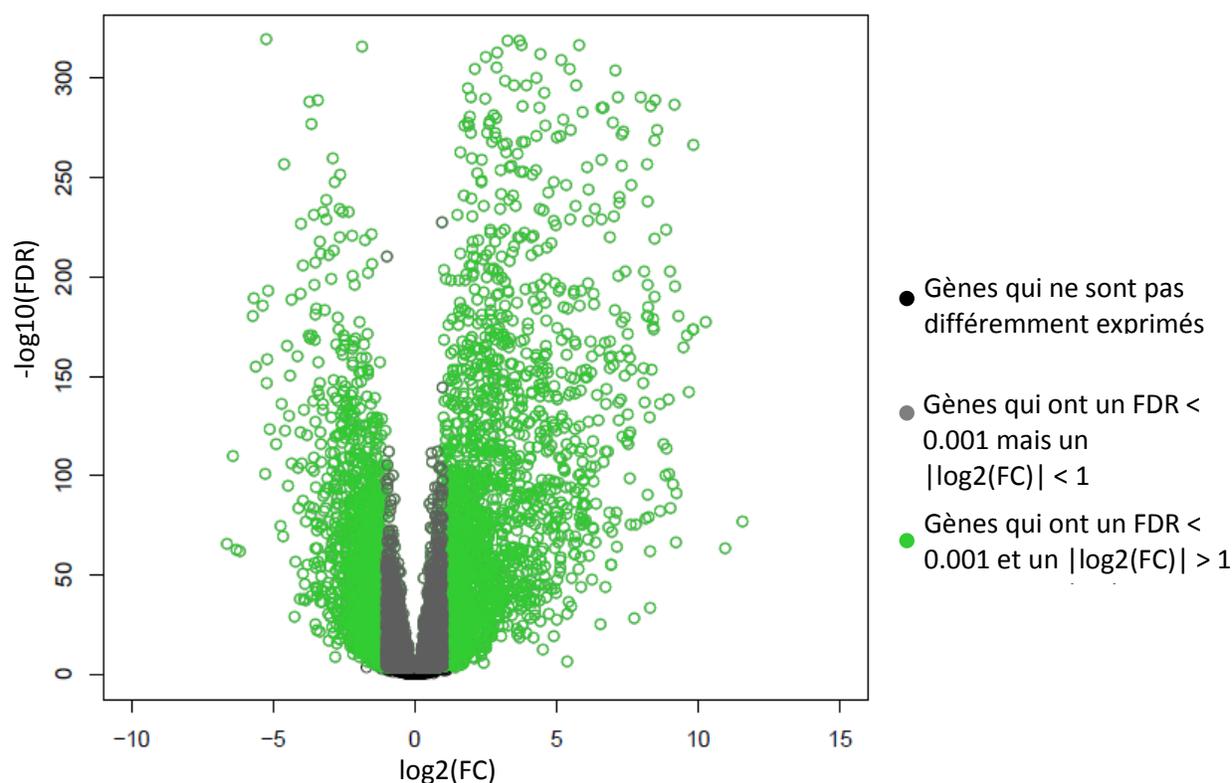


Figure 15: Graphique de type Volcano mettant en évidence les gènes différemment exprimés (FDR < 0.001 et valeur absolue de $\log_2(\text{FC}) > 1$) entre les conditions *Salmonella typhimurium* et non infectée.

Afin de mieux comprendre le rôle des gènes différemment exprimés suite aux infections bactériennes, nous avons fait une analyse de *Gene Ontology* (GO) des groupes de gènes suivants : les gènes *up-régulés* par l'infection à *Listeria monocytogenes* ou *Salmonella typhimurium* (Figure 16) et les gènes *down-régulés* par l'infection à *Listeria monocytogenes* ou *Salmonella typhimurium* (Figure 17). Pour ces analyses, nous avons utilisé le programme GeneTrail et nous avons conservé uniquement les 20 termes GO les plus significatifs.⁸¹ Ces analyses démontrent bien que les gènes *up-régulés* par les infections sont enrichis pour des fonctions liées à l'immunité, et ce, indépendamment de la bactérie observée. On y retrouve entre autres les termes « *immune system process* », « *cytokines activity* » et « *inflammatory response* ». En comparant les termes GO associés aux gènes *down-régulés* par l'une ou l'autre des bactéries, nous avons constaté que, contrairement aux gènes *up-régulés*, ils sont enrichis pour accomplir des fonctions différentes. Par exemple, *Salmonella typhimurium* semble avoir un effet inhibiteur sur la transcription puisqu'on y retrouve les

termes « *negative regulation of transcription* », « *nucleoside, nucleotide and nucleic acid metabolic process* » et « *DNA binding* ». De son côté, *Listeria monocytogenes* semble plutôt avoir un effet inhibiteur sur l'expression membranaire de plusieurs types de récepteurs puisqu'on y retrouve les termes « *G-protein coupled receptor activity* », « *transforming growth factor beta receptor complex* » et « *transmembrane-ephrin receptor activity* ».

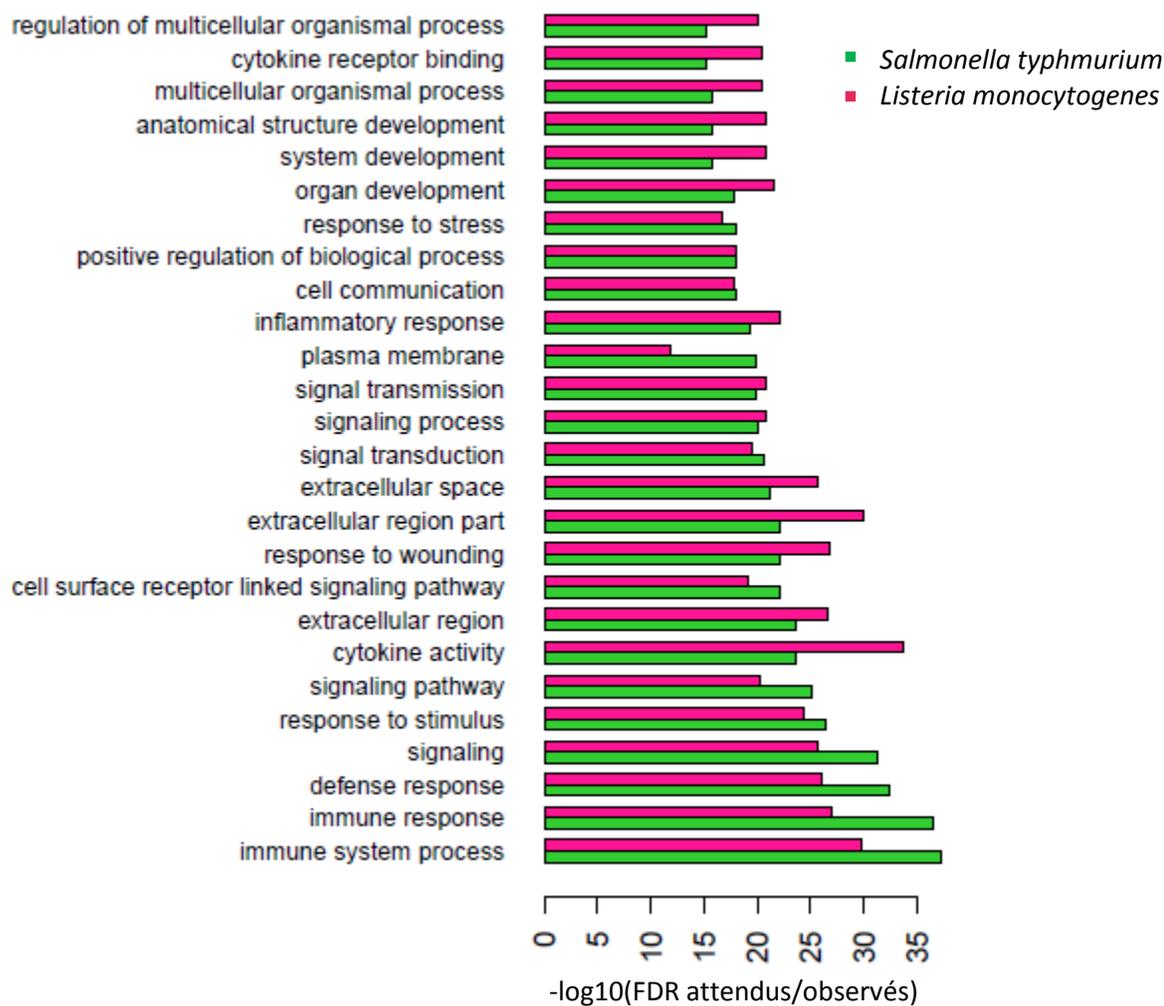


Figure 16: Graphique à barres des 20 termes GO les plus significatifs (FDR < 5 %) pour les gènes *up-régulés* suite aux infections bactériennes.

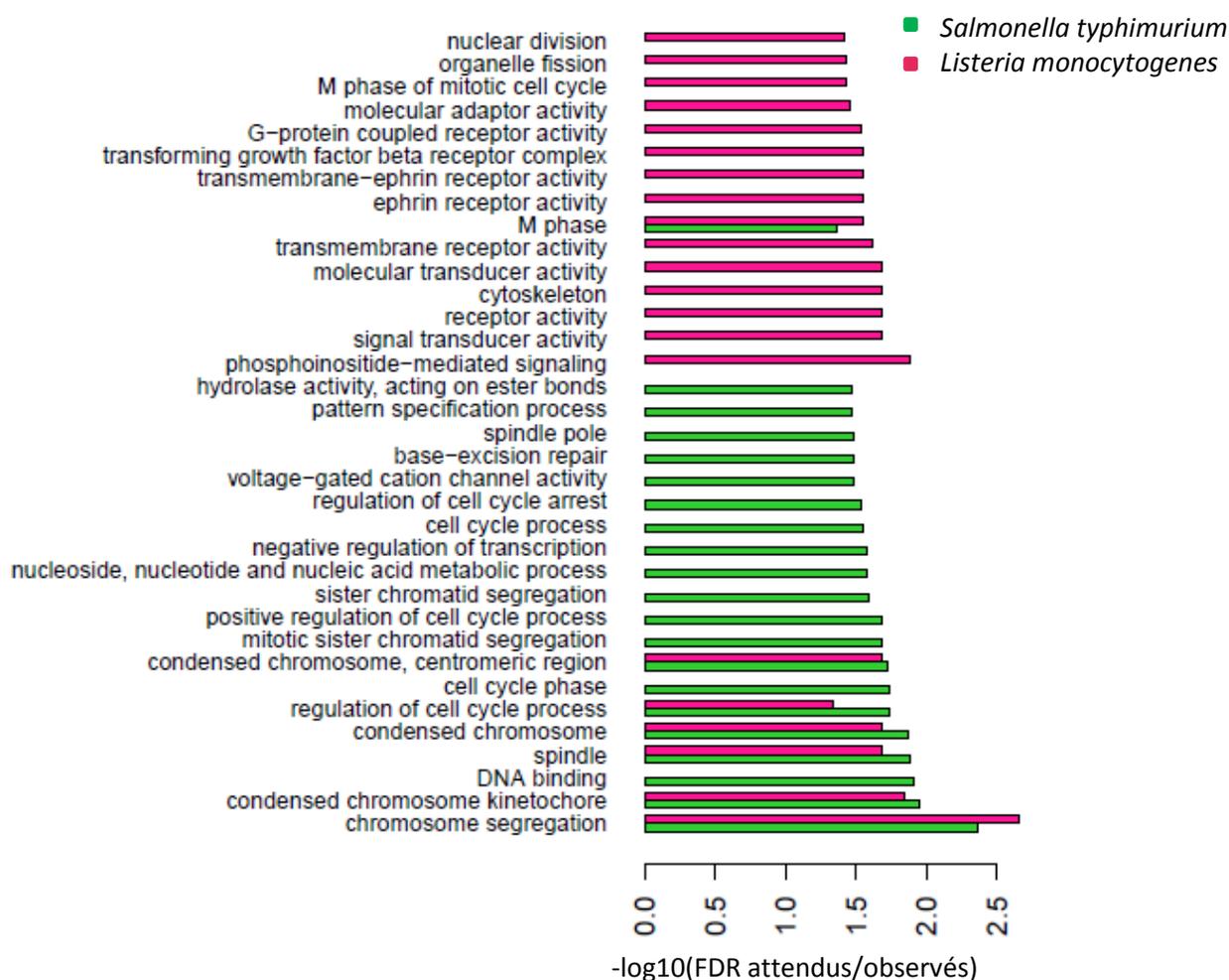


Figure 17: Graphique à barres des 20 termes GO les plus significatifs (FDR < 5 %) pour les gènes *down-régulés* suite aux infections bactériennes.

4.1.3 Caractérisation des différences dans les niveaux d'expression génétique entre les conditions *Listeria monocytogenes* et *Salmonella typhimurium*

Afin de mieux caractériser les gènes qui sont spécifiquement régulés par l'une ou l'autre des bactéries, nous avons calculé le FC et le FDR entre les conditions *Salmonella typhimurium* et *Listeria monocytogenes*. Un $\log_2(\text{FC})$ positif signifie que le gène est plus exprimé lors d'une infection avec *Salmonella typhimurium* que lors d'une infection avec *Listeria monocytogenes*. Nous avons identifié

1 447 gènes significativement plus exprimés par une infection avec *Listeria monocytogenes* et nous avons identifié 1789 gènes significativement plus exprimés par une infection avec *Salmonella typhimurium* (Figure 18).

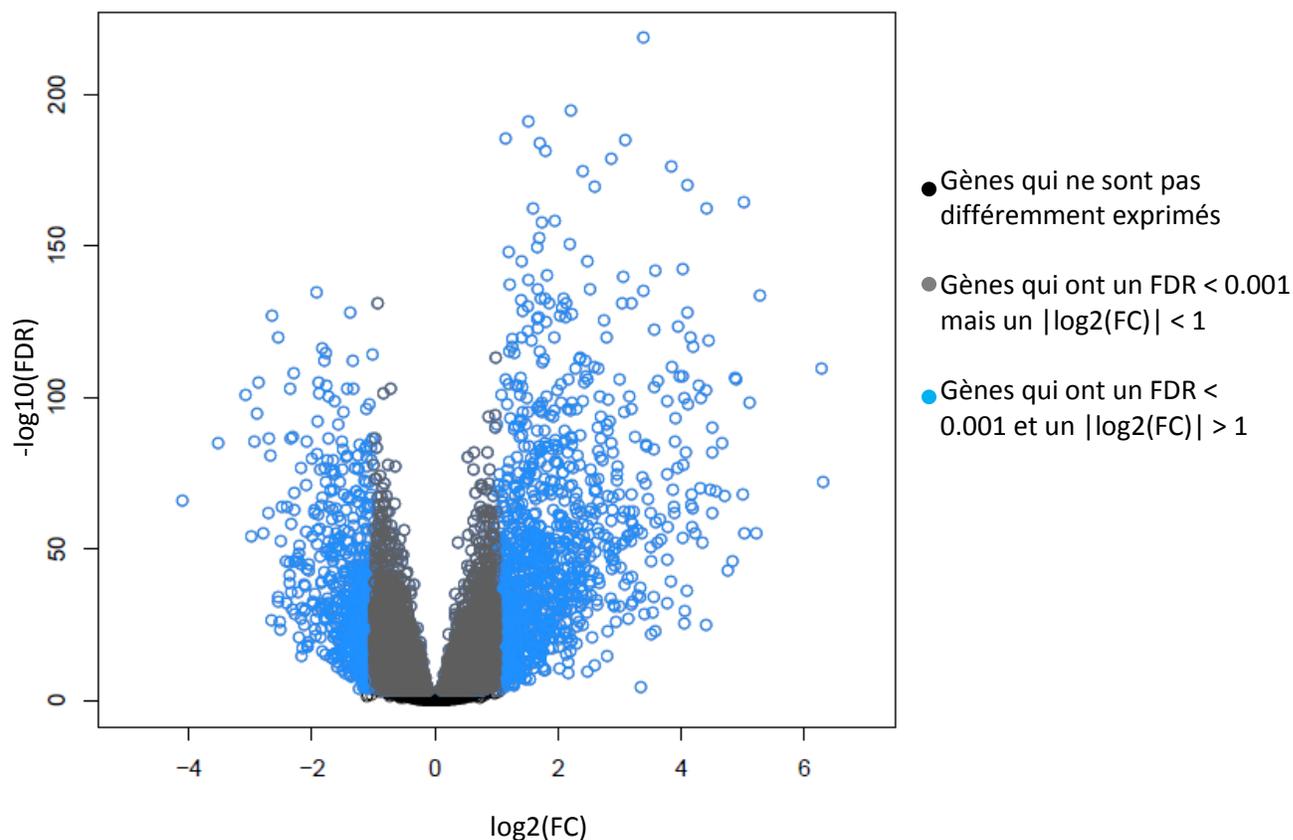


Figure 18: Graphique de type Volcano mettant en évidence les gènes différemment exprimés ($\text{FDR} < 0.001$ et valeur absolue de $\log_2(\text{FC}) > 1$) entre les conditions *Salmonella typhimurium* et *Listeria monocytogenes*.

Parmi les différences observées dans les niveaux d'expression génétique lors d'une infection à l'une ou l'autre des bactéries, nous avons identifié plusieurs gènes de la voie de signalisation de la production d'interféron-gamma qui sont plus exprimés chez les macrophages infectés avec *Salmonella typhimurium* que chez les macrophages infectés avec *Listeria monocytogenes* ($\text{FDR} < 1\%$) (Figure 19). On y retrouve entre autres les gènes IL12A, JAK2, STAT3, STAT4, JUN et IFNG.

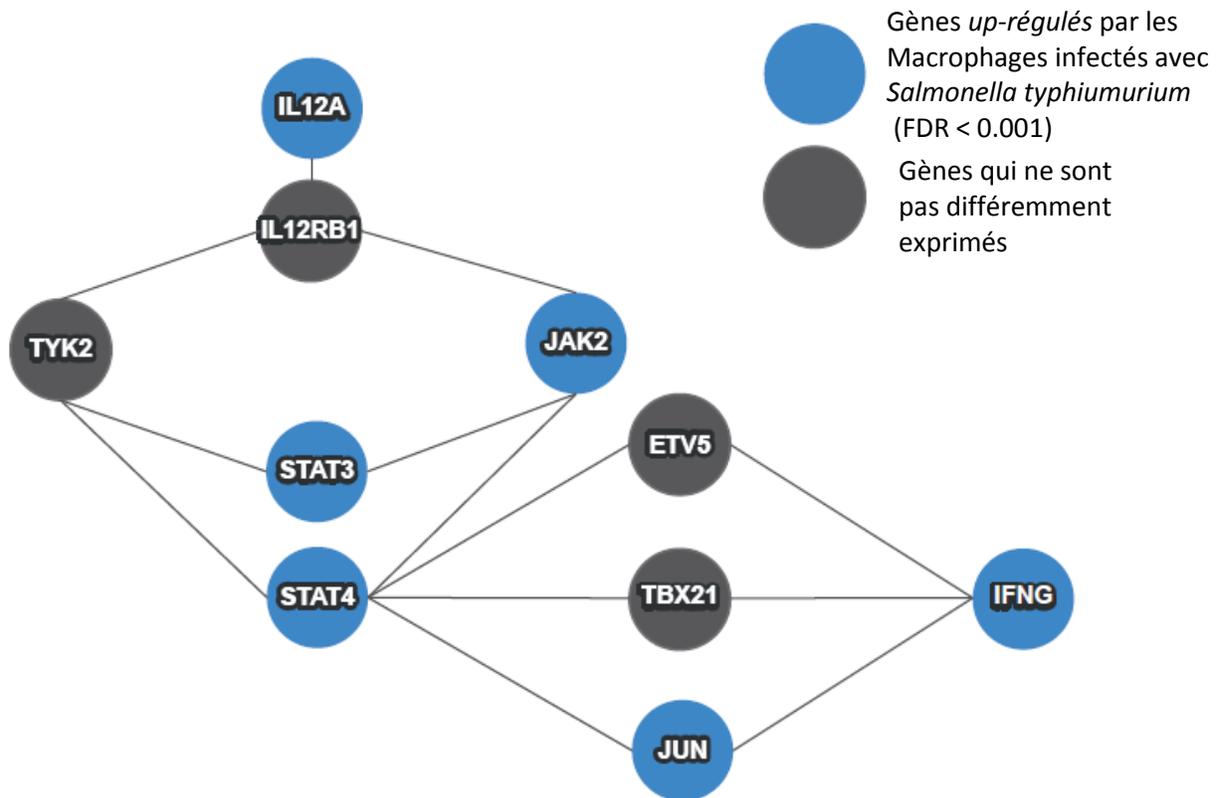


Figure 19: Voie de signalisation de la production d'interféron-gamma. Les gènes en bleu sont significativement plus exprimés chez les macrophages infectés avec *Salmonella typhimurium* que chez les macrophages infectés avec *Listeria monocytogenes* (FDR < 0.001).

4.2 Caractérisation interpopulationnelle de la réponse immunitaire des macrophages aux infections bactériennes

4.2.1 Caractérisation interpopulationnelle de l'efficacité du contrôle des bactéries par les macrophages

Puisque nous supposons qu'il existe des variations interpopulationnelles dans la réponse immunitaire aux infections bactériennes, il est logique de penser que ces différences devraient se refléter dans un caractère fonctionnel. Nous avons mesuré l'efficacité du contrôle des bactéries par les macrophages à 2, 4 et 24 heures post-infection. Pour ce faire, nous avons lysé les macrophages infectés pour récolter les bactéries. Ces bactéries ont été diluées, puis étalées sur des géloses afin de compter le nombre d'unités formant des colonies (CFU). Nous avons aussi dilué et étalé sur des géloses le bouillon d'inoculation, c'est-à-dire la solution de milieu de culture contenant les bactéries qui a été utilisée pour inoculer les macrophages, afin de compter le nombre CFU. Les valeurs de CFU post-infection (CIP) ont été divisées par les valeurs de CFU au moment de l'inoculation (CI) pour permettre de corriger pour la valeur de MOI qui varie légèrement entre chaque infection. Ces ratios ont subi une normalisation quantile et sa distribution nous a permis d'observer des différences significatives entre les Européens et Africains dans leur capacité à contrôler les bactéries à l'aide d'un test de Student (Figures 20-21). Les Africains parviennent à contrôler les bactéries plus rapidement que les Européens à 4 et 24 heures post-infection, indépendamment de la bactérie. Bien que cette tendance soit aussi observée à 2 heures post-infection, la différence n'est pas significative.

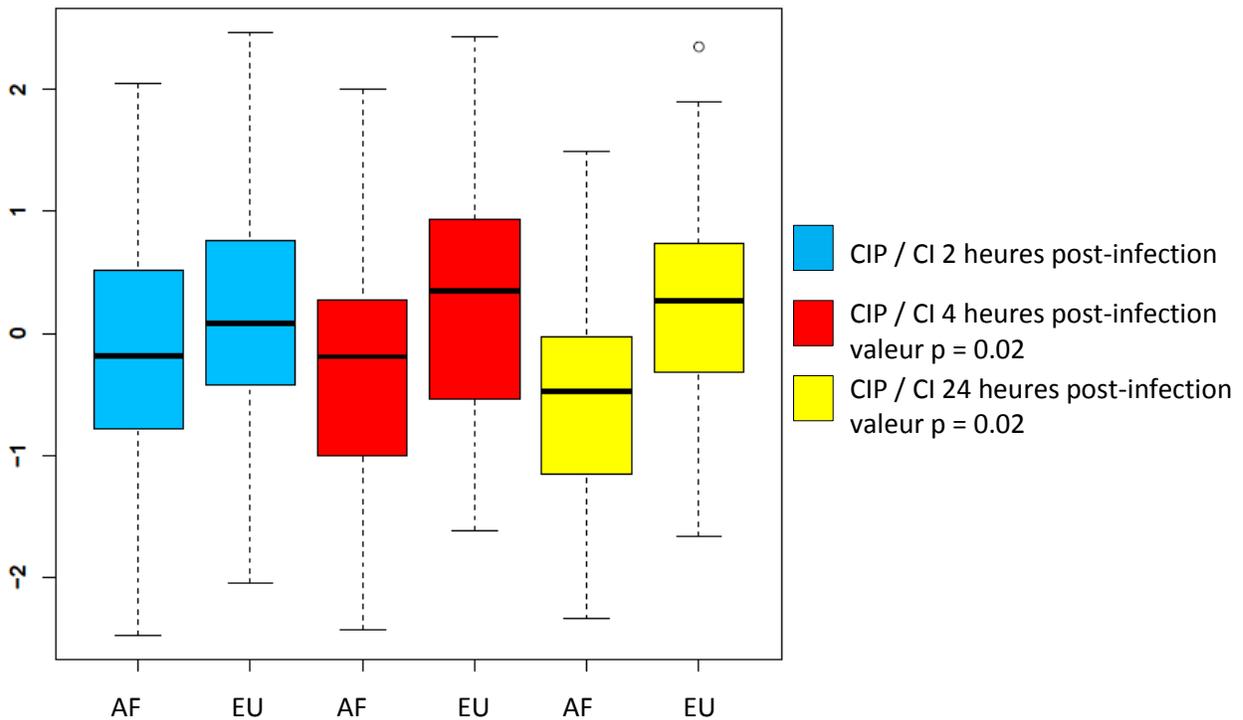


Figure 20: Boîte à moustaches des ratios CIP / CI pour les infections avec *Listeria monocytogenes*. (AF : Africains, EU : Européens). Les valeurs p ont été obtenues à l'aide d'un test de Student Africains vs Européens.

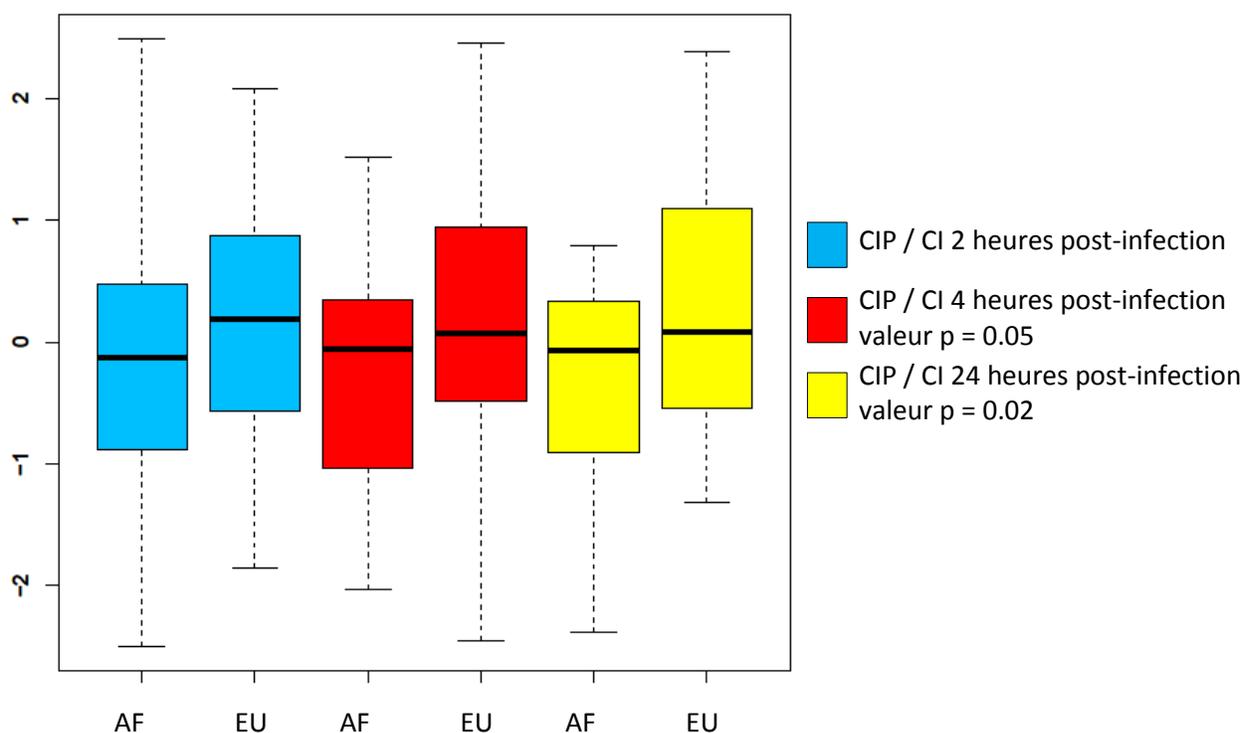


Figure 21: Boîte à moustaches des ratios CIP / CI pour les infections avec *Salmonella typhimuirum*. (AF : Africains, EU : Européens). Les valeurs p ont été obtenues à l'aide d'un test de Student Africains vs Européens.

4.2.2 Caractérisation interpopulationnelle des différences dans les niveaux de cytokines sécrétées

Nous avons mesuré en multiplexe les niveaux de sécrétion de 27 cytokines dans les surnageants des macrophages infectés et non infectés. La cytokine CCL4 a dû être éliminée des analyses parce que sa concentration dépassait le seuil maximal de détection. Les concentrations des 26 autres cytokines ont subi une normalisation quantile, puis elles ont été intégrées dans un modèle de régression linéaire avec les paramètres suivants : Niveau de sécrétion = Ethnicité + Condition + *Batch effect*. Ce modèle nous a permis d'identifier des cytokines qui sont différemment sécrétées entre les deux populations. Afin de corriger pour les tests multiples, nous avons recalculé les valeurs p en faisant 1000 permutations aléatoires des niveaux de sécrétion des cytokines dans le modèle de régression linéaire (Tableau II). Il est intéressant de constater que 3 des 4 cytokines significativement différemment sécrétées entre les populations dans les 2 conditions infectées (*Salmonella typhimuirum* et *Listeria*

monocytogenes) sont des cytokines anti-inflammatoires : IL-1RN, IL-10, IL-13. IL-1RN et IL-10 sont aussi significativement différemment sécrétés entre les populations dans la condition non-infectée.

	<i>Listeria monocytogenes</i>	<i>Salmonella typhimurium</i>	Non-infectée
IL-1B	0.929	0.977	0.591
IL-1RN	0.016	0.065	0.002
IL-2	0.899	0.27	0.104
IL-4	0.533	0.562	0.076
IL-5	0.171	0.112	0.003
IL-6	0.808	0.572	0.299
IL-7	0.002	0.164	0.008
IL-8	0.121	0.217	0.634
IL-9	0.206	0.008	0
IL-10	0.021	0.042	0
IL-12	0.527	0.12	0.013
IL-13	0.003	0.011	0.494
IL-15	0.599	0.806	0.187
IL-17A	0.451	0.046	0.002
CCL11	0.671	0.377	0.008
FGF2	0.519	0.387	0.502
CSF3	0.112	0.336	0.165
CSF2	0.96	0.217	0.002
IFNG	0.711	0.032	0.087
CXCL10	0.399	0.174	0.198
CCL2	0.027	0.183	0.04
CCL3	0.359	0.936	0.812
PDGFB	0.012	0.032	0.005
CCL5	0.032	0.371	0.006
TNFA	0.359	0.538	0.221
VEGFA	0.934	0.703	0.323

Tableau II: Tableau des valeurs p corrigées pour les tests multiples de la probabilité de commettre une erreur en rejetant l'hypothèse nulle que l'ethnicité n'est pas associée au niveau de sécrétion des cytokines dans les trois conditions.

Une étude plus approfondie des niveaux de cytokines anti-inflammatoires sécrétés ont permis de constater que IL-1RN, IL-10 et IL-13 sont plus fortement sécrétés chez les Européens de façon constante après une infection à *Salmonella typhimurium* ou *Listeria monocytogenes* (Figure 22). Ces

cytokines jouent un rôle de balance en régulant la sécrétion des cytokines pro-inflammatoires, créant ainsi une sorte d'équilibre dynamique pro-inflammatoire/anti-inflammatoire.⁸² Ceci suggère que les Européens présentent une réponse anti-inflammatoire plus forte que les Africains. La cytokine PDGFB est aussi différemment exprimée entre les populations dans les trois conditions : elle est plus exprimée chez les Européens dans les conditions non-infectée et *Salmonella typhimurium*, et elle est plus exprimé chez les Africains dans la condition *Listeria monocytogenes*.

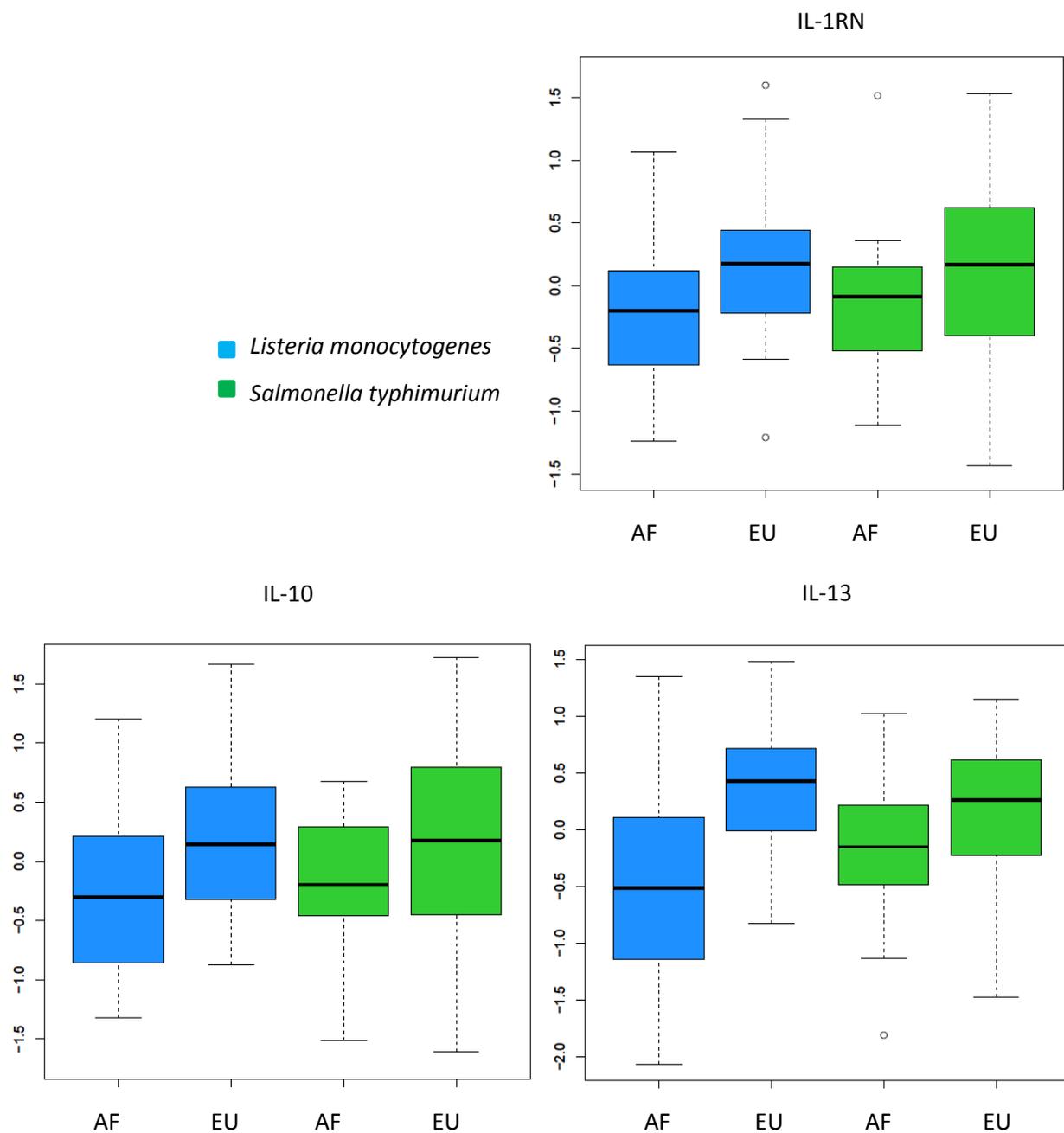


Figure 22: Niveaux normalisés de façon quantile des cytokines IL-1RN, IL-10 et IL-13 sécrétées dans le surnageant des macrophages chez les Africains et les Européens dans les conditions infectées (valeurs p des comparaisons interpopulationnelles < 0.05) (AF : Africains, EU : Européens)

4.2.3 Caractérisation interpopulationnelle des niveaux d'expression génétique

Pour comparer les résultats de séquençage du transcriptome entre les Africains et les Européens, nous avons utilisé le programme DESeq2 en intégrant les paramètres suivants dans le modèle : Niveau d'expression d'un gène = Condition + Ethnicité + *Batch effect* + CI. Le paramètre CI correspond au nombre d'unité formant des colonies (CFU) dans la solution de milieu de culture contenant les bactéries qui a été utilisées pour inoculer les macrophages. Nous avons calculé le FC et le FDR de chaque gène entre les Africains et les Européens dans chaque condition. Nous avons considéré les gènes comme étant différemment exprimés lorsqu'ils avaient un FDR inférieur à 5 % et une valeur absolue de $\log_2(\text{FC})$ supérieure à 0.5. Nous avons identifié 381 gènes différemment exprimés entre les populations dans la condition *Listeria monocytogenes*, 488 dans la condition *Salmonella typhimurium* et 399 dans la condition non infectée (Figure 23). Nous avons fait des analyses GO avec le programme GeneTrail pour mieux caractériser les gènes différemment exprimés entre les populations dans chacune des trois conditions (Tableau III).⁸¹ Parmi les 5 termes GO les plus significatifs entre les populations pour chacune des conditions, nous avons remarqué que les gènes différemment exprimés sont enrichis pour accomplir des fonctions liées à la communication extracellulaire puisque le terme « *extracellular region* » se retrouve dans les trois conditions. Ce terme GO regroupe, entre autres, les gènes CXCL1 et IL8, qui codent pour des chimiokines ayant un effet sur les neutrophiles polymorphonucléaires, et le gène CD40, un récepteur jouant un rôle dans l'activation des macrophages par les lymphocytes T.^{83,84}

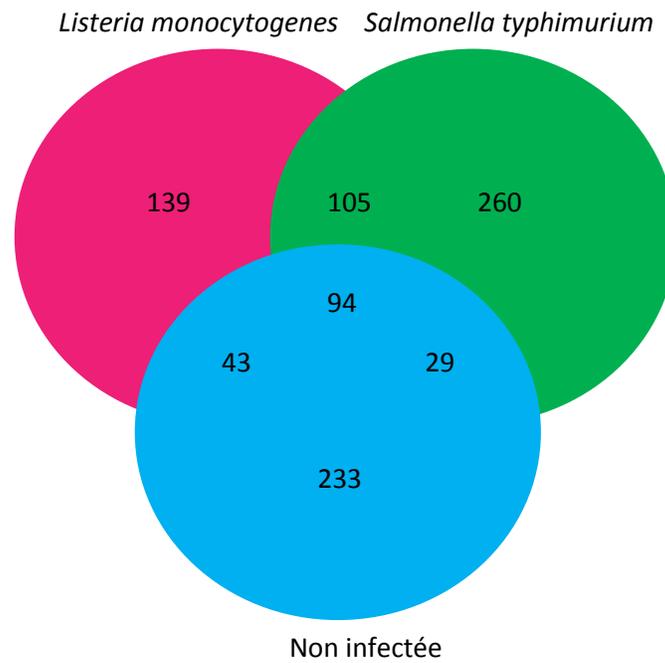


Figure 23: Diagramme de Venn des gènes différemment exprimés entre les Africains et les Européens (FDR < 0.05 et valeur absolue de $\log_2(\text{FC}) > 0.5$) dans chacune des conditions.

Conditions	Termes GO
<i>Listeria monocytogenes</i>	<i>extracellular region</i>
	<i>multicellular organismal process</i>
	<i>receptor activity</i>
	<i>plasma membrane</i>
	<i>membrane</i>
<i>Salmonella typhimurium</i>	<i>extracellular region</i>
	<i>extracellular region part</i>
	<i>extracellular space</i>
	<i>chemokine activity</i>
	<i>multicellular organismal process</i>
Non infectée	<i>plasma membrane</i>
	<i>receptor activity</i>
	<i>transmembrane receptor activity</i>
	<i>intrinsic to membrane</i>
	<i>extracellular region</i>

Tableau III: Tableau des 5 groupes ontologiques les plus significatifs (FDR < 0.05) pour les gènes différemment exprimés entre les Africains et les Européens dans chacune des conditions.

Parmi les gènes différemment exprimés entre les populations, certains ont retenu notre attention. C'est le cas du gène IL1R1, qui est plus exprimé chez les Africains que chez les Européens dans les trois conditions (FDR < 0.05) et qui code pour un récepteur de l'interleukine 1 (Figure 24). À la Figure 24, l'axe des x représente la position du gène sur le chromosome, en nombre de bases, et les exons sont indiqués par les rectangles noirs. L'axe des y représente le nombre moyen de *reads* par base divisé par le nombre total de *reads* alignés par échantillon.

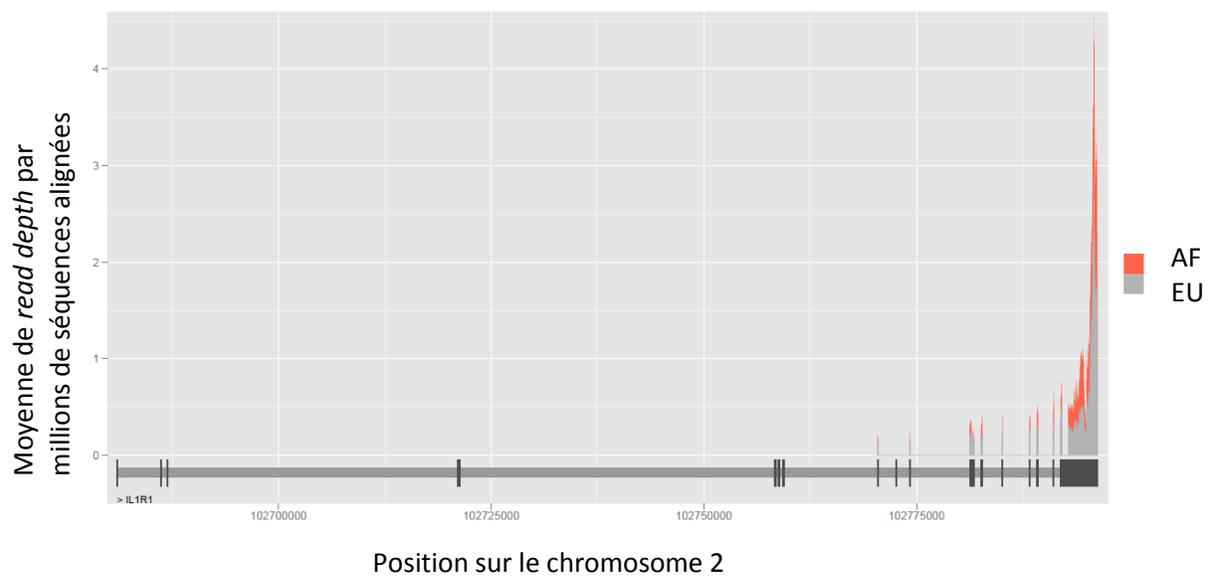


Figure 24: Le gène IL1R1 est plus exprimé chez les Africains que chez les Européens dans les 3 conditions (FDR < 0.05, log2(FC) > 0.5). C'est la condition non infectée qui est présentée ici.

Le gène mir155HG, qui code pour le micro ARN 155, est plus exprimé chez les Africains que chez les Européens dans les macrophages infectés avec *Salmonella typhimurium* et *Listeria monocytogenes* (FDR < 0.05) (Figure 25).

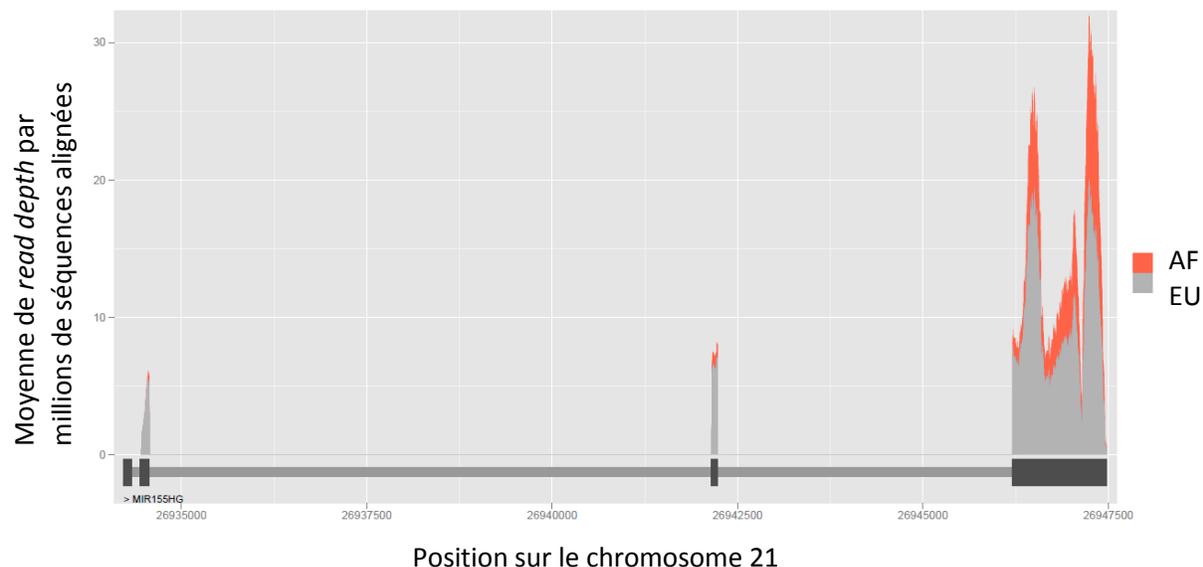


Figure 25: Le gène mir155HG est plus exprimé chez les Africains que chez les Européens dans les conditions *Listeria monocytogenes* et *Salmonella typhimurium* (FDR < 0.05, $\log_2(FC) > 0.5$). C'est la condition *Salmonella typhimurium* qui est présentée ici.

4.2.4 Identification d'eQTLs

Nous avons vérifié si les gènes différemment exprimés entre les populations sont aussi associés à des polymorphismes génétiques situés en cis (cis-eQTL) afin de valider le fait que les différences d'expression observées sont régulées génétiquement. Pour ce faire, pour chaque gène dont nous avons mesuré l'expression, nous avons sélectionné les SNPs se situant dans une fenêtre de 200 kb centrée sur le site d'initiation de la transcription (TSS) et nous avons éliminé ceux qui ont une fréquence allélique mineure inférieure à 10 % parmi tous les individus. Nous avons utilisé un modèle de régression linéaire incluant les paramètres suivants : Niveau d'expression = Génotype + Ethnicité. Ce modèle nous a permis d'identifier des variants génétiques associés à l'expression des gènes (i.e., eQTL). Pour corriger pour les tests multiples, nous avons calculé le FDR en faisant 10 permutations aléatoires des niveaux d'expression du modèle de régression linéaire. Seuls les eQTLs avec un FDR

inférieur à 5 % ont été retenus. Nous avons observé le fait que les gènes différemment exprimés entre les populations, dans chacune des trois conditions, sont significativement enrichis pour des eQTLs avec des valeurs $p < 0.01$ (test du khi-carré) par rapport aux gènes qui ne sont pas différemment exprimés (Figure 26). Nous avons obtenu 28 gènes différemment exprimés qui ont aussi un eQTL dans la condition non infectée, 31 dans la condition *Listeria monocytogenes* et 26 dans la condition *Salmonella typhimurium*.

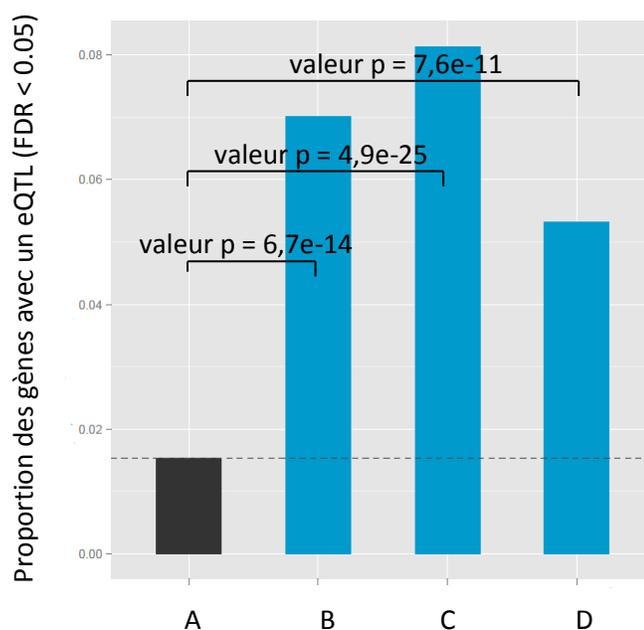


Figure 26: Proportion des gènes qui ont un eQTL (FDR < 0.05) parmi A) les gènes qui ne sont pas différemment exprimés entre les populations; B) les gènes qui sont différemment exprimés entre les populations dans la condition non infectée; C) les gènes qui sont différemment exprimés entre les populations dans la condition *Listeria monocytogenes*; D) les gènes qui sont différemment exprimés entre les populations dans la condition *Salmonella typhimurium*. Les gènes différemment exprimés entre les populations, dans chacune des trois conditions, sont significativement enrichis pour des eQTLs par rapport aux gènes qui ne sont pas différemment exprimés. Un test du khi-carré a été utilisé pour calculer les valeurs p.

Parmi les gènes différemment exprimés entre les populations qui ont aussi un eQTL, le gène *RETN*, qui code pour l'hormone résistine, a retenu notre attention. Ce gène est plus exprimé chez les Africains que chez les Européens dans les trois conditions ($FDR < 0.05$, $\log_2(FC) > 0.5$) (Figure 27).

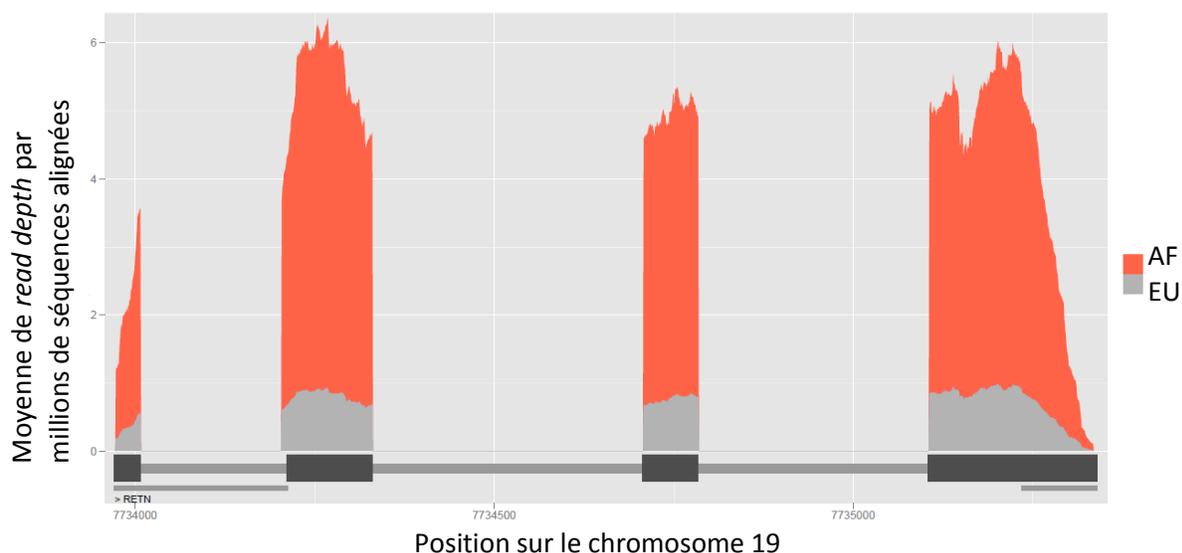


Figure 27: Le gène *RETN* est plus exprimé chez les Africains que chez les Européens dans les 3 conditions ($FDR < 0.05$, $\log_2(FC) > 0.5$), et a un eQTL ($FDR < 0.05$). C'est la condition *Salmonella typhimurium*, qui est présentée ici.

4.3 Signatures de sélection naturelle dans les gènes différemment exprimés entre les populations

Afin d'évaluer la contribution de la sélection naturelle dans l'explication de la variabilité de la réponse immunitaire aux infections bactériennes observée entre les Africains et les Européens, nous avons utilisé deux paramètres indicateurs de sélection naturelle : le coefficient F_{ST} et les valeurs iHS.

4.3.1 Différenciation des populations

Pour le calcul des coefficients F_{ST} , nous avons utilisé les génotypes Hapmap des populations CEU (descendants des Européens du Nord et de l'Ouest) et YRI (Yorubas du Nigéria en Afrique) parce que

ces individus ne présentent pas d'*admixture*.⁸⁵ Cela nous a permis d'améliorer notre pouvoir de détection. Nous avons retiré les positions avec plus de 10 % de données manquantes. Nous avons calculé les coefficients F_{ST} afin d'obtenir une mesure du niveau de différenciation des populations européennes et africaines pour chaque position des SNPs HapMap. Le coefficient F_{ST} varie entre 0 et 1. Une valeur de 0 signifie que les populations sont identiques pour une position donnée tandis qu'une valeur de 1 indique que les populations sont complètement divergentes pour cette position.

Afin de vérifier si nos gènes différemment exprimés entre les populations sont enrichis pour des valeurs élevées de F_{ST} , nous avons sélectionné le SNP avec la valeur de F_{ST} maximum dans une fenêtre de 100 kb centrée sur le TSS de chaque gène dont nous avons mesuré l'expression. Nous avons utilisé le 95^e percentile de cette distribution, qui correspond à un coefficient F_{ST} de 0.55, comme valeur seuil indiquant un niveau de différenciation des populations élevé. Nous avons calculé, dans les trois conditions, la proportion de gènes différemment exprimés entre les populations qui ont un SNP avec un coefficient F_{ST} supérieur à ce seuil. Nous avons calculé cette même proportion pour les gènes qui ne sont pas différemment exprimés à partir de 1000 rééchantillonnages. Enfin, nous avons compté les occurrences de proportions supérieures à nos observations dans les gènes qui sont différemment exprimés afin d'obtenir une valeur p. Nous avons observé le fait que les gènes différemment exprimés entre les populations dans la condition *Salmonella typhimurium*, les gènes communément différemment exprimés entre les populations dans les conditions *Salmonella typhimurium* et *Listeria monocytogenes* ainsi que les gènes communément différemment exprimés entre les populations dans les conditions *Salmonella typhimurium*, *Listeria monocytogenes* et non infectée sont significativement enrichis pour des valeurs élevées de F_{ST} (Figure 28). Bien que l'enrichissement pour des valeurs de F_{ST} élevées ne soit pas significatif dans la condition *Listeria monocytogenes*, la valeur p de 0.06 indique la même tendance.

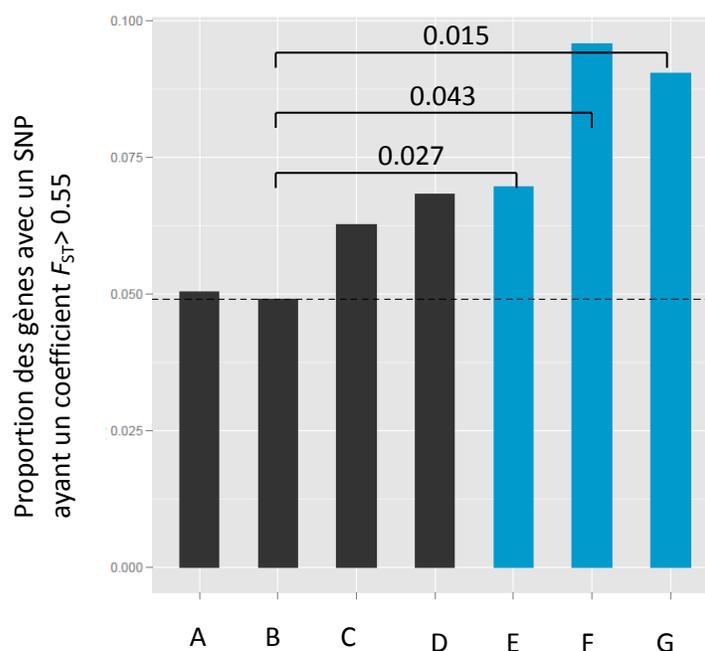


Figure 28: Proportion des gènes qui ont au moins un SNP avec un coefficient $F_{ST} > 0.55$ parmi A) tous les gènes dont nous avons mesuré l'expression; B) les gènes qui ne sont pas différemment exprimés entre les populations; C) les gènes qui sont différemment exprimés dans la condition non infectée; D) les gènes qui sont différemment exprimés dans la condition *Listeria monocytogenes*; E) les gènes qui sont différemment exprimés dans la condition *Salmonella typhimurium*; F) les gènes communément différemment exprimés dans les trois conditions; G) les gènes communément différemment exprimés dans les conditions *Listeria monocytogenes* et *Salmonella typhimurium*.

4.3.2 Identification de signatures de sélection positive récentes

Pour le calcul des valeurs d'iHS, nous avons utilisé les génotypes obtenus par nos puces de génotypage pour identifier, au sein de chaque population, des événements récents de sélection positive. Les valeurs d'iHS sont calculées en observant la longueur du déséquilibre de liaison autour d'un SNP en supposant que les SNPs positivement sélectionnés siègeront sur un haplotype anormalement long. Nous avons *phasé* les génotypes pour obtenir des haplotypes avec le programme SHAPIT, nous avons identifié l'allèle ancestral ou les allèles dérivés avec le pipeline EPO, puis nous avons calculé les iHS avec le module reh. ⁸⁶⁻⁸⁹ Une valeur absolue d'iHS supérieure à 2 est communément utilisée pour identifier des signatures de sélection naturelle.

Afin de vérifier si les gènes différemment exprimés entre les populations sont enrichis pour des valeurs élevées d'iHS, nous avons sélectionné tous les SNPs dans une fenêtre de 100 kb centrée sur le TSS de chaque gène dont nous avons mesuré l'expression. Pour chaque gène, nous avons calculé la proportion de SNPs qui ont une valeur absolue d'iHS supérieure à 2. À partir de cette distribution de proportions, nous avons identifié le 95^e percentile, qui correspond à une proportion de 0.22 dans les deux populations, c'est-à-dire les gènes qui présentent une forte proportion de SNPs avec des valeurs d'iHS élevées. Nous avons calculé, dans les trois conditions, la proportion de gènes différemment exprimés entre les populations qui présentent une forte proportion de SNPs avec des valeurs d'iHS élevées. Nous avons calculé, à partir de 1000 rééchantillonnages, cette même proportion chez les gènes qui ne sont pas différemment exprimés. Enfin, nous avons compté les occurrences de proportions supérieures à nos observations chez les gènes qui sont différemment exprimés pour obtenir une valeur p. Les gènes différemment exprimés ne sont pas significativement enrichis pour des valeurs d'iHS élevées chez les Africains dans aucun des sous-groupes testés. Les gènes communément différemment exprimés dans les trois conditions sont enrichis pour des valeurs d'iHS élevées chez les Européens (Figure 29).

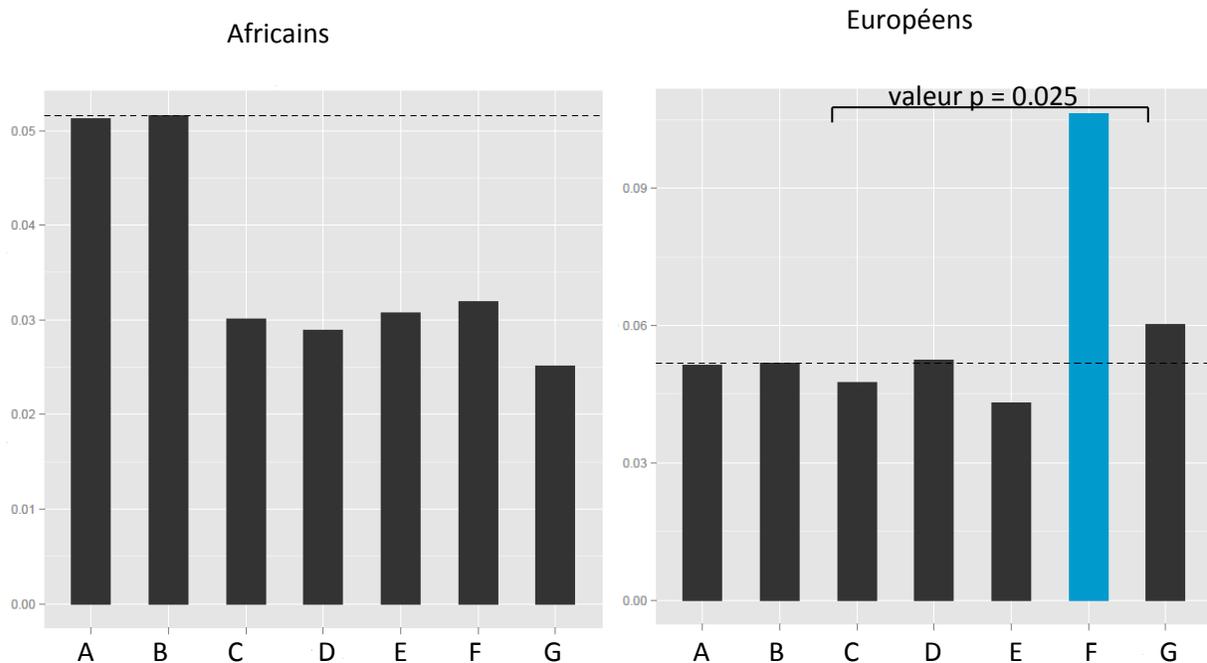


Figure 29: Proportion des gènes qui ont une forte proportion de SNPs avec des valeurs d'iHS élevés (> 0.22) chez les Africains et chez les Européens parmi A) tous les gènes dont nous avons mesuré l'expression; B) les gènes qui ne sont pas différemment exprimés entre les populations; C) les gènes qui sont différemment exprimés dans la condition non infectée; D) les gènes qui sont différemment exprimés dans la condition *Listeria monocytogenes*; E) les gènes qui sont différemment exprimés dans la condition *Salmonella typhimurium*; F) les gènes communément différemment exprimés dans les trois conditions; G) les gènes communément différemment exprimés dans les conditions *Listeria monocytogenes* et *Salmonella typhimurium*.

5. Discussion

Dans ce projet, nous avons étudié la variabilité de la réponse immunitaire des Africains et des Européens à des infections bactériennes *in vitro* puis nous avons évalué la contribution de la sélection naturelle dans l'explication de ce phénomène. Les résultats de notre étude seront discutés dans cette section.

5.1 *Listeria monocytogenes* et *Salmonella typhimurium* ont induit une réponse immunitaire innée spécifique

La caractérisation de la réponse immunitaire engendrée par une infection avec *Salmonella typhimurium* ou *Listeria monocytogenes* confirme que l'utilisation des deux bactéries a permis d'obtenir un portrait plus large des gènes susceptibles d'être activés lors d'une infection bactérienne, comme l'indique la Figure 13. Chaque bactérie a développé des mécanismes spécifiques pour survivre aux défenses antimicrobiennes de la cellule hôte, et cela s'est notamment traduit par la *down-régulation* de gènes aux fonctions différentes par *Salmonella typhimurium* et par *Listeria monocytogenes*, comme nous pouvons l'observer à la Figure 17. Ainsi, *Salmonella typhimurium* et *Listeria monocytogenes*, grâce à leur structure moléculaire distincte et leur capacité à interagir avec la cellule hôte d'une façon qui leur est propre, ont induit une réponse immunitaire innée spécifique.

5.2 Les Européens éliminent les bactéries intracellulaires plus lentement que les Africains

Nous avons observé que les macrophages des Africains parviennent à mieux contrôler les bactéries que les macrophages des Européens. Cette tendance est significativement observable après 4 et 24 heures d'infection, mais n'est pas significative après seulement 2 heures d'infection. Ceci indique que c'est bien le mécanisme de contrôle qui est plus efficace chez les Africains, et non pas l'invasion des bactéries dans les macrophages qui est plus limitée chez les Africains. Cette mesure, qui est directement liée à la réponse immunitaire de l'hôte, est une bonne indication qu'il existe une différence interpopulationnelle dans la réponse immunitaire innée aux infections bactériennes.

5.3 Les Européens sécrètent une plus grande quantité de cytokines anti-inflammatoires que les Africains

La quantification des cytokines sécrétées dans le surnageant des macrophages infectés et non infectés a mis en évidence plusieurs différences entre les Africains et les Européens. L'une des plus remarquables est le niveau d'IL1RN, d'IL10 et d'IL13, qui est significativement plus élevé chez les Européens de façon constante après une infection à *Salmonella typhimurum* ou à *Listeria monocytogenes* (Figure 22). Ces résultats indiquent que les Européens présentent une réponse anti-inflammatoire plus forte que les Africains. Le système immunitaire est régulé par un réseau complexe de molécules interagissant constamment ensemble dans le but de moduler une réponse immunitaire appropriée et efficace. Les cytokines anti-inflammatoires jouent un rôle important au sein de ce réseau en limitant le potentiel dommageable d'une réponse inflammatoire excessive ou soutenue. Un déséquilibre du système immunitaire vers un excès de réponse anti-inflammatoire peut mener à une inhibition du système immunitaire causant un risque d'infection systémique.⁹⁰ Par ailleurs, un déséquilibre du système immunitaire vers un excès de réponse pro-inflammatoire peut causer un environnement favorable au développement de maladies inflammatoires et auto-immunes.⁹¹

IL-1RN est un inhibiteur de la famille de l'IL-1 puisqu'il bloque l'action de l'IL-1 α et de l'IL-1 β en se liant au récepteur IL-1R1.⁹² La liaison de l'IL-1 α et de l'IL-1 β au récepteur IL1R1 active une cascade de signalisation qui mène à l'activation du facteur de transcription *nuclear factor-kappa B* (NF- κ B). Cette activation stimule la synthèse de cyclooxygénase de type 2, de phospholipase A de type 2 et d'oxyde nitrique synthase, ce qui induit la production de prostaglandine-E2 et d'oxyde nitrique. Il en résulte généralement de la fièvre, une vasodilatation et une hypotension, qui sont des caractéristiques classiques de la réponse inflammatoire.⁹³ La liaison non-productive de l'antagoniste IL1RN au récepteur IL1R1, puisqu'elle n'active pas la transcription de NF- κ B bloque l'action de l'IL-1 α et de l'IL-1 β . Il existe au moins un polymorphisme génétique associé à la régulation de la synthèse de l'IL-1RN et il a été démontré qu'un excès de sécrétion d'IL-1RN, par rapport à la sécrétion d'IL-1 α et d'IL-1 β , est associé à une augmentation de la susceptibilité à la tuberculose et aux septicémies.⁹⁴⁻⁹⁶

L'IL-10 joue un rôle important dans l'inhibition de la réponse pro-inflammatoire afin de limiter son effet dommageable sur l'hôte. La sécrétion de l'IL-10 se fait à la suite de l'activation du facteur de transcription NF- κ B.⁹⁷ L'induction de son effet anti-inflammatoire est médiée par sa liaison sur le récepteur IL-10R et par sa signalisation par l'entremise du *signal transducer and activator of*

transcription 3 (STAT3).⁹⁸ La sécrétion d'IL-10 par les macrophages entraîne une inhibition de la réponse par les lymphocytes T_H1 et T_H2.⁹⁷ L'action de l'IL-10 sur le macrophage lui-même est liée à une inhibition des fonctions effectrices innées du macrophage ainsi qu'à une diminution de l'expression des protéines du complexe majeur d'histocompatibilité de classe II.⁹⁹ D'ailleurs, des études ont démontré que des niveaux élevés de sécrétion d'IL-10 sont associés à une plus grande mortalité causée par la méningococcie.¹⁰⁰ Plus récemment, un groupe a démontré que des niveaux élevés de la cytokine anti-inflammatoire IL-10 diminuent la réponse bactéricide du macrophage, ce qui concorde parfaitement avec nos observations.¹⁰¹

L'IL-13 est une cytokine anti-inflammatoire sécrétée par plusieurs types cellulaires et qui est bien caractérisée pour sa modulation de l'inflammation causée par les allergies. L'induction de son effet anti-inflammatoire est médiée par sa liaison sur le récepteur IL-4R α et par sa signalisation par l'entremise du *signal transducer and activator of transcription 6* (STAT6).¹⁰² La sécrétion d'IL-13 a un effet modulateur sur les monocytes et les lymphocytes B.¹⁰³ L'action de l'IL-13 sur le macrophage lui-même induit une inhibition du facteur de transcription NF- κ B qui mène à une diminution des protéines du complexe majeur d'histocompatibilité de classe II et à une *down-régulation* de la sécrétion de TNF, d'IL-1 et d'IL-8.^{104,105} L'IL-13 semble donc jouer un rôle important dans la modulation de la réponse pro-inflammatoire du macrophage.

Ces résultats indiquent que les Européens présentent une réponse immunitaire plus fortement régulée par l'action des cytokines anti-inflammatoire par rapport aux Africains. Ceci est une piste intéressante pour expliquer le mécanisme plus efficace des Africains pour éliminer les bactéries intracellulaires et, plus globalement, elle pourrait expliquer, en partie, les différences dans la susceptibilité aux maladies infectieuses entre les populations européennes et africaines.

5.4 Les Africains et les Européens présentent des différences dans le niveau d'expression de certains gènes

L'étude du transcriptome nous a permis d'identifier, à travers les trois conditions d'analyse, plusieurs gènes différemment exprimés entre les populations. Ces gènes sont enrichis pour des fonctions liées à la membrane cellulaire et à la communication extracellulaire. Néanmoins, il est important de noter que les nombres de différences que nous avons observés entre les populations sont sous-estimés puisqu'il ne considère pas les isoformes. En effet, comme l'a démontré Wang et al., nous savons que

92 à 94% des gènes produisent des isoformes et que ceux-ci dépendent du type de tissu ou du contexte biologique.¹⁰⁶ De plus, comme l'a décrit Lappalainen et al., la proportion de gènes dans les lignées lymphoblastoïdes du *Consortium 1000 genomes Project* qui présentent une utilisation différente des transcrits d'ARN entre les populations européennes et la population africaine varie entre 75 et 85%.⁶⁷ L'étude des gènes qui présentent des différences dans l'utilisation des transcrits d'ARN entre les populations est donc un point important qui nécessitera de s'y attarder davantage dans le futur.

Parmi les gènes différemment exprimés entre les populations, le récepteur IL1R1 à la Figure 24 a attiré notre attention puisqu'il fait partie de la famille de l'IL-1 dont il a été question un peu plus tôt. Ce gène est différemment exprimé entre les populations dans les trois conditions mais répond aussi à l'infection bactérienne puisque les $\log_2(\text{FC})$ entre les conditions *Salmonella typhimurium* et non-infecté, et *Listeria monocytogenes* et non-infecté, est plus grand que 1 ($\log_2(\text{FC}) > 1$, $\text{FDR} < 0.001$). Ce récepteur peut être lié par l'IL-1 α et par l'IL-1 β pour activer le facteur de transcription NF- κ B, ou par l'IL-1RN pour bloquer l'action de l'IL-1 α et de l'IL-1 β . Nous avons observé précédemment que l'IL-1RN est davantage sécrété dans le surnageant des macrophages européens tandis qu'ici le gène IL1R1 est davantage exprimé chez les Africains. Cette observation confirme l'hypothèse que, pour la voie de signalisation de l'IL-1, la balance penche plutôt du côté de la réponse anti-inflammatoire chez les Européens tandis qu'elle penche plutôt du côté pro-inflammatoire chez les Africains.

Le gène hôte du miR-155, qui est davantage exprimé chez les Africains dans les conditions *Salmonella typhimurium* et *Listeria monocytogenes* (Figure 25), est impliqué dans plusieurs processus physiologiques, dont notamment la réponse immunitaire.^{107,108} Les micro-ARNs sont des modulateurs importants de la régulation post-transcriptionnelle de l'expression génétique. Le miR-155 est connu pour être hautement présent dans les macrophages et les cellules dendritiques à la suite de l'activation des récepteurs TLR.¹⁰⁹ Il a été démontré qu'un large éventail de médiateurs pro-inflammatoires pouvaient déclencher la transcription du gène hôte du miR-155 par l'entremise des facteurs de transcription NF- κ B ou AP-1.¹⁰⁷ De plus, il a été démontré que l'IL-10 réduit l'activité du promoteur du gène MIR155HG.¹¹⁰ Ceci explique, en partie, les différences d'expression du MIR155HG entre les Africains et les Européens puisque nous avons démontré précédemment que l'IL10 est plus fortement sécrétée dans le surnageant des macrophages Africains. Il semble donc qu'il y ait un lien étroit entre l'expression du miR-155, l'inflammation et la réponse immunitaire innée.

Nous avons validé le fait que les gènes différemment exprimés entre les populations sont enrichis pour des eQTLs dans les trois conditions, indiquant que les gènes différemment régulés le sont par l'entremise d'un mécanisme génétique. Le gène *RETN*, qui a un eQTL, est différemment exprimé entre les populations dans les trois conditions mais répond aussi à l'infection bactérienne puisque les $\log_2(\text{FC})$ entre les conditions *Salmonella typhimurium* et non-infecté, et *Listeria monocytogenes* et non-infecté, est plus grand que 1 ($\log_2(\text{FC}) > 1$, $\text{FDR} < 0.001$). Il a longtemps été associé à la résistance à l'insuline et à l'obésité (Figure 27).^{111,112} Plus récemment, une étude a observé que les patients avec des signes cliniques d'inflammation sévère présentent une concentration de résistine plus élevée que les patients en santé. De plus, une corrélation positive significative entre le niveau de résistine et le niveau d'autres marqueurs inflammatoires a été mise en évidence.¹¹³ La résistine est donc un marqueur de l'inflammation.

5.5 Les Africains présentent une réponse inflammatoire plus forte que les Européens

L'ensemble de ces résultats indique que les Africains présentent une réponse inflammatoire plus forte, autant sur le plan protéique qu'au niveau de l'expression génétique. Parmi les différences interpopulationnelles présentées précédemment, on peut remarquer que certaines le sont aussi dans la condition non-infectée, comme les gènes *IL1R1* et *RETN* et les cytokines *IL1RN* et *IL10*. Ceci indique donc que, non seulement la réponse pro-inflammatoire des Africains aux infections bactériennes est plus forte que celle des Européens, mais que le profil d'expression des macrophages à l'état basal est aussi plus fortement pro-inflammatoire chez les Africains. Ceci se traduit par un contrôle des bactéries intracellulaires plus efficace chez les Africains et, plus globalement, pourrait expliquer en partie les différences dans la susceptibilité aux maladies infectieuses entre les populations européennes et africaines. À première vue, cette information peut sembler en contradiction avec le Tableau I, qui indique que les Afro-Américains ont un taux de mortalité significativement plus élevé que les Américains d'origine européenne pour certaines maladies infectieuses. Néanmoins, une étude antérieure a démontré qu'un polymorphisme situé dans le gène du *TLR1* est responsable d'un mauvais trafic intracellulaire de ce récepteur vers la membrane cellulaire et ainsi d'une mauvaise signalisation via ce récepteur. Étonnement ce polymorphisme confère une résistance contre la lèpre.¹¹⁴ Plus récemment, Barreiro et al. ont démontré que ce même polymorphisme est associé avec une diminution de l'activation de *NF-κB* et est sous l'effet de la

sélection positive chez les Européens.¹¹⁵ Ces informations indiquent qu'une réponse inflammatoire atténuée peut conférer un avantage, suggérant qu'une réponse pro-inflammatoire excessive pourrait entraîner chez l'hôte des dommages qui pourraient expliquer les différences dans la susceptibilité aux maladies infectieuses entre les populations européennes et africaines.

Comme mentionné précédemment, un système immunitaire trop actif induit des troubles immunologiques tels que des maladies auto-immunes ou des allergies. D'un autre côté, un système immunitaire trop axé vers la réponse anti-inflammatoire n'est pas efficace pour éliminer les infections. De ce fait, dans un milieu fortement pathogénique, il semble être approprié d'avoir une réponse pro-inflammatoire forte. Le fardeau imposé par les maladies infectieuses est plus important en Afrique que partout ailleurs. Une des explications est attribuable à la présence de l'espèce humaine et de ses ancêtres en Afrique depuis des millions d'années. Cette présence a permis aux pathogènes locaux de s'adapter à l'espèce humaine sur une longue période de temps. Ceci peut expliquer pourquoi les descendants de populations africaines présentent une réponse pro-inflammatoire plus active que les descendants de populations européennes.²⁸

Des résultats de différentes études abondent dans le même sens que nos observations. Par exemple, Ness et al. ont étudié les variants alléliques dans les gènes de cytokines chez des femmes américaines d'origine européenne et africaine. Ils ont déterminé que les femmes américaines d'origine africaine étaient significativement plus susceptibles d'être porteuse d'un variant allélique connu pour up-réguler la production de cytokines pro-inflammatoires que les américaines d'origine européennes.¹¹⁶ Une étude sur le protéine réactive C, un marqueur de l'inflammation associé avec l'IL6, indiquent que les Afro-Américains présentent de niveaux sérologiques significativement plus élevé que les autres ethnicités de l'étude (hispaniques, chinois, japonais, caucasiens), même après avoir corrigé les données pour l'âge, le statut socioéconomique et l'indice de masse corporel.^{117,118} La caractérisation des variants pro-inflammatoires du TLR4, tel que revue par Pennington et al. indique que les Africains présente un profil pro-inflammatoire plus fort que les autres groupes ethniques. D'ailleurs 2 polymorphismes de ce récepteur présents à une fréquence plus élevé chez les Africains est associé avec une inflammation de la membrane chorionique, augmentant les risques d'accouchements prématurés.¹¹⁹ L'ensemble de ces observations, combinées avec notre étude, supportent l'idée que les descendants des populations africaines présentent un profil pro-inflammatoire plus marqué que les descendants des populations européennes.

5.6 Des gènes différemment exprimés entre les populations sont marqués par la sélection naturelle chez les Européens

Les tests de détection de la sélection naturelle ont indiqué que les gènes différemment exprimés entre les populations étaient enrichis pour des signatures de sélection naturelle. Dans un premier temps, nous avons démontré que les gènes différemment exprimés entre les populations étaient enrichis pour des SNPs avec un niveau de différenciation des populations élevé. Puis nous avons démontré que les gènes communément différemment exprimés entre les populations dans les trois conditions étaient enrichis pour une forte proportion de SNPs avec des valeurs d'iHS élevées seulement chez les Européens. Le test d'iHS permet de détecter des événements de sélection positive récents. Le fait que nous détectons des valeurs d'iHS élevées seulement dans la population européenne indique que des événements de sélection positive sont probablement survenus en dehors du continent africain après la migration des individus qui ont peuplé l'Europe. Après leur déplacement hors de l'Afrique, l'environnement différent a créé un changement de la pression sélective. Il semble que, dans ce nouvel environnement, une réponse pro-inflammatoire forte n'était plus adéquate. Ainsi, afin de limiter les effets potentiellement dommageables d'une réponse pro-inflammatoire trop forte, la sélection naturelle aurait favorisée des mécanismes anti-inflammatoires afin de moduler une réponse immunitaire plus appropriée au nouvel environnement.²⁸ Cette hypothèse est soutenue par la plus forte prévalence de certaines maladies auto-immunes chez les Africains.²⁸ En effet, les maladies auto-immunes sont associées à une réponse pro-inflammatoire inadéquate.³⁰ Par ailleurs, les Africains sont plus susceptibles que les Caucasiens ou les Asiatiques de subir un rejet de greffe lors d'une allogreffe rénale.¹²⁰ Cependant, nous savons que le test d'iHS manque de pouvoir statistique dans la population africaine à cause des patrons de déséquilibre de liaison qui y sont plus courts.¹²¹ Il est donc pour l'instant impossible d'écarter l'hypothèse que le manque de pouvoir statistique dans la population Africaine empêche de déceler des événements de sélection positive.

5.7 Perspectives

Il reste de nombreuses avenues à explorer afin de mieux caractériser les différences interpopulationnelles de la réponse immunitaire. La prochaine étape consistera à approfondir nos analyses sur les eQTLs. Pour ce faire, le nombre d'échantillons sera augmenté, jusqu'à l'obtention de 70 individus dans chaque population, afin de permettre un meilleur pouvoir de détection. Ainsi, il sera

possible d'identifier des eQTLs directement liés à la réponse immunitaire, c'est-à-dire des SNPs associés à l'expression de gènes qui présentent une variation de leur expression entre les états non infecté et infectés. Ces eQTLs associés à la réponse immunitaire seront de bons candidats pour affecter la réponse immunitaire de l'individu.¹²² Une fois que ces eQTLs associés à la réponse seront identifiés, il sera intéressant de voir s'ils sont spécifiques à une population et s'ils sont associés à des signatures de sélection naturelle. Ceci permettra de caractériser des gènes, ou plus globalement des voies de signalisation, associés à des différences interpopulationnelles dans l'expression de gènes après des infections bactériennes.

Par la suite, l'épissage alternatif sera étudié afin d'identifier des gènes dont l'utilisation des exons varie entre les populations dans le contexte immunitaire. Le programme MiSo sera utilisé afin de détecter les variations entre les populations des transcrits d'ARN produits à partir d'un même gène.¹²³ La même logique que celle utilisée pour les eQTLs sera appliquée afin d'identifier des SNPs associés à l'expression particulière d'un transcrit d'ARN. Encore une fois, il sera intéressant de voir si ces eQTLs sont spécifiques d'une population et s'ils sont associés à des signatures de sélection naturelle. Ces analyses nous permettront d'affiner notre compréhension et d'obtenir un portrait plus global des différences de la réponse immunitaire entre les populations.

6. Conclusion

Ce projet a permis de valider le fait qu'il existe des différences d'ordre génétique entre les populations qui expliquent des variabilités de la réponse immunitaire entre les populations d'origine européenne et africaine. Nous avons remarqué que les macrophages d'Africains ont une fonction bactéricide plus efficace que les macrophages d'Européens puisqu'ils éliminent les bactéries intracellulaires plus rapidement. En accord avec cette observation, nous avons constaté que les Européens sécrètent des quantités de cytokines anti-inflammatoires plus élevées que les Africains. L'étude du transcriptome des macrophages infectés et non infectés des Africains et des Européens a permis d'identifier plusieurs gènes différemment exprimés entre les populations, parmi lesquels nous avons relevé des gènes pro-inflammatoires significativement plus exprimés chez les Africains. Ces résultats indiquent que les Africains présentent une réponse immunitaire globale davantage axée sur la réponse pro-inflammatoire tandis que, du côté des Européens, elle est davantage axée sur la réponse anti-inflammatoire. Les tests de sélection naturelle ont indiqué que les gènes différemment exprimés entre les populations dans le contexte immunitaire sont enrichis pour des événements de sélection positive récents uniquement chez les Européens. Ces résultats démontrent que des événements de sélection naturelle ont eu lieu chez les individus qui sont sortis d'Afrique afin de leur permettre de s'adapter à leur nouvel environnement. Notre étude a permis d'identifier des mécanismes immunologiques qui peuvent expliquer les différences observées au niveau de l'incidence de certains troubles immunitaires entre les Européens et les Africains. La caractérisation des différences interindividuelles et interpopulationnelles de la réponse immunitaire représente une avancée essentielle vers le développement d'outils cliniques permettant une amélioration du diagnostic et du traitement des maladies infectieuses.

Bibliographie

1. Casanova, J.L. & Abel, L. Inborn errors of immunity to infection: the rule rather than the exception. *J Exp Med* **202**, 197-201 (2005).
2. The world health report 2004. (2004).
3. Frodsham, A.J. & Hill, A.V. Genetics of infectious diseases. *Hum Mol Genet* **13 Spec No 2**, R187-94 (2004).
4. Hill, A.V. Aspects of genetic susceptibility to human infectious diseases. *Annu Rev Genet* **40**, 469-86 (2006).
5. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).
6. Wang, E.T., Kodama, G., Baldi, P. & Moyzis, R.K. Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc Natl Acad Sci U S A* **103**, 135-40 (2006).
7. Wolfe, N.D., Dunavan, C.P. & Diamond, J. Origins of major human infectious diseases. *Nature* **447**, 279-83 (2007).
8. Eyre-Walker, A. & Keightley, P.D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344-7 (1999).
9. Kryukov, G.V., Pennacchio, L.A. & Sunyaev, S.R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* **80**, 727-39 (2007).
10. Hurst, L.D. Fundamental concepts in genetics: genetics and the understanding of selection. *Nat Rev Genet* **10**, 83-93 (2009).
11. Weir, B.S. & Hill, W.G. Estimating F-statistics. *Annu Rev Genet* **36**, 721-50 (2002).
12. Sabeti, P.C. et al. Positive natural selection in the human lineage. *Science* **312**, 1614-20 (2006).
13. Sabeti, P.C. et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-7 (2002).
14. Bamshad, M. & Wooding, S.P. Signatures of natural selection in the human genome. *Nat Rev Genet* **4**, 99-111 (2003).
15. Vallender, E.J. & Lahn, B.T. Positive selection on the human genome. *Hum Mol Genet* **13 Spec No 2**, R245-54 (2004).
16. Nielsen, R. et al. Genomic scans for selective sweeps using SNP data. *Genome Res* **15**, 1566-75 (2005).
17. Hughes, A.L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167-70 (1988).
18. Takahata, N., Satta, Y. & Klein, J. Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130**, 925-38 (1992).
19. Prugnolle, F. et al. Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol* **15**, 1022-7 (2005).
20. Barreiro, L.B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet* **11**, 17-30 (2010).
21. Beutler, E. G6PD deficiency. *Blood* **84**, 3613-36 (1994).
22. Allison, A.C. Two lessons from the interface of genetics and medicine. *Genetics* **166**, 1591-9 (2004).
23. Miller, L.H., Mason, S.J., Clyde, D.F. & McGinniss, M.H. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. *N Engl J Med* **295**, 302-4 (1976).

24. Tournamille, C., Colin, Y., Cartron, J.P. & Le Van Kim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet* **10**, 224-8 (1995).
25. Livingstone, F.B. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. *Hum Biol* **56**, 413-25 (1984).
26. Hamblin, M.T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* **66**, 1669-79 (2000).
27. Quintana-Murci, L. & Barreiro, L.B. The role played by natural selection on Mendelian traits in humans. *Ann N Y Acad Sci* **1214**, 1-17 (2010).
28. Pennington, R., Gatenbee, C., Kennedy, B., Harpending, H. & Cochran, G. Group differences in proneness to inflammation. *Infect Genet Evol* **9**, 1371-80 (2009).
29. Sorlie, P.D., Backlund, E. & Keller, J.B. US mortality by economic, demographic, and social characteristics: the National Longitudinal Mortality Study. *Am J Public Health* **85**, 949-56 (1995).
30. Richardus, J.H. & Kunst, A.E. Black-white differences in infectious disease mortality in the United States. *Am J Public Health* **91**, 1251-3 (2001).
31. Haraga, A., Ohlson, M.B. & Miller, S.I. Salmonellae interplay with host cells. *Nat Rev Microbiol* **6**, 53-66 (2008).
32. Francis, C.L., Starnbach, M.N. & Falkow, S. Morphological and cytoskeletal changes in epithelial cells occur immediately upon interaction with *Salmonella typhimurium* grown under low-oxygen conditions. *Mol Microbiol* **6**, 3077-87 (1992).
33. Jones, B.D., Ghori, N. & Falkow, S. *Salmonella typhimurium* initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. *J Exp Med* **180**, 15-23 (1994).
34. Takeuchi, A. & Sprinz, H. Electron-Microscope Studies of Experimental *Salmonella* Infection in the Preconditioned Guinea Pig: II. Response of the Intestinal Mucosa to the Invasion by *Salmonella typhimurium*. *Am J Pathol* **51**, 137-61 (1967).
35. Jepson, M.A., Collares-Buzato, C.B., Clark, M.A., Hirst, B.H. & Simmons, N.L. Rapid disruption of epithelial barrier function by *Salmonella typhimurium* is associated with structural modification of intercellular junctions. *Infect Immun* **63**, 356-9 (1995).
36. McCormick, B.A. et al. Surface attachment of *Salmonella typhimurium* to intestinal epithelia imprints the subepithelial matrix with gradients chemotactic for neutrophils. *J Cell Biol* **131**, 1599-608 (1995).
37. Garai, P., Gnanadhas, D.P. & Chakravorty, D. *Salmonella enterica* serovars Typhimurium and Typhi as model organisms: revealing paradigm of host-pathogen interactions. *Virulence* **3**, 377-88 (2012).
38. Hansen-Wester, I. & Hensel, M. *Salmonella* pathogenicity islands encoding type III secretion systems. *Microbes Infect* **3**, 549-59 (2001).
39. Janssen, R., van der Straaten, T., van Diepen, A. & van Dissel, J.T. Responses to reactive oxygen intermediates and virulence of *Salmonella typhimurium*. *Microbes Infect* **5**, 527-34 (2003).
40. Chakravorty, D., Hansen-Wester, I. & Hensel, M. *Salmonella* pathogenicity island 2 mediates protection of intracellular *Salmonella* from reactive nitrogen intermediates. *J Exp Med* **195**, 1155-66 (2002).
41. Gunn, J.S., Ryan, S.S., Van Velkinburgh, J.C., Ernst, R.K. & Miller, S.I. Genetic and functional analysis of a PmrA-PmrB-regulated locus necessary for lipopolysaccharide modification,

- antimicrobial peptide resistance, and oral virulence of *Salmonella enterica* serovar typhimurium. *Infect Immun* **68**, 6139-46 (2000).
42. Garcia-del Portillo, F., Zwick, M.B., Leung, K.Y. & Finlay, B.B. Salmonella induces the formation of filamentous structures containing lysosomal membrane glycoproteins in epithelial cells. *Proc Natl Acad Sci U S A* **90**, 10544-8 (1993).
 43. Uchiya, K. et al. A Salmonella virulence protein that inhibits cellular trafficking. *EMBO J* **18**, 3924-33 (1999).
 44. Shotland, Y., Kramer, H. & Groisman, E.A. The Salmonella SpiC protein targets the mammalian Hook3 protein function to alter cellular trafficking. *Mol Microbiol* **49**, 1565-76 (2003).
 45. Kuhle, V., Jackel, D. & Hensel, M. Effector proteins encoded by Salmonella pathogenicity island 2 interfere with the microtubule cytoskeleton after translocation into host cells. *Traffic* **5**, 356-70 (2004).
 46. Sixl, W., Stunzner, D. & Withalm, H. [Epidemiologic and serologic study of listeriosis in man and domestic and wild animals in Austria]. *J Hyg Epidemiol Microbiol Immunol* **22**, 460-9 (1978).
 47. Gellin, B.G. & Broome, C.V. Listeriosis. *JAMA* **261**, 1313-20 (1989).
 48. Pamer, E.G. Immune responses to *Listeria monocytogenes*. *Nat Rev Immunol* **4**, 812-23 (2004).
 49. Lecuit, M. Human listeriosis and animal models. *Microbes Infect* **9**, 1216-25 (2007).
 50. Stavru, F., Archambaud, C. & Cossart, P. Cell biology and immunology of *Listeria monocytogenes* infections: novel insights. *Immunol Rev* **240**, 160-84 (2011).
 51. Freitag, N.E. From hot dogs to host cells: how the bacterial pathogen *Listeria monocytogenes* regulates virulence gene expression. *Future Microbiol* **1**, 89-101 (2006).
 52. Camejo, A. et al. The arsenal of virulence factors deployed by *Listeria monocytogenes* to promote its cell infection cycle. *Virulence* **2**, 379-94 (2011).
 53. Freitag, N.E., Port, G.C. & Miner, M.D. *Listeria monocytogenes* - from saprophyte to intracellular pathogen. *Nat Rev Microbiol* **7**, 623-8 (2009).
 54. Gaillard, J.L., Berche, P., Frehel, C., Gouin, E. & Cossart, P. Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. *Cell* **65**, 1127-41 (1991).
 55. Bielecki, J., Youngman, P., Connelly, P. & Portnoy, D.A. *Bacillus subtilis* expressing a haemolysin gene from *Listeria monocytogenes* can grow in mammalian cells. *Nature* **345**, 175-6 (1990).
 56. Domann, E. et al. A novel bacterial virulence gene in *Listeria monocytogenes* required for host cell microfilament interaction with homology to the proline-rich region of vinculin. *EMBO J* **11**, 1981-90 (1992).
 57. Kocks, C. et al. *L. monocytogenes*-induced actin assembly requires the actA gene product, a surface protein. *Cell* **68**, 521-31 (1992).
 58. Tilney, L.G. & Portnoy, D.A. Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J Cell Biol* **109**, 1597-608 (1989).
 59. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
 60. Nica, A.C. et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6**, e1000895 (2010).
 61. Cheung, V.G. et al. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365-9 (2005).

62. Monks, S.A. et al. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* **75**, 1094-105 (2004).
63. Doss, S., Schadt, E.E., Drake, T.A. & Lusis, A.J. Cis-acting expression quantitative trait loci in mice. *Genome Res* **15**, 681-91 (2005).
64. Schadt, E.E. et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297-302 (2003).
65. Stranger, B.E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet* **1**, e78 (2005).
66. Stranger, B.E. et al. Population genomics of human gene expression. *Nat Genet* **39**, 1217-24 (2007).
67. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506-11 (2013).
68. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63 (2009).
69. Pickrell, J.K. et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768-72 (2010).
70. Metzker, M.L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46 (2010).
71. FastQC. (2010).
72. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
73. HTSeq: Analysing high-throughput sequencing data with Python. (2010).
74. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**, 480 (2011).
75. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
76. Butler, H. & Ragoussis, J. BeadArray-based genotyping. *Methods Mol Biol* **439**, 53-74 (2008).
77. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
78. Zakharia, F. et al. Characterizing the admixed African ancestry of African Americans. *Genome Biol* **10**, R141 (2009).
79. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655-64 (2009).
80. Leng, S.X. et al. ELISA and multiplex technologies for cytokine measurement in inflammation and aging research. *J Gerontol A Biol Sci Med Sci* **63**, 879-84 (2008).
81. Backes, C. et al. GeneTrail--advanced gene set enrichment analysis. *Nucleic Acids Res* **35**, W186-92 (2007).
82. Opal, S.M. & DePalo, V.A. Anti-inflammatory cytokines. *Chest* **117**, 1162-72 (2000).
83. Kobayashi, Y. The role of chemokines in neutrophil biology. *Front Biosci* **13**, 2400-7 (2008).
84. Grewal, I.S. & Flavell, R.A. CD40 and CD154 in cell-mediated immunity. *Annu Rev Immunol* **16**, 111-35 (1998).
85. The International HapMap Project. *Nature* **426**, 789-96 (2003).
86. Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10**, 5-6 (2013).
87. Paten, B. et al. Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* **18**, 1829-43 (2008).

88. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* **18**, 1814-28 (2008).
89. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176-7 (2012).
90. Kasai, T. et al. Anti-inflammatory cytokine levels in patients with septic shock. *Res Commun Mol Pathol Pharmacol* **98**, 34-42 (1997).
91. Munoz, C. et al. Dysregulation of in vitro cytokine production by monocytes during sepsis. *J Clin Invest* **88**, 1747-54 (1991).
92. Dinarello, C.A. Biologic basis for interleukin-1 in disease. *Blood* **87**, 2095-147 (1996).
93. Sims, J.E. & Smith, D.E. The IL-1 family: regulators of immunity. *Nat Rev Immunol* **10**, 89-102 (2010).
94. Danis, V.A., Millington, M., Hyland, V.J. & Grennan, D. Cytokine production by normal human monocytes: inter-subject variation and relationship to an IL-1 receptor antagonist (IL-1Ra) gene polymorphism. *Clin Exp Immunol* **99**, 303-10 (1995).
95. Witkin, S.S., Gerber, S. & Ledger, W.J. Influence of interleukin-1 receptor antagonist gene polymorphism on disease. *Clin Infect Dis* **34**, 204-9 (2002).
96. Fang, X.M., Schroder, S., Hoefl, A. & Stuber, F. Comparison of two polymorphisms of the interleukin-1 gene family: interleukin-1 receptor antagonist polymorphism contributes to susceptibility to severe sepsis. *Crit Care Med* **27**, 1330-4 (1999).
97. Saraiva, M. & O'Garra, A. The regulation of IL-10 production by immune cells. *Nat Rev Immunol* **10**, 170-81 (2010).
98. Murray, P.J. Understanding and exploiting the endogenous interleukin-10/STAT3-mediated anti-inflammatory response. *Curr Opin Pharmacol* **6**, 379-86 (2006).
99. Opal, S.M., Wherry, J.C. & Grint, P. Interleukin-10: potential benefits and possible risks in clinical infectious diseases. *Clin Infect Dis* **27**, 1497-507 (1998).
100. Lehmann, A.K., Halstensen, A., Sornes, S., Rokke, O. & Waage, A. High levels of interleukin 10 in serum are associated with fatality in meningococcal disease. *Infect Immun* **63**, 2109-12 (1995).
101. Lee, K.S. et al. IL-10 suppresses bactericidal response of macrophages against *Salmonella* Typhimurium. *J Microbiol* **49**, 1050-3 (2011).
102. Wynn, T.A. IL-13 effector functions. *Annu Rev Immunol* **21**, 425-56 (2003).
103. Zurawski, G. & de Vries, J.E. Interleukin 13, an interleukin 4-like cytokine that acts on monocytes and B cells, but not on T cells. *Immunol Today* **15**, 19-26 (1994).
104. Mijatovic, T., Kruys, V., Caput, D., Defrance, P. & Huez, G. Interleukin-4 and -13 inhibit tumor necrosis factor- α mRNA translational activation in lipopolysaccharide-induced mouse macrophages. *J Biol Chem* **272**, 14394-8 (1997).
105. Di Santo, E. et al. IL-13 inhibits TNF production but potentiates that of IL-6 in vivo and ex vivo in mice. *J Immunol* **159**, 379-82 (1997).
106. Wang, E.T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-6 (2008).
107. Elton, T.S., Selemon, H., Elton, S.M. & Parinandi, N.L. Regulation of the MIR155 host gene in physiological and pathological processes. *Gene* **532**, 1-12 (2013).
108. Bushati, N. & Cohen, S.M. microRNA functions. *Annu Rev Cell Dev Biol* **23**, 175-205 (2007).
109. Taganov, K.D., Boldin, M.P., Chang, K.J. & Baltimore, D. NF- κ B-dependent induction of microRNA miR-146, an inhibitor targeted to signaling proteins of innate immune responses. *Proc Natl Acad Sci U S A* **103**, 12481-6 (2006).

110. McCoy, C.E. et al. IL-10 inhibits miR-155 induction by toll-like receptors. *J Biol Chem* **285**, 20492-8 (2010).
111. Silha, J.V. & Murphy, L.J. Serum resistin (FIZZ3) protein is increased in obese humans. *J Clin Endocrinol Metab* **89**, 1977; author reply 1977-8 (2004).
112. McTernan, C.L. et al. Resistin, central obesity, and type 2 diabetes. *Lancet* **359**, 46-7 (2002).
113. Stejskal, D., Adamovska, S., Bartek, J., Jurakova, R. & Proskova, J. Resistin - concentrations in persons with type 2 diabetes mellitus and in individuals with acute inflammatory disease. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* **147**, 63-9 (2003).
114. Johnson, C.M. et al. Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. *J Immunol* **178**, 7520-4 (2007).
115. Barreiro, L.B. et al. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet* **5**, e1000562 (2009).
116. Ness, R.B., Haggerty, C.L., Harger, G. & Ferrell, R. Differential distribution of allelic variants in cytokine genes among African Americans and White Americans. *Am J Epidemiol* **160**, 1033-8 (2004).
117. Pradhan, A.D., Manson, J.E., Rifai, N., Buring, J.E. & Ridker, P.M. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *JAMA* **286**, 327-34 (2001).
118. Kelley-Hedgpeth, A. et al. Ethnic differences in C-reactive protein concentrations. *Clin Chem* **54**, 1027-37 (2008).
119. Simhan, H.N., MacPherson, T., Caritis, S.N. & Krohn, M.A. Maternal and fetal Toll-like receptor 4 genotype and chorionic plate inflammatory lesions. *Am J Obstet Gynecol* **199**, 400 e1-4 (2008).
120. Young, C.J. & Gaston, R.S. Renal transplantation in black Americans. *N Engl J Med* **343**, 1545-52 (2000).
121. Liu, X. et al. Detecting and characterizing genomic signatures of positive selection in global populations. *Am J Hum Genet* **92**, 866-81 (2013).
122. Barreiro, L.B. et al. Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proc Natl Acad Sci U S A* **109**, 1204-9 (2012).
123. Katz, Y., Wang, E.T., Airoidi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat Methods* **7**, 1009-15 (2010).