



Université
de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Pathologie, Toxicologie, Génétique et Nutrition

Présentée et soutenue par :

M. NATHAN CEBRON

le mardi 14 décembre 2021

Titre :

Analyse intégrative des données multi-dimensionnelles pour l'étude de la vaccination vis-à-vis des infections mammaires et pulmonaires chez les bovins

Ecole doctorale :

Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries (SEVAB)

Unité de recherche :

Interactions Hôtes - Agents Pathogènes (IHAP)

Directeur de Thèse :

M. GILLES FOUCRAS

Rapporteurs :

M. FRANÇOIS MEURENS, ONIRIS - NANTES

MME FLORENCE JAFFREZIC, INRAE JOUY-EN-JOSAS

Membres du jury :

M. DIDIER CONCORDET, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Président

M. GILLES FOUCRAS, ECOLE NATIONALE VETERINAIRE DE TOULOUSE, Membre

MME SABINE RIFFAULT, INRAE JOUY-EN-JOSAS, Membre

M. PIERRE GERMON, INRAE VAL DE LOIRE, Membre

Table des matières

Liste des abréviations	3
Liste des figures	5
Liste des tableaux	5
1. Introduction générale	6
1.1. Santé des bovins en élevage et enjeux	6
1.1.1. Les principales maladies bovines	6
1.1.2. Les mesures de maîtrise des maladies des bovins	27
1.1.3. Les enjeux autour de la santé des bovins.....	40
1.2. Sciences animales et biostatistiques	43
1.2.1. Historique des biostatistiques et de l'analyse de données	43
1.2.2. Les champs d'application	49
1.2.3. Démarche expérimentale et méthodes statistiques appliquées à la santé des bovins	52
1.3. Méthodes statistiques appliquées aux données de grandes dimensions	56
1.3.1. De la biologie moléculaire à la biologie multi-niveaux.....	56
1.3.2. Les différents types de données omiques collectées à haut-débit	60
1.3.3. Méthodes d'analyse des données transcriptomiques par séquençage de l'ARN (RNA-seq)	71
1.3.4. Méthodes d'analyse de données non transcriptomiques.....	87
1.3.5. La représentation et l'intégration de données	90
2. Etude expérimentale.....	108
2.1. Objectif de la thèse	108
2.2. Etude 1 : Etude transcriptomique des mécanismes de protection permis par la vaccination vis-à-vis d'une mammité à <i>E. coli</i> chez la vache laitière	109
2.2.1. Objectif et protocole	109
2.2.2. Résultats.....	110
2.2.3. Conclusion générale	112

2.3. Etude 2 : Etude d'un protocole de préparation sur les paramètres sanitaires, immunitaires et les performances zootechniques de veaux dans des ateliers d'engraissement	126
2.3.1. Objectif et protocole	126
2.3.2. Résultats.....	127
2.3.3. Discussion.....	128
Results	134
3. Discussion	161
3.1. La vaccination, encore beaucoup de limites ?	161
3.2. La transcriptomique et la protéomique ciblée, comme moyens d'étude de la réponse immunitaire.....	163
3.3. Intégration des données multidimensionnelles	166
3.4. Réduction des données expérimentales	168
3.5. Et après ?!	169
4. Bibliographie.....	170
5. Productions scientifiques	196
5.1. Publications	196
5.2. Communications orales	196
5.3. Posters.....	196

Liste des abréviations

ACP : Analyse en composantes principales	CLR : Récepteurs de lectine de type C
ADCC : Cytotoxicité cellulaire dépendante des anticorps	CMH : complexe majeur d'histocompatibilité
Adj : Adjuvant	CV : Coefficient de variation
ADN : Acide DésoxyriboNucléique	DBSCAN : density-based spatial clustering of applications with noise
AFC : Analyse factorielle des correspondances	<i>E. coli</i> : <i>Escherichia coli</i>
AFDM : Analyse factorielle des données mixtes	ELISA : enzyme-linked immunosorbent assay
AFM : Analyse factorielle multiple	FDR : False Discovery Rate
ANOVA : Analyse de la variance	GCN : réseaux de co-expression
ARN : Acide RiboNucléique	GMQ : Gain moyen quotidien
ARNm : Acide RiboNucléique messenger	GO : Gene ontology
ARNr : Acide RiboNucléique ribosomique	GSEA : Gene Set Enrichment Analysis
ARNt : Acide RiboNucléique transfert	GTF : Gene transfer format
BAM : Binary alignment map	<i>H.somni</i> : <i>Histophilus somni</i>
<i>BAV</i> : <i>Adenovirus bovin</i>	HCPC : Classification Hiérarchique sur Composantes Principales
BCL : Binary base call	HGNC : HUGO Gene Nomenclature Committee
<i>BCoV</i> : <i>Coronavirus bovin</i>	Hpi : Heure post infection
BH : Benjamin-Hochberg	HUGO : Human Genome Organisation
BMAP : Bovine Myeloid Antimicrobial Peptide	IA : Intelligence artificiel
BNBD : β -défensine	IFN : Interferon
<i>BoHV-1</i> : <i>Herpès virus bovin</i>	Ig : Immunoglobuline
BPI : Bronchopneumonies infectieuses	IL : Interleukine
<i>BPI3</i> : <i>Parainfluenza virus 3</i>	ILC : Innate lymphoid cells
<i>BRSV</i> : <i>Virus respiratoire syncytial bovin</i>	IM : Intramusculaire
<i>BVDV</i> : <i>Virus de la diarrhée virale bovine</i>	IMM : Intramammaire
CAH : Classification ascendante hiérarchique	iNOS : Nitric oxide synthase
CCA : Analyse canonique des corrélations	IPA : Ingenuity pathway analysis
CCS : Cellules somatiques	<i>K.pneumoniae</i> : <i>Klebsiella pneumoniae</i>
CEMITool : Co-expression modules identification Tool	KEGG : Kyoto Encyclopedia of Genes and Genomes

LAP : Lingual Antimicrobial Peptide

LPS : Lipopolysaccharide

M. bovis : *Mycobacterium bovis*

MEC : Cellules épithéliales mammaires

MERFISH : Multiplexed error-robust fluorescence in situ hybridization

MFI : Mean fluorescence intensity

Mh : *Mannheimia haemolytica*

miRNA : Micro acide RiboNucléique

MSigDB : Molecular Signatures Database

NCBI : National Center for Biotechnology Information

NK : Natural killer

NLR : Récepteurs de type Nod

P. multocida : *Pasteurella multocida*

PADOG : Pathway Analysis with Down Weighting of Overlapping Genes

PAMP : Pathogen-Associated Molecular Patterns

PBMC : Peripheral blood mononuclear cell

PCR : Polymerase Chain Reaction

PLS : Partial Least Squares Regression

PRR : Pattern recognition recepteurs

RIN : RNA integrity number

RLE : Relative log-expression

RLR : Récepteurs de type RIG-1

RNA-seq : séquençage de l'ARN

RSEM : RNA-Seq by Expectation-Maximization

S. aureus : *Staphylococcus aureus*

S. uberis : *Streptococcus uberis*

S. agalactiae : *Streptococcus agalactiae*

S. dysgalactiae : *Streptococcus dysgalactiae*

SAA : Sérum amyloïde A

SAAC : Slime associated antigenic complex

SAM : Sequence alignment map

SCS : Score cellules somatiques

SEA : Simpler Enrichment Analysis

SeqFISH : sequential Fluorescence In Situ Hybridization

SGSE : Spectral Gene Set Enrichment

siRNA : Acide RiboNucléique interférent

SNP : Polymorphisme d'un seul nucléotide

STAR : Spliced Transcripts Alignment to a Reference

String : Search tool for the retrieval of interacting genes / proteins

SUAM : Streptococcus uberis adhesion molecule

TAP : Tracheal Antimicrobial Peptide

TB / TP : taux butyreux / taux protéique

Th : T helper

TLR : Toll-like récepteurs

TMM : Trimmed Mean of M-values

TNF : tumor necrosis factor

WGCNA : Weighted correlation network analysis

Liste des figures

Figure 1 : Morphologie d'une glande mammaire de vache

Figure 2 : Compartimentation pulmonaire d'un veau

Figure 3 : Représentation schématique des relations entre les différents termes liés à l'intelligence artificiel

Figure 4 : Les différences entre les études épidémiologiques et les études individus centrées

Figure 5 : Protocole détaillé de la technologie Luminex

Figure 6 : Schéma des relations entre les différentes couches d'informations au sein d'un échantillon biologique

Figure 7 : Etapes pour la préparation d'échantillons pour le séquençage

Figure 8 : Bilan des méthodes de purification des ARNm

Figure 9 : Fonctionnement du séquençage

Figure 10 : Fonctionnement du demultiplexage de séquences

Figure 11 : Effet du nettoyage des séquences via le logiciel Trimalore

Figure 12 : Schéma résumant les différentes méthodes de réduction et visualisation de données multivariées et leur condition d'utilisation

Figure 13 : Fonctionnement de la déconvolution cellulaire

Liste des tableaux

Table 1 : Bilan des caractéristiques des pathogènes vis-à-vis des mammites

Table 2 : Différents types d'agents intervenant dans les BPI

Table 3 : Comparaison des méthodes de classifications les plus utilisées

Table 4 : Détails sur les cytokines et chimiokines les plus connues

Table 5 : Etude entre les différents types d'aligneurs utilisées dans l'analyse RNA-seq

Table 6 : Principales méthodes de normalisation de données RNA-seq

Table 7 : Recapitulatif des méthodes de déconvolution cellulaire les plus utilisées dans l'étude immunitaire

1.Introduction générale

1.1. Santé des bovins en élevage et enjeux

1.1.1. *Les principales maladies bovines*

1.1.1.1. Glande mammaire : mammites

1.1.1.1.1. Introduction

La mammite est une inflammation de la glande mammaire généralement due à une colonisation bactérienne à travers l'extrémité du canal du trayon. La glande mammaire est une glande spécialisée, spécifique des mammifères permettant la production du lait, pour l'alimentation du jeune au cours des premières semaines de vie. Chez les bovins, qui est notre espèce d'étude, elle est formée de quatre quartiers non communicants, mais qui ne sont pas totalement indépendants du point de vue de la réponse inflammatoire et immunitaire. Il n'y a pas de passage direct de bactéries entre les quartiers. Lors d'une mammite, c'est le plus souvent un seul quartier qui est atteint.

La mamelle est composée d'alvéoles où a lieu la synthèse du lait ; elles sont reliées entre elles à la citerne par des canaux galactophores. Dans l'intervalle de temps entre les traites ou les tétées, le volume de lait sécrété est distribué dans 3 régions : 60% du volume est contenu dans les alvéoles, 20% à l'intérieur des canaux et seulement 20% est présent dans la citerne. L'excrétion du lait se fait par le sinus du trayon correspond à un canal de 3 à 10 cm avec de nombreux vaisseaux et nerfs. Celui-ci est suivi par le canal du trayon d'une longueur de 0,5 à 1,3 cm (**Figure 1**). Tout ce système est très innervé et vascularisé, avec un débit moyen de 300 litres de sang par heure qui transitent par la mamelle. Ce canal peut être fermé avec un diamètre de 0,4 mm ou ouvert avec un diamètre de 1 à 2 mm. La glande permet la production de lait dont la qualité peut varier en fonction du contexte environnemental et sanitaire de l'animal, et de son potentiel de production comme résultat de la sélection génétique sur ce critère.

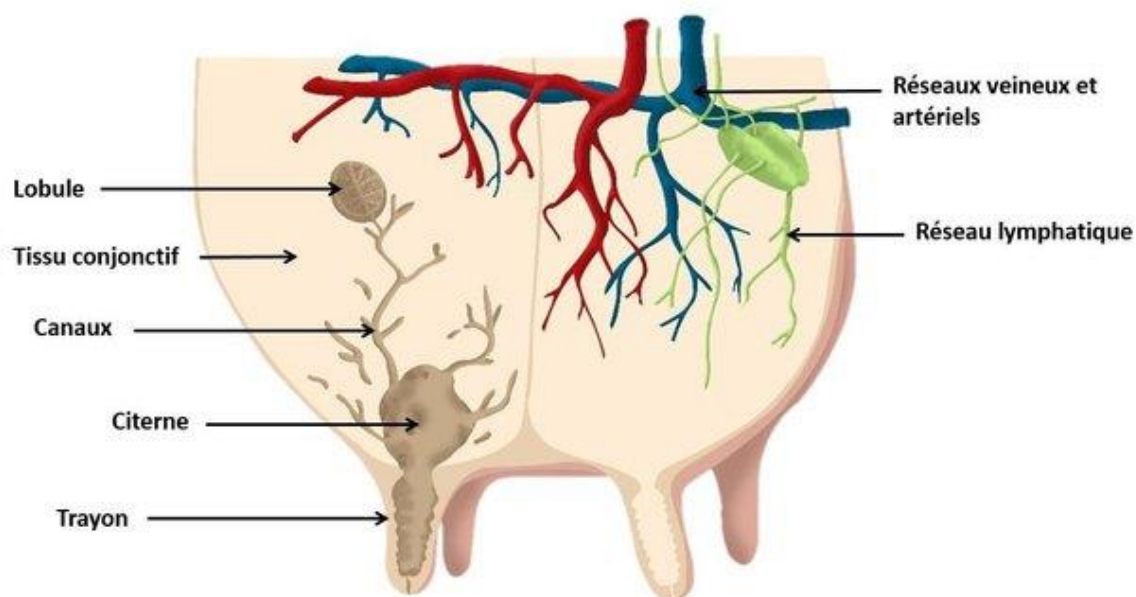


Figure 1 : Morphologie d'une glande mammaire de vache (C.Charton, thèse)

L'un des principaux indicateurs permettant d'étudier la qualité sanitaire du lait est le dénombrement de cellules somatiques qu'il contient. Ces cellules ont 2 origines possibles : soit ce sont des cellules sanguines correspondant aux globules blancs, soit il s'agit de cellules épithéliales du tissu qui s'exfolient au cours du processus de renouvellement de l'épithélium. Néanmoins, ces cellules épithéliales ne représentent généralement qu'au plus 10% des cellules somatiques totales. Les leucocytes présents dans le lait sont composés de nombreux types cellulaires différents, tels que les macrophages, les lymphocytes, ou encore les granulocytes dont principalement les neutrophiles. Lors de mammite, le nombre de cellules augmente très fortement indiquant une activation de la réponse inflammatoire et une réponse immunitaire vis-à-vis de la bactérie qui l'a initiée. La recherche de la présence de ces cellules est systématiquement réalisée à chaque traite pour la détection des mammites. Diverses méthodes sont utilisées : de l'observation de la présence de grumeaux constitués de l'agrégation de cellules jusqu'à la quantification automatisée du nombre de cellules. Au-delà d'un certain seuil du nombre de cellules somatiques, le lait est considéré comme impropre à la consommation humaine. Une vache atteinte de mammite peut avoir plusieurs millions de cellules somatiques par mL de lait alors que la norme est considérée à 150 000 cellules par mL. Le seuil réglementaire européen est fixé à 400 000 cellules par mL alors que plus de 200 000 cellules par mL indiquent une mammite.

Les mammites peuvent être de présentation assez diverse en fonction de l'observation de signes cliniques : ainsi, elles sont dites cliniques si des symptômes locaux ou généraux sont présents, alors qu'elles sont sub-cliniques dans les autres cas. Les mammites cliniques peuvent être aiguës, voire suraiguës si la durée d'apparition est très courte, ou bien chroniques voire récurrentes, si les symptômes persistent ou reviennent à plusieurs reprises. Les mammites cliniques sont caractérisées par la présence de symptômes généraux (hyperthermie, perte d'appétit, arumination), et/ou locaux (rougeur, œdème, chaleur, douleur) et surtout par la modification de l'aspect du lait dont la couleur peut être modifiée et qui contient des grumeaux. En fonction de la nature et de l'intensité des symptômes, elles peuvent être considérées comme bénignes en l'absence d'inflammation de la mamelle malgré l'altération de l'aspect du lait, modérées s'il y a inflammation, voire sévères si des signes systémiques sont présents et que l'évolution vers le décubitus, voire la mort s'ensuit.

Les mammites subcliniques sont caractérisées par une augmentation de la quantité de cellules somatiques dans le lait comme conséquence d'une activation de la réponse inflammatoire [1].

Les infections mammaires peuvent avoir une source environnementale ou évoluer selon un mode contagieux. Les mammites dites d'environnement sont dues à la contamination du trayon au contact du sol ou de la litière entre les traites. Les mammites contagieuses se transmettent au moment de la traite entre les vaches via l'équipement de traite ou les mains de l'éleveur. L'un des facteurs de risque majeurs est constitué par l'hygiène de l'environnement de vie des animaux, avec la présence d'une densité anormalement grande de bactéries avec un pouvoir infectieux pour la mamelle. Les conditions de réalisation de la traite (qualité des équipements, hygiène, technique) jouent également un rôle déterminant dans l'occurrence des mammites évoluant selon un mode contagieux.

Les mammites sont la première affection d'origine infectieuse en élevage laitier, avec plus de 40% des animaux en production qui sont atteints. Plus des 2/3 des mammites ont lieu au cours du premier mois de la lactation [2].

1.1.1.1.2. Pathogènes responsables

La majorité des mammites est causée par six espèces bactériennes. La sévérité et l'évolution de la mammite peut varier en fonction de la bactérie, mais d'autres facteurs dont ceux liés à la réponse de l'hôte font varier la présentation clinique (**Table 1**).

- *Escherichia coli* cause des mammites dites d'environnement car la grande majorité d'entre elles est acquise par une contamination liée au milieu de vie, avec une sévérité

généralement marquée pour des mammites évoluant selon un mode aigu à suraigu. Elle est l'un des pathogènes majeurs responsables des mammites dans les élevages laitiers de bon niveau technique dans lesquels les mammites contagieuses sont correctement maîtrisées. Une infection par *E. coli* peut causer une mammite létale si elle n'est pas traitée, et elle correspond à la cause de mortalité la plus courante en relation avec une mammite. La sévérité clinique à la suite de l'infection est liée à de nombreux facteurs tels que le caractère plus ou moins pathogène de la souche et la réponse de l'hôte via l'activation des neutrophiles [3]. Ce pathogène n'infecte pas le tissu profond de la mamelle comme le parenchyme alvéolaire et le tissu mammaire et reste dans les canaux. Les périodes où le risque d'infection est majeur correspondent aux deux semaines après le tarissement, les deux dernières semaines avant la mise bas mais aussi le début de la lactation [4]. En effet, 50% des cas de mammites cliniques sont dues à des infections acquises durant la période sèche [5]. Lors d'une infection, en parallèle du développement de bactéries au sein de la glande mammaire, une augmentation de la production de cytokines inflammatoires (IL-6, IL-22, TNF- α) et anti-inflammatoires sont présentes ainsi qu'un fort recrutement cellulaire, notamment de neutrophiles. Une déplétion des neutrophiles induit une augmentation de la charge bactérienne. Concernant la réponse immunitaire adaptative, la contribution de la réponse lymphocytaire T CD4+ produisant de l'IL-17A est majeure. En son absence, la multiplication bactérienne est plus élevée. La cytokine IL-17A a pour fonction l'activation des neutrophiles. Cette réponse a donc un intérêt majeur dans le développement de méthodes de vaccination [6].

- *Klebsiella pneumoniae* est considérée comme une bactérie causant des mammites dont la source infectieuse est environnementale, mais elle peut aussi être transmise à partir d'une vache infectée. C'est l'une des bactéries les plus abondantes dans les lieux de vie des animaux. Elle a une capacité plus grande à infecter la glande mammaire que *E. coli*. Tout comme les infections à *E. coli*, celles provoquées par *K. pneumoniae* commencent à la fin de la période sèche sous la forme de mammites sub-cliniques avant d'évoluer en mammites cliniques au début de la lactation. La physiopathologie de cette infection est peu documentée par comparaison aux infections à *E. coli*. Néanmoins, une étude récente a montré que la réponse immunitaire induite était similaire entre ces deux types d'agents [7]. La sévérité des mammites induites par *K. pneumoniae* est souvent grande, avec notamment des cas létaux qui ont été documentés [8]. Ceux-ci sont dus à une forte dissémination de la bactérie dans la mamelle, avec même la possibilité d'une septicémie.

Dans ces cas, une forte diminution de la production de lait est généralement observée [9].

Ces deux premières bactéries sont la cause de mammites cliniques pouvant être sévères liées à l'environnement au contraire de *S. aureus*, qui elle entraîne des mammites sub-cliniques.

- *Staphylococcus aureus* est responsable de mammites évoluant selon un mode contagieux et une sévérité qui est généralement peu marquée, mais elle établit des infections chroniques dont la durée d'évolution est généralement longue, avec des mammites difficilement curables. Cette bactérie est la principale cause des mammites chroniques. L'utilisation d'antibiotiques est le seul traitement possible avec un taux de guérison faible car *S. aureus* provoque la formation de micro-abcès dans lesquels la diffusion de l'antibiotique est insuffisante pour éliminer la bactérie [10]. De plus, cette bactérie a une forte capacité à résister à l'activité des neutrophiles [11]. Une infection par *S. aureus* cause une baisse de la production laitière sur le long terme car cette bactérie cause des dommages irréversibles du tissu mammaire sécrétoire. Une majorité des infections sont asymptomatiques. La transmission se fait majoritairement d'animal à animal. Une étude a montré que 12 à 15% des vaches en première lactation sont contaminées avec pour la majorité, peu de symptômes. Ces animaux peuvent constituer une source de contamination pour leurs congénères [12]. La prévalence des mammites cliniques est de 10 à 12% ; néanmoins la majorité des mammites restent sub-cliniques et difficilement détectables sans recours aux examens complémentaires [13].
- *Streptococcus uberis* est responsable de mammites qui peuvent évoluer dans les troupeaux laitiers, selon les deux modes de transmission, avec une sévérité assez forte mais une persistance assez faible. Selon certaines études [14], [15], la transmission par contagion serait prédominante. Toutefois, une source environnementale ne peut pas être exclue car cette bactérie est présente dans le milieu de vie des animaux, et le portage cutané est possible. La sévérité de la mammite dépend de la réponse immunitaire de l'animal et de l'isolat bactérien, certaines étant plus contagieuses et virulentes que d'autres. La majorité des infections à *S. uberis* ont lieu durant la période sèche avec une forme sub-clinique. La pathogénie est proche de celle de *S. aureus* avec des dommages du parenchyme alvéolaire ; en outre, la réponse aux traitements antibiotiques est souvent assez médiocre [16]. Dans les cas où l'infection persiste, une réinfection ou une réactivation est possible [17].

- *Streptococcus agalactiae* est l'agent typique d'une transmission selon le mode contagieux, notamment au moment de la collecte du lait. Les isolats issus de mammites et isolés chez l'homme partagent 58% de similarités, indiquant une possible transmission liée au personnel [18]. Cette bactérie infecte presque exclusivement la mamelle, et peut survivre quelques temps sur les mains du personnel. La prévalence est grande dans les élevages avec une faible hygiène. Elle est majoritairement présente dans les pays pauvres ou en voie de développement mais a quasiment disparu dans les pays développés du fait des conditions d'hygiène de la traite et des traitements antibiotiques appliqués lors d'infection. Cette bactérie peut être facilement éradiquée lorsque des bonnes pratiques de traite sont appliquées [19].
- *Streptococcus dysgalactiae* est à l'origine de mammites avec une sévérité assez marquée et une persistance faible à modérée. Des études ont montré que la voie d'infection par cette bactérie fait encore débat [20]. En effet, celle-ci est présente de façon diffuse dans la glande, mais aussi dans le tube digestif et les locaux hébergeant les animaux. Elle provoque des mammites avec une forte inflammation.

Une majorité des mammites détectées durant la période de lactation ont été contractées durant la période sèche [21].

Table 1 : Bilan des caractéristiques des pathogènes

Bactérie	Mode de contamination	Type de mammité	Sévérité
<i>E.coli</i>	Environnementale	Clinique	+++
<i>K.pneumoniae</i>	Environnementale / Contagieuse	Clinique	+++
<i>S.aureus</i>	Contagieuse	Sub-clinique	+
<i>S.uberis</i>	Environnementale / Contagieuse	Clinique	++
<i>S.agalactiae</i>	Contagieuse	Clinique	++
<i>S.dysgalactiae</i>	Environnementale ou Contagieuse	Clinique	++

Légende : +++ Forte ++ Modérée + Bégnine

1.1.1.2. Sphère pulmonaire et broncho-pneumonies infectieuses

1.1.1.2.1. Définition

Les bronchopneumonies d'origine infectieuse (ou BPI) sont des troubles respiratoires touchant majoritairement les jeunes bovins. Cet ensemble de maladies respiratoires entraînent une augmentation de la morbidité, de la mortalité, et une diminution de la croissance, du gain de poids et de la qualité de la carcasse à l'abattage, avec un impact économique marqué dans la filière de production de viande. Par ailleurs, elles sont responsables de l'altération de la santé et du bien-être de l'animal. La prévalence globale des troubles respiratoires est d'environ 2%, avec néanmoins une prévalence plus élevée entre la naissance et les six premiers mois de vie. Cette période est considérée comme celle où le risque de développer une broncho-pneumonie est le plus grand, à cause d'un système immunitaire encore immature chez les veaux de moins de six mois et une augmentation marquée des facteurs d'exposition et de prédisposition liés au système d'élevage. Ainsi, la saison, notamment l'hiver, mais aussi le sexe mâle sont autant de facteurs favorisant l'apparition des BPI [22]. La mortalité qui en résulte varie de 4 à 20% en fonction du contexte et du(es) agent(s) responsable(s). En cas d'infection virale, le taux de morbidité peut être élevé alors que la mortalité reste faible ; c'est généralement le contraire lorsque l'infection est de nature bactérienne [23]. Lors de co-infections par les deux types d'agents, les taux de morbidité et de mortalité sont généralement élevés.

Ces troubles touchent principalement les jeunes bovins élevés en stabulation car les conditions d'ambiance tel que la température, l'hygrométrie et la ventilation sont propices au développement des infections respiratoires. L'étiopathogénie de ces maladies est complexe et multifactorielle avec un effet des facteurs liés aux agents pathogènes, à la physiologie du bovin et de son environnement de vie, ainsi que des modalités de conduite de l'élevage. Comme déjà indiqué précédemment, ces maladies sont provoquées par des virus, des bactéries et dans un certain nombre de cas, par la co-infection avec ces deux types d'agents. Les principaux virus responsables d'infections respiratoires chez les bovins sont : *virus respiratoire syncytial bovin (BRSV)*, le *virus parainfluenza 3 (BPIV3)*, le *l'herpès virus bovin de type 1 (BoHV-1)*, le *coronavirus bovin (BCoV)*, et de façon plus indirecte, le *virus de la diarrhée virale bovine (BVDV)*. Pour les bactéries, les agents pathogènes majoritaires sont les *Mycoplasmes* dont *Mycoplasma bovis* et surtout des pasteurelles comme *Mannheimia haemolytica*, *Pasteurella multocida* et *Histophilus somni*, ainsi que d'autres agents qui peuvent également être isolés des voies respiratoires comme des streptocoques des staphylocoques, ou des colibacilles.

L'anatomie pulmonaire des bovins possède plusieurs caractéristiques, le poumon gauche est constitué de 2 lobes (lobe cranial, lui-même divisé en 2 segments et le lobe caudal). Le poumon droit est divisé en 4 lobes (cranial, medium, caudal et accessoire) (**Figure 2**) [24].

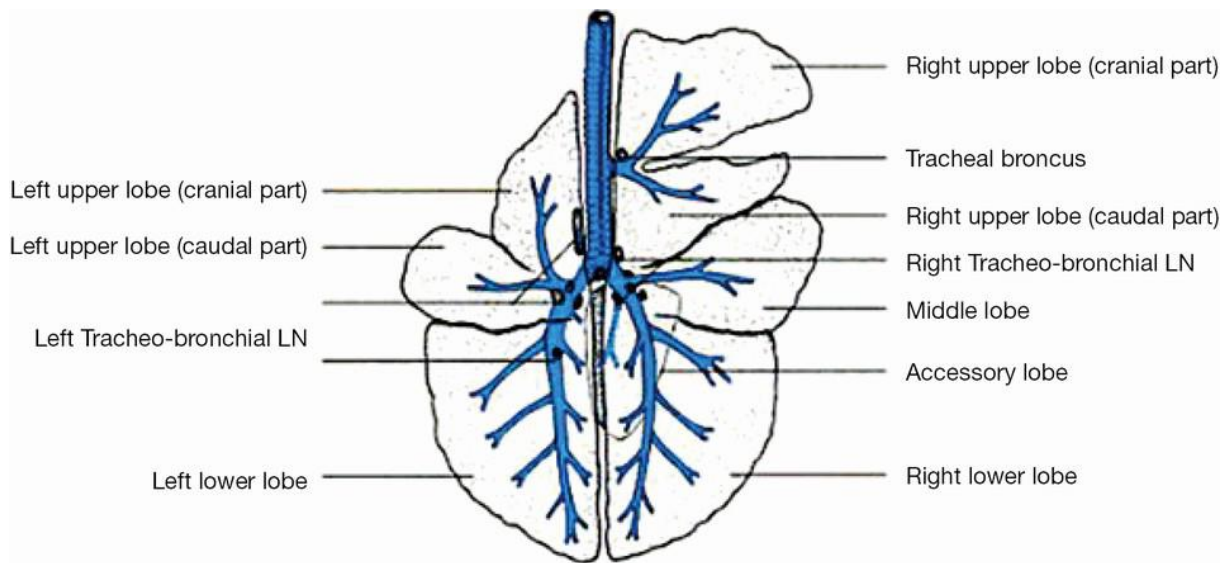


Figure 2 : Compartimentation pulmonaire d'un veau ([25])

Il existe de nombreux facteurs qui augmentent la prédisposition des jeunes bovins à ces troubles, en plus du rôle prépondérant des agents infectieux. Les conditions d'élevage et les modalités d'organisation de la filière d'élevage contribuent à l'apparition de ces troubles. Le transport est l'un des facteurs de risque reconnu comme important, avec pour effet la déshydratation des animaux ainsi qu'un ensemble de facteurs dits de stress qui surviennent concomitamment (sevrage, changement du lieu de vie et des rapports sociaux, ...). Les conditions météorologiques ont aussi un fort impact sur l'apparition de ces troubles.

Les caractéristiques physiologiques du jeune bovin sont également à considérer. En effet, *Bos taurus* est une espèce prédisposée aux infections respiratoires et à leurs conséquences morbides [26].

Les BPI sont caractérisées par des signes systémiques tels que la fièvre, l'abattement, ou la diminution de l'appétit et de l'ingestion, ainsi que par des signes locaux comme la toux, le jetage, les changements de la fréquence et de l'amplitude respiratoires, indiquant la présence d'une dyspnée.

1.1.1.2.2. Pathogénie des infections respiratoires

Les infections virales sont fréquemment responsables de BPI. Elles peuvent aussi prédisposer à l'apparition d'une infection bactérienne secondaire à cause des modifications de la physiologie des voies respiratoires et du tissu pulmonaire facilitant ainsi la colonisation bactérienne, mais aussi en inhibant la capacité de réponse immunitaire. Une infection par le *virus de la diarrhée virale bovine (BVDV)* par exemple provoque une diminution des fonctions des macrophages alvéolaires entraînant ainsi une immunodépression transitoire [27]. L'infection virale peut provoquer la perte de l'activité ciliaire et une destruction de cellules épithéliales de l'appareil respiratoire superficiel et profond, à l'origine d'une inflammation. Le *BRSV* provoque une diminution des capacités de phagocytose et d'opsonisation des macrophages alvéolaires. Le *BoHV-1* infecte les cellules épithéliales de l'appareil respiratoire proximal, dont la trachée avec *in fine* la destruction de celles-ci provoquant une altération de l'activité ciliaire qui facilite la colonisation de l'appareil respiratoire distal par *Mannheimia haemolytica* par exemple [28]. Les BPI provoquées par ce virus ne sont pas cliniquement différenciables de celles provoquées par d'autres agents pathogènes respiratoires viraux (*BRSV*, *BPIV3* et *BVDV*) ou bactériens (*M. haemolytica* et *P. multocida*) (**Table 2**).

1.1.1.2.2.1. Agents viraux

Le *virus respiratoire syncytial bovin (BRSV)* est l'un des pathogènes les plus virulents touchant la sphère respiratoire dans l'espèce bovine. L'infection par ce virus peut être asymptomatique, mais elle cause assez souvent une maladie respiratoire sévère. Il peut aussi participer à l'initiation des BPI seul ou en association avec les autres pathogènes viraux. Tout comme le *BPIV3*, il est présent dans la sphère respiratoire. Ce virus appartient à la famille des Paramyxoviridae. Les espèces sensibles sont les ruminants en général dont la chèvre, la brebis et surtout l'espèce bovine ; cette dernière semble être le réservoir. Il touche majoritairement les jeunes individus entre 3 et 12 mois ; il peut aussi dans une moindre mesure infecter les adultes [29], [30]. La contamination a lieu via les aérosols et les sécrétions issues du tractus respiratoire, principalement nasales. Il est responsable seul d'une pneumonie ou après co-infection avec des agents bactériens tels que *M. haemolytica* ou *P. multocida* [31]–[33]. Ce virus infecte les cellules épithéliales de la muqueuse nasale et des bronches, mais aussi les pneumocytes de type II et les macrophages alvéolaires (2). Au cours de l'infection, la disparition de l'activité ciliaire et la nécrose des cellules épithéliales sont observées. Les signes cliniques sont majoritairement

respiratoires avec de la toux et de la fièvre, et peuvent aboutir à une détresse respiratoire sévère dans certains cas.

Le *virus parainfluenza 3 bovin (BPIV3)* cause des infections communes dans l'espèce bovine. Elles sont souvent asymptomatiques, mais peuvent conduire à l'apparition de signes cliniques respiratoires. Ce virus est très souvent identifié dans les primo-infections lors de BPI car il altère le fonctionnement du système respiratoire, mais également celui du système immunitaire. Contrairement à d'autres virus (*BCoV*, *BVDV*) qui infectent aussi dans le tube digestif, il est exclusivement isolé de la sphère respiratoire. Il fait partie de la famille des Paramyxoviridae. Le réservoir et la cible principale sont l'espèce bovine et certaines espèces de ruminants tels que les ovins [34]. Ce virus cible majoritairement les cellules épithéliales. L'infection de ces cellules provoque une nécrose de l'épithélium, et une altération de l'activité ciliaire. Il en découle une diminution des mécanismes de défense de l'hôte. Lors d'une mono-infection, ce virus va entraîner une infection respiratoire avec des signes cliniques locaux et systémiques tels que la toux et la fièvre, respectivement. Lors d'une infection avec d'autres virus, la maladie est exacerbée avec une pneumonie plus sévère.

Le *virus de la diarrhée bovine (BVDV)* appartient à la famille des Flaviviridae et comprend 2 génotypes (*BVDV1* et *BVDV2*), et différents sous-types dont les majoritaires sont *BVDV1a*, *BVDV1b* et *BVDV2a*. [35], [36]. Historiquement très prévalent dans la population bovine, sa contribution à la survenue des BPI tend à diminuer grâce aux mesures de police sanitaire qui ont été mises en œuvre et qui ont conduit à une forte diminution de sa prévalence. Le réservoir majeur restant l'espèce bovine, et en particulier les bovins infectés de façon permanente, même si le virus est capable d'infecter d'autres espèces animales. Il se transmet par voie aérienne essentiellement, par contact direct entre les animaux. Chez l'animal infecté, le virus est présent dans l'appareil respiratoire supérieur. Il est excrété avec les sécrétions nasales après infection de l'épithélium respiratoire ; il est également présent dans l'appareil digestif. La sévérité de l'infection virale varie en fonction de la souche virale, du statut immunitaire de l'animal et de l'apparition d'infections secondaires. L'infection par le *BVDV* provoque une immunodépression, en lien avec une leucopénie. Cette modification de la compétence immunitaire contribue à augmenter la sensibilité de l'animal vis à vis d'autres agents pathogènes provoquant l'apparition de BPI plus sévères [37]. La leucopénie est plus marquée lors d'une infection par le *BVDV2* par rapport au *BVDV1* [38]–[42]. Alors que l'infection par le *BVDV* est fréquemment asymptomatique, la co-infection avec le *BRSV* ou le *BoHV-1* peut conduire à des tableaux cliniques sévères. La présence d'anticorps spécifiques du *BVDV* réduit

la morbidité en cas d'infection indiquant qu'ils sont protecteurs vis-à-vis de l'expression des signes cliniques associés à l'infection [43].

Le *coronavirus bovin (BCoV)* a été identifié pour la première fois lors d'épisodes de diarrhée néonatale chez des veaux, puis dans des épisodes diarrhéiques chez des vaches laitières adultes, indiquant une prépondérance digestive. Néanmoins, il est également capable de provoquer des infections respiratoires symptomatiques et participer activement au développement des BPI. C'est un virus enveloppé appartenant à la famille des Coronaviridae. Les virus de cette famille sont connus pour causer des infections respiratoires dans d'autres espèces comme l'Homme, comme dans le cas de la Covid-19. Le *BCoV* est retrouvé dans toutes les régions du monde. Il n'y a qu'un seul sérotype mais il existe un certain degré de variabilité antigénique et l'existence de différentes souches [29]. Le premier lien entre coronavirus et BPI fut l'isolation du virus dans des écouvillons nasaux d'animaux malades.

L'*herpès virus bovin (BoHV)* fut le premier virus identifié parmi les agents infectant le système respiratoire supérieur des bovins [44]. Actuellement, ce virus a été pratiquement éradiqué du territoire et la maladie n'existe quasiment plus. L'infection est parfois détectée dans les ateliers d'engraissement de jeunes bovins. Il touche non seulement la sphère respiratoire, mais aussi le tractus génital et possiblement le système nerveux. Ce virus appartient à la famille des Alphaherpesviridae avec divers sous types, le *BoHV-1* est associé à des pathologies respiratoires, tout comme le *BoHV-5* alors que le *BoHV-4*, lui est impliqué dans des infections du tractus génital [28]. Ces sous-types partagent certaines propriétés. Il est présent dans toutes les régions du monde et touche les différentes espèces bovines [29], mais aussi d'autres espèces de ruminants. Ce virus est retrouvé chez les animaux âgés souvent de plus de 6 mois après la disparition de l'immunité maternelle, qui ne prévient cependant pas l'infection. Celui-ci peut établir une latence sans symptômes après infection, ou rester sous forme active et donc être excrété et détecté dans les sécrétions nasales. Cette réactivation peut entraîner un effet stimulant sur la réponse, dont la production d'anticorps, et donc réduire à nouveau l'excrétion [28]. La sévérité des infections respiratoires par ce virus peut être modérée ou sévère, avec parfois jusqu'à 100% de morbidité. Cette sévérité est également due aux pathogènes qui infectent l'animal de façon concomitante. La maladie respiratoire due à ce virus cause des signes cliniques respiratoires (fièvre, rhinite) [45] [30].

1.1.1.2.2.2. Agents bactériens

Mannheimia haemolytica (*Mh*) est une bactérie à Gram négatif très fréquemment isolée lors de BPI chez les bovins. Elle est une cause majeure de broncho-pneumonie avec une prévalence supérieure à 65% [46]. Elle a un fort pouvoir pathogène respiratoire dans cette espèce. Celle-ci est parfois retrouvée au sein du microbiote respiratoire chez des animaux sains, indiquant que le portage oropharyngé est possible chez un animal n'exprimant pas de symptôme [47]. Il existe 12 sérotypes (A1 à A12) et les sérotypes A1 et A2 sont prépondérants dans l'appareil respiratoire des bovins. Cette bactérie est un agent primaire de BPI. Le sérotype A1 est le plus souvent isolé du tissu pulmonaire, mais d'autres sont également retrouvés comme A6 en fonction des régions du monde.

Cette bactérie possède de nombreux facteurs de virulence associés à son caractère pathogène ; ceux-ci sont notamment utilisés pour l'invasion et l'adhérence aux cellules épithéliales respiratoires. Certaines protéines exprimées à sa surface sont capables de protéger la bactérie vis à vis des cellules du système immunitaire, en empêchant la phagocytose.

Pasteurella multocida comprend 5 sérogroupes, (de A à F) et 16 sérotypes (de 1 à 16). Le plus fréquemment retrouvé lors de BPI chez les bovins est le type A:3 [48], dans les infections respiratoires chez le veau. Néanmoins, sa détection est souvent associée à des facteurs prédisposants tels que des facteurs dits de stress (transport, changement alimentaire, conditions météorologiques), et la présence d'autres agents infectieux. Comme indiqué pour *M. haemolytica* précédemment, cette espèce est retrouvée dans l'appareil respiratoire superficiel d'animaux sains avec une prévalence comprise entre 20 et 60% [49], [50].

Histophilus somni est une bactérie à Gram négatif parfois commensale du tractus respiratoire [51], [52]. La physiopathologie des infections respiratoires par cette bactérie est similaire à celle décrite pour *M. haemolytica* car les facteurs de virulence sont similaires, avec néanmoins quelques particularités comme la production d'un exopolysaccharide supplémentaire. Cette bactérie est responsable d'infections pulmonaires seule ou en association avec d'autres agents pathogènes.

Mycoplasma bovis peut provoquer des pneumonies qui peuvent être aiguës, mais elles sont le plus souvent chroniques à la suite de l'infection par d'autres agents respiratoires [53], [54]. La prévalence de cette bactérie est variable selon les élevages et les systèmes ; néanmoins l'augmentation de sa prévalence est liée à la présence de facteurs de stress et surtout reliée aux mouvements d'animaux (achat, regroupement). La bactérie peut infecter d'autres organes comme la glande mammaire ou les articulations, indiquant la possibilité d'une bactériémie [53].

	Prévalence	Pouvoir pathogène
Bactéries		
<i>M.haemolytica</i>	+++	+++
<i>P.multocida</i>	++	++
<i>H.somni</i>	++	++
<i>M.bovis</i>	+++	+++
Virus		
<i>BRSV</i>	+++	+++
<i>BPI3</i>	+++	++
<i>BVDV</i>	++	++
<i>BoHV</i>	++	+++
<i>BCoV</i>	++	++

Légendes : +++ Forte ++ Modérée + Faible

Table 2 : Différents types d'agents intervenant dans les BPI

1.1.1.3. Réponse immunitaire

1.1.1.3.1. Générale

La défense vis-à-vis des pathogènes comportent plusieurs types de mécanismes. Le premier correspond aux barrières physico-chimiques tels que la peau ou les muqueuses. Dans le cadre des mammites et des broncho-pneumonies infectieuses, ce sont les muqueuses qui jouent le rôle de barrière. Elles sont constituées de cellules épithéliales qui protègent les organes des pathogènes provenant de l'extérieur. Elles constituent une barrière à la fois mécanique, chimique et même biologique. En effet, via la présence de cils à la surface de ces cellules, elles sont capables d'éliminer des pathogènes captés par la production de mucus. Ces cellules sont aussi capables de produire des molécules permettant la destruction des pathogènes tels que le lysozyme. Au sein des muqueuses, la présence d'un microbiote empêche la colonisation par des pathogènes étrangers. Néanmoins tous ces mécanismes ne sont pas toujours suffisants, le pathogène va infecter l'organisme et induire l'activation du système immunitaire innée.

De façon générale, l'activation de la réponse immunitaire passe par une étape de reconnaissance de la bactérie via des récepteurs spécifiques que sont les *Pattern recognition receptor* (ou PRR). Ceux-ci sont divisés en quatre grandes familles, les TLR (*Toll-like récepteurs*), les RLR

(Récepteurs de type *RIG-1*), les NLR (Récepteurs de type *Nod*) et les CLR (Récepteurs de lectine de type *C*). Tous ces récepteurs ont pour rôle de reconnaître le pathogène via des structures spécifiques des agents pathogènes nommées *Pathogen-Associated Molecular Patterns* (ou PAMP). Ces derniers sont présents majoritairement à la surface des cellules, hormis quelques récepteurs qui sont localisés dans leur cytoplasme. Cette reconnaissance va induire un ensemble de cascade de signalisation entraînant une production de molécules que sont les cytokines et chimiokines permettant le recrutement de cellules de l'immunité innée tels que les neutrophiles ou encore les macrophages. Toute cette production de molécules et ce recrutement cellulaire sont responsables du phénomène d'inflammation. Ce mécanisme intervient moins de 24 heures après l'infection. Si l'ensemble de ces réponses ne suffisent pas à la destruction du pathogène, l'immunité adaptative va prendre le relais via l'activation des cellules lymphocytaires T ou B.

Tous ces mécanismes sont communs aux infections mais il existe des particularités en fonction du type d'infection ou de sa localisation.

1.1.1.3.2. Vis-à-vis des mammites

La première défense vis-à-vis des mammites sont les barrières physiques et biologiques de l'extrémité du trayon, car elle est la voie utilisée pour l'entrée des bactéries dans la mamelle. Durant la période sèche, le trayon est obstrué par un bouchon de kératine permettant une protection physique contre l'invasion bactérienne. La kératine a des effets anti-bactériens via une altération de l'osmose des bactéries. Cette défense est modulée par de nombreux facteurs tels que la taille du canal, la quantité de kératine, ou encore des facteurs physico-chimiques qui ont des propriétés anti-bactériennes [55].

Le sphincter présent à cet endroit se ferme entre les traites environ 20 minutes à 2 heures après la fin de l'extraction du lait. Cette période est très favorable à l'invasion des bactéries. En cas de franchissement de la barrière, une seconde ligne de défense est mobilisée, qui correspond à la réponse immunitaire de l'hôte. La qualité de cette réponse et la sévérité de la mammite sont étroitement corrélées. Elle varie en fonction de facteurs génétiques de l'animal mais aussi de l'état immunitaire de l'animal (stress, concomitant, vaccination, âge, ...) au moment de l'infection, [56], du type de bactérie infectante [57] et de l'expression de facteurs de virulence. Dans le cas particulier des mammites, par exemple, TLR2 reconnaît le peptidoglycane présent chez certaines bactéries alors que TLR4 reconnaît le LPS, molécule retrouvée chez la bactérie

E. coli notamment. Cette reconnaissance est à l'origine d'une cascade de signalisation entraînant la production de molécules permettant l'activation cellulaire [58].

La réponse immunitaire comporte à la fois des acteurs solubles et d'autres de nature cellulaire. La réponse humorale correspond au système du Complément, à diverses molécules antibactériennes (lactoferrine, transferrine, lysozyme), à des enzymes (lactoperoxydase, xanthine oxydase), et à la réponse anticorps. Le complément est un système activé avec plusieurs effets comme la lyse bactérienne et le recrutement cellulaire [59]. Dans le cas des mammites, la production de la molécule C5a est augmentée en cas d'infection par *E. coli* et *S. uberis* à l'inverse de l'infection par *S. aureus* ; elle participe au recrutement des neutrophiles en plus des chimiokines qui sont sécrétées [60]–[63].

La lactoferrine est une molécule connue pour avoir un effet chélateur qui lui confère une activité bactériostatique. Elle est produite par les cellules épithéliales mammaires (ou MEC) en majorité, mais elle est aussi présente dans les granules des neutrophiles. Pour leur multiplication, les bactéries ont besoin d'une grande quantité de fer ; elles sont donc très sensibles aux effets des molécules captant ce métal [64]. Toutes les bactéries causant des mammites ne sont cependant pas sensibles. *E. coli* fait partie des plus sensibles, alors que les bactéries du genre *Streptococcus* sont résistantes car elles vont lier la lactoferrine à leur surface leur permettant ainsi de capter directement le fer via la lactoferrine [65], [66]. L'effet bactériostatique de cette molécule est temporaire car il permet uniquement de stopper la multiplication bactérienne mais ne permet pas la destruction des bactéries. La lactoferrine a un effet bactéricide vis-à-vis d'*E. coli* en concentrant le fer [67].

La transferrine est une molécule similaire à la lactoferrine et elle est présente dans la circulation sanguine. Elle n'est pas produite au sein de la glande mammaire et provient donc du sang via le phénomène de transcytose [68]. Il a été montré lors d'une mammite expérimentale à *E. coli* que la concentration en transferrine était augmentée avant celle de lactoferrine [69]. Ces deux dernières molécules peuvent entraîner des dommages de la paroi des bactéries Gram négatives entraînant une perméabilité [68].

Le lysozyme est une protéine antibactérienne qui clive les peptidoglycanes présents à la surface des bactéries. L'action isolée de cette molécule n'est généralement pas suffisante. Néanmoins elle peut agir en synergie avec les autres molécules anti-infectieuses décrites précédemment. Certaines enzymes ont des activités antimicrobiennes ; c'est le cas pour l'enzyme lactoperoxydase qui permet l'inhibition ou la destruction des pathogènes en présence de

thiocyanate et d'hydrogène dans le milieu. La xanthine oxidase permet la formation d'oxide de nitrile à partir de nitrite inorganique et a des fonctions antibactériennes [70].

La sérum amyloïde A (ou SAA) est une protéine présente dans le lait en cas d'infection mammaire ; elle peut aussi être retrouvée dans le sérum. Cette molécule peut être utilisée comme marqueur d'une infection mammaire [71], [72]. En plus de produire les molécules anti-microbiennes décrites précédemment, certaines cellules tels que les granulocytes, les monocytes/macrophages et les cellules épithéliales sécrètent des peptides aux effets antimicrobiens. Les peptides les plus connus dans l'espèce bovine sont les défensines produites par les neutrophiles tels que la β -défensine (BNBD) [73]. L'effet des défensines sur *E. coli* et *S. aureus* a été démontré *in vitro* [74]. Néanmoins leur activité anti-bactérienne n'est pas systématiquement présente *in vivo* [75] à cause du contexte local plus ou moins favorable à leur activité. Ils possèdent aussi des propriétés de chimiotactisme vis-à-vis de plusieurs types de cellules immunitaires (monocytes, cellules dendritiques) et peuvent également permettre la production de cytokines par les cellules épithéliales [76], [77].

La réponse cellulaire repose sur divers types de cellules immunitaires que sont les macrophages, les neutrophiles, les cellules Natural Killer (NK) et les cellules dendritiques mais aussi les cellules épithéliales. Le type cellulaire dominant est le macrophage, correspondant à des cellules capables de phagocytose des pathogènes bactériens permettant ainsi leur destruction [78]–[80]. L'activité phagocytaire dans le lait est moins efficace que celles des neutrophiles. Cependant, ces cellules peuvent aussi avoir une fonction de sentinelle permettant la détection de la présence des pathogènes [81]. Les macrophages activés peuvent produire la *nitric oxide synthase* (iNOS) à l'origine de la formation de réactifs dérivés de l'oxygène ayant des propriétés anti-microbiennes [82]. La production de ces dérivés est faible dans la mamelle saine mais elle augmente dans le lait en cas d'infection par *E. coli* ou *S. aureus* [83]–[85]. L'activation des macrophages est souvent liée à des facteurs solubles comme les cytokines et les chimiokines, notamment IFN- γ et TNF- α . Les cellules du lait ont la capacité de signaler la présence de bactéries via des médiateurs qui sont pour le moment mal définis. Les macrophages du lait ne possèdent pas la même activité que les monocytes issus du sang, et avec notamment une production d'IL-1 plus faible [86].

L'activité phagocytaire est surtout portée par les neutrophiles, mais leur concentration et leur viabilité dans le lait d'une glande saine sont faibles, ne permettant pas cette activité. Néanmoins, une étude a montré une corrélation entre la présence de ces cellules dans le lait et la réduction du risque d'infection [87].

Les cellules NK sont des lymphocytes ayant une activité cytotoxique indépendante du complexe majeur d'histocompatibilité (CMH), mais dépendante des anticorps via la cytotoxicité cellulaire dépendante des anticorps (ou ADCC). Les cellules NK sont cruciales dans le cas d'infection par des bactéries intra-cellulaires, ce qui est rarement le cas des infections mammaires, mais elles sont aussi capables de participer à la destruction des pathogènes via la production de protéines similaires à la Saposine, qui ont des activités anti-bactériennes. Ces cellules présentes dans le parenchyme mammaire et le lait, ne sont pas les seules à intervenir lors d'infection ; d'autres infiltrent la glande et ont des activités anti-bactériennes [88]. C'est le cas de lymphocytes T de type CD4+ et CD8+ au niveau des alvéoles mais cela peut aussi être des lymphocytes B, ou des cellules dendritiques, dont le rôle dans l'immunité mammaire est mal défini.

Les MEC possèdent elles-aussi des récepteurs permettant la reconnaissance des principaux pathogènes causant des mammites. Elles peuvent donc participer à l'élaboration des défenses en produisant des molécules de défense comme les cytokines, chimiokines ou encore des peptides anti-microbiens. Elles jouent possiblement un rôle dans le recrutement des neutrophiles. Il a été montré *in vitro* que ces cellules produisaient la chimiokine IL-8 en cas de stimulation avec du LPS [89]. Elles produisent aussi d'autres chimiokines permettant le recrutement cellulaire, notamment celui des neutrophiles. En plus de la production de chimoattracteurs, elles synthétisent des cytokines jouant un rôle dans l'inflammation (comme IL-6, IL-1, et TNF α) lors d'une stimulation avec du LPS. Le TNF α jouerait un rôle principal dans la réponse à une infection mammaire par *E. coli* [90]. L'ensemble des molécules produites par les MEC, mais aussi probablement par les macrophages, qu'ils soient présents dans le tissu ou dans le lait, vont permettre le recrutement des neutrophiles depuis le sang vers le parenchyme, et ultérieurement le lait. Ceux-ci vont phagocyter, détruire les bactéries pathogènes et produire des molécules anti-bactériennes (lactoferrine, peptides anti-microbiens) permettant ainsi la destruction des bactéries selon plusieurs modalités. Le recrutement des neutrophiles est primordial et déterminera l'issue plus ou moins favorable de l'infection. Il a été montré que le délai entre l'apparition de l'infection bactérienne et le recrutement des neutrophiles était d'environ une heure, la quantité de bactéries est ainsi nettement supérieure au nombre de neutrophiles (8 fois plus grand) [3]. Le recrutement cellulaire varie en fonction de la bactérie responsable de l'infection. Dans le cas d'une mammite causé par *E. coli*, l'activation forte de la réponse inflammatoire avec une production de cytokines pro-inflammatoires (IL-1 et TNF α) dans le lait, entraîne un fort recrutement de neutrophiles entre 3 et 12 heures post-infection. Cette phase correspond au pic aigu d'une mammite à *E. coli* avec

une destruction des bactéries via l'activation des neutrophiles [63]. En revanche pour les mammites provoquées par *S. aureus*, celle-ci sont généralement moins sévères, souvent subcliniques. Les signes de l'infection n'apparaissent que 24 à 48 heures après l'infection avec une faible production initiale de cytokines inflammatoires ou de chimiokines comme l'IL-8 et le TNF α [60], [62]. Ces caractéristiques favorisent probablement le passage à la chronicité de cette infection initialement moins sévère. Le recrutement cellulaire se poursuit et entraîne *in fine* des dommages du tissu mammaire [91]. Lors de mammites causées par d'autres bactéries Gram-positives tels que les streptocoques, la réponse est très dépendante de l'animal [92]. Lors de ces infections, les réponses induites par la réponse adaptative tels que l'activation des neutrophiles ou la production de peptides antimicrobiens ont un effet majeur.

La réponse adaptative permise par les lymphocytes a aussi un rôle dans les infections mammaires, même si elle est globalement moins bien connue que les autres aspects de la réponse immunitaire mammaire, car leur rôle exact a été peu investigué. La présence de lymphocytes de type mémoire au sein du tissu mammaire a été établie [93]. La réponse de type 3 correspondant aux lymphocytes de type Th17 jouerait un rôle dans les mammites. En effet, via la production d'IL-17, ces cellules favorisent le recrutement des neutrophiles. Il a été établi que lors d'infections par *E. coli* et *S. aureus*, la réponse médiée par les lymphocytes CD4⁺ Th17 permet le contrôle de l'infection et qu'elle est suivie d'une activation de la réponse anti-inflammatoire médiée par les lymphocytes T régulateurs produisant de l'IL-10 [6], [94]. Cette réponse a un effet bénéfique dans le cas d'une mammité à *E. coli* et un effet modéré dans le cas d'une mammité à *S. aureus*. La production d'IL-17 a aussi un effet sur les cellules épithéliales bovines en induisant la production de molécules antimicrobiennes directement par celles-ci [95].

Les mécanismes de la vaccination fondée sur la mobilisation de la réponse adaptative sont encore mal définis ce qui nuit au développement de vaccins efficaces. De plus, les études de la réponse associée à la production d'IL-17 sont encore peu répandues.

1.1.1.3.3. Vis-à-vis des infections respiratoires

La défense des animaux vis-à-vis des agents pathogènes respiratoires est permise par une protection par des mécanismes physiques, des interactions microbiologiques avec le microbiote respiratoire, et par la réponse immunitaire qu'elle soit de nature humorale ou cellulaire.

La première ligne de défense est portée par l'épithélium des voies respiratoires, au sein duquel certaines cellules dotées d'une activité ciliaire permettent l'élimination des particules et de ce

fait de certains pathogènes. Ainsi, 90% des bactéries *M. haemolytica* sont éliminées via cette activité [96]. Néanmoins, les pathogènes viraux tels que les virus respiratoires *BRSV* et *BoHV-1* provoquent une altération de cette activité par la destruction de ces cellules [97], [98]. Cette propriété serait une des raisons pour laquelle les infections virales prédisposent aux infections bactériennes.

Le système respiratoire superficiel est colonisé par des bactéries dites commensales. Cette colonisation masque les récepteurs et utilise les nutriments empêchant l'adhérence et la multiplication des bactéries pathogènes par compétition.

Au début de l'infection, les récepteurs présentés précédemment ont une activité augmentée, comme cela a été montré *in vitro* pour *M. haemolytica* et *BoHV-1* avec une expression plus forte de TLR-2 [99]. *In vivo*, une infection par le virus *BoHV-1* entraîne une expression augmentée des TLR-3, -7, -8 et -9 dans les régions superficielle et profonde du système respiratoire [100]. L'activation de ces récepteurs va permettre la mise en place de la réponse immunitaire innée, puis adaptative, via la sécrétion de molécules entraînant le recrutement de cellules.

Les cellules épithéliales du système respiratoire participent à la reconnaissance du pathogène via l'expression de PRR, à la défense via la production de molécules antimicrobiennes tels que la lactoferrine et divers peptides anti-microbiens (dont TAP pour Tracheal Antimicrobial Peptide, LAP pour Lingual Antimicrobial Peptide, et BMAP-28 pour Bovine Myeloid Antimicrobial Peptide. Le peptide TAP a une activité bactéricide sur *M. haemolytica*, *H. somni* ainsi que sur *P. multocida* [101], alors que le peptide BMAP-28 présente une action uniquement sur cette dernière bactérie [102]. Tout comme l'activité ciliaire, la production de peptides antimicrobiens peut être réduite en présence de pathogènes viraux comme le virus *BVDV* qui empêche la production de lactoferrine et de LAP par les cellules épithéliales de la trachée [103]. Les cellules épithéliales ont donc un rôle dans la défense vis-à-vis des pathogènes bactériens mais c'est aussi le cas pour les pathogènes viraux avec une induction de la réponse anti-virale conduisant à la production d'interférons de type I qui est plus forte à la suite d'une infection par *BoHV-1* [104]. Certains pathogènes viraux comme *BRSV*, *BVDV* et *BPI3* possèdent des mécanismes inhibant cette réponse notamment dans le cadre de co-infections [105], [106], [107], [108]. L'invasion des cellules épithéliales par des agents bactériens entraînent une réponse inflammatoire et un recrutement de cellules immunitaires entraînant la production d'IL-6, IL-8 et de TNF- α . Cette production est exacerbée en cas de co-infection, notamment entre *M. haemolytica* et *BoHV-1* [99], [109] et entre *BRSV* et *P. multocida* [110]. Les cellules

épithéliales peuvent aussi jouer un rôle dans la réponse anti-inflammatoire par la production d'Annexine A1 et A2 qui sont des protéines ayant une fonction anti-inflammatoire [111].

La seconde ligne de défense correspond à la réponse cellulaire de l'immunité innée tout d'abord correspondant à l'activation de cellules telles que les monocytes/macrophages, les neutrophiles et les cellules dendritiques. Les premières cellules recrutées sont les neutrophiles issus du sang en réponse à la production d'IL-8 et de cytokines pro-inflammatoires. Ces cellules sont à double tranchant ; elles possèdent une forte activité phagocytaire permettant la destruction des pathogènes mais la production de réactifs oxygénés peut aussi causer des lésions tissulaires secondaires, comme cela a été montré dans l'infection par *M. haemolytica* [112]. Un autre mécanisme de défense permis par les neutrophiles est la NETose, phénomène durant lequel les neutrophiles libèrent une partie de l'ADN (pour former des filets ou *nets*) dans le milieu afin de piéger les pathogènes. Certains pathogènes à l'origine des BPI comme *M. haemolytica*, *M. bovis* et *H. somni* ont la capacité de déclencher la nétose des neutrophiles. Il a ainsi été montré *in vitro* que ce mécanisme pouvait tuer *M. haemolytica* et *H. somni* [113]–[115]. Ce mécanisme a un fort effet sur le pathogène car dans le cas d'une administration de DNase permettant la destruction des nets, il y a une réduction des signes cliniques, et une diminution de l'obstruction des voies respiratoires lors d'une infection par *BRSV* [116].

Les macrophages peuvent à la fois être des cellules permettant l'éducation des lymphocytes mais aussi réaliser l'élimination des pathogènes ou des cellules mortes destinées à les éliminer, par la phagocytose, mais aussi participer à la production de molécules actives permettant la destruction directe du pathogène. Néanmoins l'activité de ces cellules peut être diminuée par les pathogènes comme le *BVDV*. Ce pathogène a des effets immunodépresseurs sur les macrophages alvéolaires qui sont favorables à l'apparition d'une BPI. Il a été montré que ce virus était capable de supprimer la production de cytokines pro-inflammatoires, de diminuer l'activité phagocytaire et la reconnaissance via une diminution de l'expression des TLR [117]. D'autres pathogènes ont des effets immunomodulateurs sur ce type cellulaire comme le *BRSV* entraînant une inhibition de la phagocytose [118]. Une co-infection par les virus *BVDV* et *BoHV-1* entraîne le recrutement d'un nombre plus grand de macrophages que l'infection par un seul de ces virus, mais la production de cytokines pro-inflammatoires et de radicaux oxygénés est réduite [119].

Les cellules dendritiques permettent l'activation des lymphocytes ; c'est une étape essentielle de la réponse adaptative. La réponse adaptative comprend 2 modes d'action, avec la réponse

humorale, et la production d'anticorps surtout, et la réponse cellulaire via l'activation des lymphocytes T auxiliaires (CD4+) et cytotoxiques (CD8+).

Les lymphocytes B sont capables de reconnaître de façon directe des déterminants de surface des pathogènes, conduisant à leur activation et à la différenciation en cellules effectrices productrices d'anticorps. Ces anticorps jouent un rôle de neutralisation, d'activation du Complément, dans la phagocytose médiée par les récepteurs Fc, et la cytotoxicité dépendante des anticorps (ADCC). Dans le cadre d'une infection par *M. haemolytica*, les anticorps vis-à-vis des protéines de membrane permettent l'activation du Complément et la destruction de la bactérie [120]. Après vaccination contre *H. somni*, les taux d'IgG2 ainsi produits sont inversement corrélés à la sévérité de la maladie lors d'une infection expérimentale [121]. Les anticorps développés lors d'une infection expérimentale par *M. bovis* sont de type IgG1 et ont une activité vis-à-vis du pathogène qui est plus faible que celle des IgG2 [122]. Les glycoprotéines de surface des pathogènes viraux sont aussi la cible de certains anticorps, avec des réponses variables selon la protéine comme cela a été montré dans le cas du *BoHV-1* [123]. Les lymphocytes T $\gamma\delta$ sont des cellules intervenant à la fois dans la réponse innée et dans la réponse adaptative. Elles sont retrouvées en grande proportion, jusqu'à 60% de l'ensemble des lymphocytes, dans le sang des jeunes bovins [124], mais elles sont aussi présentes en grandes quantités dans les régions muqueuses. Elles ont un rôle de défense notamment par la production d'IL-17, cytokine permettant le recrutement des neutrophiles. Cette production cytokinique est augmentée dans un modèle d'infection *in vitro* avec *BRSV* et *M. haemolytica* [125]. Néanmoins cet accroissement n'est pas observé dans le cadre d'une co-infection avec *BRSV* et *P. multocida* [126]. Lors d'une infection par *M. bovis*, ces cellules sont la source principale d'IFN γ [127]. L'activation de celles-ci est possible lors d'une vaccination avec un virus *BoHV-1* atténué [128], mais aussi dans la réponse après une épreuve infectieuse avec *BVDV-1* [129]. Néanmoins, une déplétion de ces cellules n'a pas d'impact significatif sur la réponse clinique des veaux dans le cadre d'une infection avec *BRSV* [130].

Le rôle exact des lymphocytes T est encore peu documenté. Cependant, une infection avec *P. multocida* entraîne une augmentation du nombre de lymphocytes T CD8+ activés dans le lavage broncho-alvéolaire [126]. Lors d'une infection de cellules sanguines par *M. bovis in vitro*, les lymphocytes T CD4+ et CD8+ sont activés et produisent de l'IFN γ et de l'IL-4 [127]. Les lymphocytes T CD4+ sont essentiels pour la protection vis-à-vis d'une infection par *BoHV-1* par la reconnaissance des glycoprotéines de surface [131]. Les deux types de lymphocytes sont aussi très importants dans une réponse vis-à-vis de *BRSV* avec une infiltration des lymphocytes

T CD8+ et une augmentation de l'activité cytotoxique 7 à 10 jours après l'infection [132]. Ils sont aussi identifiés dans le lavage broncho-alvéolaire en réponse à une infection par le virus BVDV [129].

1.1.2. Les mesures de maîtrise des maladies des bovins

1.1.2.1. Les méthodes d'élevages

1.1.2.1.1. Mesures générales

Lors du développement de connaissances sur les pathogènes, il a été montré que l'hygiène était un paramètre crucial pour lutter contre les infections. En effet, l'absence de conditions d'hygiène adaptées, les coûts de production augmentent à cause des dépenses liées à la santé des animaux, à la morbidité ou encore à la mortalité. En parallèle, une augmentation du temps de travail est nécessaire pour le soin des animaux. Il a donc été mis en place des mesures de gestion permettant la prévention des infections. La première consiste à diminuer le risque d'introduction et de circulation d'un pathogène au sein de l'élevage, qui est un ensemble de mesures que l'on désigne par le terme générique de biosécurité. Cela passe par le nettoyage des locaux et du lieu de vie des animaux mais aussi du matériel utilisé. Une partie des contaminations dans les élevages est liée à l'homme, un nettoyage des mains lors de la traite par exemple est indispensable, mais cette modalité de contamination est mineure par rapport aux sources animales. La seconde consiste à respecter la conduite de l'élevage afin de réduire les facteurs de risque d'apparition des maladies. L'ensemble de ces mesures sont applicables dans tous les contextes infectieux mais il en existe d'autres qui diffèrent en fonction de l'infection (mammites ou infections respiratoires).

1.1.2.1.2. Prévention des mammites

Les mammites sont multifactorielles avec des facteurs de risque pouvant être corrigés totalement ou partiellement en fonction du type de mammites (mode de transmission contagieux ou d'origine environnementale). Il a été montré qu'une faible hygiène du lieu de vie était associée à une augmentation de la quantité de cellules somatiques [133]. Le réservoir majeur de *S. aureus* correspond aux glandes infectées des vaches en lactation, mais cette bactérie peut infecter d'autres régions anatomiques de l'animal tels que la peau ou encore les cavités nasales [134]–[136]. Les animaux chez lesquels la bactérie est présente sur la peau ont 3,3 fois plus de chance de développer une mammite provoquée par cette bactérie [135]. Concernant *Streptococcus agalactiae*, ce fut l'une des premières bactéries dont il a été montré qu'elle se

transmettait via le matériel de traite [137]. La désinfection des équipements de la traite permet une baisse du risque de développement d'une mammite dont le mode de transmission est principalement contagieux, comme c'est aussi le cas notamment lors d'infection par *S. uberis*, et surtout *S. aureus* [138] [139].

La maîtrise de la nutrition est associée à une réduction de la prévalence des mammites [140]. La prévention des troubles métaboliques, comme la cétose, par l'utilisation d'ionophores qui sont des antibiotiques utilisés comme facteurs de croissance et de régulation du microbiote ruminal, permet de réduire le risque de mammites cliniques [141]–[143]. La supplémentation en sélénium et en vitamine E permet l'amélioration de l'activité phagocytaire, et elle est donc associée à une diminution du risque [144], [145]. Le cuivre a une activité anti-oxydante corrélée à une diminution des mammites cliniques alors que le zinc a un effet dans le maintien de l'intégrité des barrières épithéliales [140].

Comme vu dans la partie dédiée à la vaccination, elle peut réduire la prévalence des mammites mais elle est encore d'une efficacité modeste, notamment vis-à-vis des infections par *E. coli* et *S. aureus*. L'utilisation d'antibiotiques doit être réduite, notamment dans l'objectif de prévenir l'antibiorésistance.

Il existe des facteurs propres à l'animal tel que l'âge qui modulent l'incidence des mammites. En effet, il a été montré que les animaux âgés avaient un nombre de cellules somatiques augmenté en début de lactation [133], et un nombre de mammites dues à des pathogènes environnementaux également plus fréquentes [138]. La perte de kératine entre les périodes de lactation a été observée et contribue au risque accru de développement des mammites [146]. La génétique de l'animal a aussi un impact sur la réceptivité et la sensibilité vis-à-vis des mammites. Les vaches produisant beaucoup de lait ont un risque augmenté de développer une mammite [147], [148]. La forme des trayons de l'animal a aussi un impact car il a été montré que les animaux avec des tétines longues étaient plus fréquemment atteints de mammites [149], [150]. La période avant la mise bas, la première lactation sont des moments où le développement des mammites est plus fréquent en lien avec une augmentation du stress oxydatif, et une diminution probable des mécanismes de défense liés à l'oxygène [151]. Il a aussi été observé une augmentation du nombre de cellules somatiques et notamment des neutrophiles juste après la mise bas [152], [153].

L'incidence des mammites est liée aux conditions climatiques avec une augmentation durant la saison humide [154]. De plus, une étude a démontré une augmentation des infections de bactéries du type *E. coli* durant les périodes froides (température inférieure à 21°C) [155].

1.1.2.1.3. Dans le cadre des bronchopneumonies infectieuses

Les mesures désignées sous le terme générique de biosécurité sont un ensemble de pratiques de gestion permettant de réduire l'introduction et la transmission de agents pathogènes infectieux au sein d'un effectif. Elles reposent sur l'amélioration des conditions de vie des animaux ainsi que sur la prévention vis-à-vis des maladies respiratoires. Ces améliorations passent par une stratégie de vaccination, la biosécurité des échanges par la diminution des contacts entre animaux et de l'introduction d'agents pathogènes et notamment du *BVDV*, l'amélioration des conditions de vie comme la qualité de l'air des locaux d'hébergement.

La vaccination est aussi une méthode permettant la prévention des troubles respiratoires. Néanmoins, le développement de vaccins efficaces est difficile en raison de la complexité des mécanismes des maladies respiratoires, et du parcours de vie des animaux depuis les zones d'élevage plutôt extensif vers des ateliers d'engraissements spécialisés (voir partie vaccination). La biosécurité des échanges et de l'hébergement passe par la diminution de la transmission entre les animaux. En effet, la transmission de ces pathogènes se fait majoritairement par le contact entre les individus car contrairement aux pathogènes digestifs, la survie dans l'environnement est limitée [156]. La qualité de l'air et l'ambiance du lieu de vie des animaux sont considérées comme des facteurs qui participent au développement des maladies respiratoires [30], [157]. Des études ont montré que la limitation de la taille des groupes de 7 bovins a un effet bénéfique sur la prévalence des affections respiratoires [158]. En effet, il a été observé que les pathogènes pouvaient se transmettre malgré des distances supérieures à 4 m [159], [160]. Néanmoins, la transmission entre deux bandes d'animaux est faible, s'il n'y a pas de contact physique ou qu'un vide sanitaire est appliqué, du fait de la faible survie de ces pathogènes dans l'environnement. Une bonne ventilation a un effet bénéfique sur la santé des animaux. Elle permet de diminuer la stagnation de l'air vicié, de maintenir une température et une humidité optimales, ainsi que d'éliminer les gaz nocifs sur le tractus respiratoire, tels que le méthane ou l'ammoniac [161].

Certaines mesures peuvent être appliquées en amont de l'engraissement. En effet la nutrition est primordiale au maintien de fonctions optimales, notamment pour l'immunité. Cela passe par un transfert de l'immunité maternelle au veau dans la période post-natale immédiate. Il a été montré que si le transfert de l'immunité passive est défectueux, le développement de maladies respiratoires est augmenté [162]–[164]. Une étude a démontré aux Etats-Unis en 2007 que près d'un veau sur 5 présente un transfert défectueux [165]. Le sevrage des animaux dès la mise en

lot lors de l'engraissement est un facteur de stress ayant un impact sur le développement des maladies respiratoires [161], [166], [167]. Des études ont déterminé l'âge où le veau est prêt pour le sevrage via la consommation d'aliments concentrés qui doit être comprise entre 0,7 et 1 kg par jour durant 2 à 3 jours consécutifs [165], [166].

La désinfection du nombril permet de réduire de moitié la mortalité des veaux et de faire passer le pourcentage d'animaux traités pour des pathologies respiratoires de 19% à 5% [168]. Une diminution de la température et une augmentation de l'humidité sont des facteurs contribuant à l'augmentation de la prévalence des pathologies respiratoires [169].

Le transport des animaux est aussi considéré comme un facteur de stress. L'utilisation de la metaphylaxie lors du transport des animaux permet la réduction de l'incidence des troubles respiratoires [163]. En effet, l'utilisation de la tulathromycine permet de réduire de 50% la probabilité de développer des troubles respiratoires [170]. Les procédures de quarantaine lors de l'arrivée de nouveaux veaux sont essentielles pour éviter la contamination des individus déjà présents dans les locaux. En effet, il a été montré que l'introduction du *BVDV* dans un élevage se faisait majoritairement lors de l'arrivée de nouveaux animaux [171]

1.1.2.2. La sélection génétique

La sélection génétique est utilisée en élevage bovin laitier afin d'améliorer historiquement la production laitière, mais aussi de plus en plus souvent, la longévité, la fertilité ou encore la résistance aux mammites. Cette sélection a pour but de rechercher et d'utiliser des composantes génétiques présentes chez certains animaux ayant des caractéristiques plus favorables, comme la résistance aux mammites, puis de les utiliser pour la reproduction. Cette sélection utilise des index permettant l'évaluation génétique. L'héritabilité est un paramètre qui est utilisé pour mesurer la part attribuée à la génétique vis-à-vis des infections dans notre cas, dans une population donnée.

Il existe des index spécifiques pour différents critères et un index synthétique qui est pondéré en fonction des objectifs actuels d'amélioration de la race. Les animaux possédant un index favorable pour le caractère voulu seront utilisés pour la reproduction. La sélection génétique pour la résistance aux maladies la plus utilisée est la sélection vis-à-vis des mammites ; elle est fondée sur le score de cellules somatiques qui est calculé à l'aide de la formule suivante : $SCS = \log_2(CCS/100) + 3$. Différentes études ont montré que la valeur du SCS était un critère héritable [172], [173], et qu'il était fortement corrélé avec la fréquence des mammites cliniques dont les R^2 est compris entre 0.67 et 0.97 [174]–[176].

D'autres critères comme des molécules présentes dans le lait après infection pourraient être utilisées dans ce cadre, telle que l'activité de la N-acetyl-B-D-glucosaminidase permettant de différencier les mammites dues à des pathogènes mineurs ou majeurs [177]. L'utilisation de la conductivité électrique a aussi été réalisée permettant d'étudier les changements de concentrations en sodium, potassium et chlore. Il a été mis en évidence que la corrélation entre la conductivité et les mammites cliniques variait de 0,65 à 0,80 [178].

Une autre méthode de sélection génétique pour améliorer la résistance vis-à-vis des mammites est l'utilisation des polymorphismes correspondant à des variations mineures du génome (SNP). Le développement des puces dites SNP à plus ou moins haut débit permet de cribler le génome et de retenir les animaux porteurs d'haplotypes reliés au phénotype d'intérêt. En outre, il a été découvert des variants codant pour la lactoferrine bovine et le lysozyme qui pouvaient être potentiellement exploités [179]. L'utilisation des SNP a permis le développement de vaches transgéniques comme par exemple, une vache résistante à *S.aureus* produisant dans le lait de faibles quantités de lysostaphin qui est une enzyme permettant la destruction des bactéries de type *S.aureus* [180]. Néanmoins, cette méthode cause des problèmes éthiques au sein de la population et n'est pas encore exploitée.

Dans le cadre des bronchopneumonies, la sélection génétique un impact limité du fait que ces maladies résultent de l'interaction de nombreux facteurs et en nombre plus élevé que les mammites.

De plus, la vaccination vis-à-vis de cette maladie et l'amélioration des conditions d'élevage permettent une nette diminution de la prévalence. De ce fait, le développement d'études de sélection génétique n'a que peu impact vis-à-vis des infections respiratoires. Néanmoins, une étude a défini que la résistance à ces troubles à une part génétique avec une héritabilité faible à modérée (0 à 0.26) [181], [182]. Au sein des ateliers d'engraissement, cette héritabilité est plus faible et comprise entre 0,04 et 0,08 [183]. Pour avoir un meilleur résultat avec la sélection génétique, il est nécessaire de comprendre les mécanismes permettant une meilleure protection comme cela est réalisé avec SOCS2 chez la brebis. Il a été montré que des animaux présentant une mutation sur ce gène avait une susceptibilité à l'inflammation plus grande dans le contexte d'une infection mammaire [184].

1.1.2.3. La vaccination

1.1.2.3.1. But

Le terme *vaccination* a été introduit par Edward Jenner en 1796. La vaccination fait partie des mesures préventives, car elle est réalisée avant l'apparition de l'infection. Le but de la vaccination est d'éduquer le système immunitaire afin qu'il soit opérationnel et plus efficace au moment du contact avec l'agent pathogène vis-à-vis duquel elle est censée protéger l'organisme. Parmi les effecteurs de la réponse immunitaire, les cibles de la vaccination sont les lymphocytes B producteurs d'anticorps et les lymphocytes T de type CD4 comme chefs d'orchestre de la réponse de défense, et ceux de type CD8 pour la réponse cytotoxique vis-à-vis des agents viraux, ou certaines bactéries intracellulaires comme *Listeria*. Il existe différents types de vaccins en fonction de leur composition :

- Les vaccins vivants atténués contiennent des agents vivants qui ont perdu une partie de leur virulence. Ces agents peuvent être obtenus spontanément au fur et à mesure des passages en culture, ou bien intentionnellement via l'introduction d'une mutation génétique. Ces vaccins permettent généralement une bonne immunisation car les réponses sont dirigées vis-à-vis de l'ensemble des constituants de l'agent pathogène, dans une forme de réponse qui ressemble beaucoup à l'infection naturelle. Néanmoins ils peuvent parfois causer de fortes réactions en raison de leur immunogénicité. Ils ne contiennent pas d'adjuvant ; c'est le cas du BCG vis-à-vis de la tuberculose chez l'homme ou de nombreux virus. Il n'existe actuellement pas de vaccin utilisant une bactérie qui soit destiné aux bovins sur le marché européen.
- Les vaccins inactivés contiennent un ou plusieurs agents pathogènes qui ont été tués chimiquement ou par la chaleur. Contrairement aux précédents, ces vaccins requièrent la présence d'un adjuvant pour que la réponse soit suffisamment immunogène et proche de celle des vaccins atténués. Un vaccin inactivé peut contenir l'agent entier et permettre la production d'une réponse assez diverse vis-à-vis de ses constituants, ou bien une fraction seulement de l'agent ; il est alors dit sous-unitaire.
- Il existe d'autres types de vaccins comme les vaccins à ARN, à ADN ou encore les vecteurs recombinés. Une partie de ces nouvelles méthodes ont été utilisées dans le cadre de la vaccination vis-à-vis de la COVID-19, mais ils sont rares en médecine vétérinaire, et non commercialisés pour les bovins à l'heure actuelle.

Une majorité de vaccins nécessite une ou plusieurs injections de primo-vaccination, et ensuite des rappels à intervalles variables de 6 à 12 mois, pour une réponse forte et durable, que l'on

espère protectrice vis-à-vis de la maladie, de l'infection, et de la transmission de l'agent infectieux aux congénères. Une majorité de vaccins animaux ont des durées de protection relativement courtes, obligeant à des injections de rappel assez fréquentes.

De nombreux vaccins, contiennent des adjuvants comme l'aluminium ou divers composés lipidiques, possiblement à l'origine d'effets indésirables, entraînant une réticence de la population pour le recours à la vaccination.

La vaccination va permettre l'éducation de lymphocytes T associés à la réponse cellulaire et des lymphocytes B responsables de la production d'anticorps. Il en résulte la création d'un répertoire de cellules à caractère mémoire, qui peuvent être mobilisées et réactivées lors de l'infection par le pathogène, et ainsi conférer un certain degré de protection.

1.1.2.3.2. Vaccination contre les mammites

Comme indiqué précédemment, les mammites dans l'espèce bovine sont causées par différentes espèces bactériennes dont les majoritaires sont *E. coli*, *S. uberis* et *S. aureus*, avec pour chacune des spécificités pathogéniques et de la réponse de l'hôte. Ces mammites peuvent être résolues par les effecteurs innés de la réponse immunitaire ; la participation du système adaptatif est susceptible d'accroître ou d'aider au contrôle de ces infections. La vaccination peut donc s'avérer utile, notamment dans l'éducation des cellules innées par les lymphocytes T auxiliaires, comme les T producteurs d'IL-17 comme nous l'avons indiqué précédemment. La quantité de cellules de ce type est 100 fois moindre dans le lait que dans le sang [185].

A l'échelle mondiale, la vaccination contre *E. coli* est probablement la plus répandue dans les élevages laitiers. Divers vaccins utilisent une souche mutante de *E. coli* appelée *E. coli* J5, caractérisée par l'absence de l'antigène O de la capsule, mais qui contient tous les autres antigènes de la bactérie (LPS, protéines de membrane, lipide A). L'avantage de ces antigènes est qu'ils sont partagés entre différentes bactéries Gram-négatives [186]–[190]. Il a été déterminé que l'augmentation du taux d'anticorps vis à vis d'*E. coli* J5 était corrélée avec une diminution de l'incidence des mammites cliniques dû à ce type de bactéries [191]. Plusieurs études indiquent une diminution de la prévalence des mammites chez les animaux vaccinés [192]–[194]. Néanmoins, l'immunisation avec ce vaccin ne permet pas de prévenir totalement ces mammites [195]. En effet, l'immunisation au moyen de deux injections de ce vaccin permettait de diminuer la proportion de mammites cliniques sévères, de diviser par 3 la mortalité, mais ne diminuait pas l'incidence globale [196]. De nombreuses études ont montré que l'immunité induite était de courte durée, obligeant à des rappels très fréquents. Pour

améliorer l'efficacité vaccinale, des essais ont évalué l'augmentation du nombre d'injections ou l'effet du lieu de ces injections [197]–[199]. Un autre vaccin basé sur un mutant de *Salmonella* Typhimurium (Re-17) est commercialisé mais les études concernant ce produit sont peu nombreuses. Une des dernières a montré une diminution de l'incidence et de la morbidité après deux immunisations avec ce vaccin [200].

Le développement de vaccins contre *S. aureus* est difficile à réaliser à cause d'une grande diversité antigénique des isolats. La recherche d'un antigène universel qui offrirait une protection vis-à-vis de toutes les souches était jusqu'à présent difficile. La deuxième limite concerne les facteurs de virulence produits par ces bactéries leur permettant d'échapper au système immunitaire. Enfin, les effets, et surtout les modalités de cette vaccination sur l'apparition et l'évolution de la mammite, et la santé de l'animal, sont encore insuffisamment connus. En effet, ce type bactérien est à l'origine de l'apparition de mammites chroniques dont la résolution par la réponse immunitaire semble peu efficace. Différents développements vaccinaux vis-à-vis de ce pathogène ont été tentés, en utilisant différents facteurs de virulence ou différentes souches. Un vaccin contenant plusieurs sérotypes issus des différentes souches bactériennes est commercialisé par Boehringer (Lysigin) aux USA. Il a été montré que ce vaccin permettait de diminuer les infections mammaires causées par *S. aureus* et les staphylocoques à coagulase négative [201]. Une épreuve infectieuse après vaccination a montré que la vaccination n'empêche pas le développement d'une mammite, mais que sa sévérité est réduite avec une augmentation des anticorps de type IgG1 dirigés contre *S. aureus* et une élimination plus rapide [202]. Néanmoins, une étude sur des vaches multipares a montré que cette vaccination ne permet pas de diminuer le nombre de quartiers infectés par *S. aureus*, ni la concentration des cellules somatiques dans le lait, l'hypothèse étant que ce vaccin n'induirait pas suffisamment d'anticorps pour la phagocytose [203]. Un autre vaccin basé sur des bactéries tuées a permis de réduire le nombre de quartiers infectés ainsi que la valeur des CCS lors d'une infection à *S. aureus* [204]. Ce vaccin utilise l'adjuvant de Freund contenant en outre un adjuvant huileux, des antigènes de mycobactéries qui empêche son utilisation comme vaccin commercial car il cause des réactions fortes au site de l'injection, et interfère avec le dépistage de la tuberculose bovine. Une vaccination basée sur un polysaccharide jouant un rôle dans la formation de biofilm (SAAC) a permis une diminution de la quantité de bactéries dans les quartiers après inoculation [205]. Un vaccin utilisant cette molécule et couplé à *E. coli* a été approuvé dans l'Union Européenne avec l'argument qu'il permettait de diminuer le taux de contamination entre vaches et la concentration des cellules somatiques chez les animaux vaccinés [206], [207]. L'utilisation de la vaccination contre *S. aureus* est souvent peu

avantageuse dans les élevages, au contraire d'une amélioration de l'hygiène qui a généralement un effet bénéfique.

La vaccin STARTVAC (laboratoire HIPRA) contenant *E. coli* J5 inactivé, *S. aureus* PC8 inactivé a été commercialisé en 2009. Une étude a montré que ce vaccin permettait une activation de la réponse immunitaire adaptative. En effet, la vaccination semble induire une activation lymphocytaire de type Th2 avec une augmentation de la réponse anticorps, mais avec absence de réponse de type Th1 dans le cadre d'une infection à *S. aureus*. Dans le cadre d'une infection avec *E. coli*, la réponse immunitaire induite est moins claire avec néanmoins une forte production d'interféron gamma indiquant une polarité vers le phénotype Th1 [208].

Le développement de vaccins vis-à-vis de certaines bactéries du genre *Streptococcus* a été peu actif à cause de l'efficacité des mesures de gestion. En effet, via l'amélioration de l'hygiène des conditions de traite, les mammites provoquées par *S. agalactiae* et *S. dysgalactiae* ont été assez facilement maîtrisées et ont quasiment disparu à présent. Néanmoins, les mammites causées par *S. uberis* sont beaucoup plus fréquentes. Il a été montré que la bactérie pouvait persister au sein des cellules épithéliales mammaires de fait, l'opsonisation par les anticorps spécifiques et les facteurs antibactériens produits à la suite de l'infection sont insuffisants pour éliminer efficacement l'infection [209], [210]. Les bactéries isolées de mammites cliniques sont très diverses du point de vue génétique, alors que certains isolats sont prédominants dans les mammites sub-cliniques [14]. Des chercheurs du Royaume Uni ont évalué un protocole vaccinal associant un vaccin vivant avec l'administration intramammaire d'un extrait de protéines bactériennes. Ce vaccin permet une diminution des signes cliniques, mais il est dépendant de la souche infectante [211]. Un autre vaccin fondé sur un facteur de virulence (SUAM) correspondant à une protéine d'adhésion a été produit. Ce facteur est très conservé entre les différentes souches de *S. uberis* [212]. Une étude cherchant à démontrer son efficacité a été réalisée, et a montré que la réponse inflammatoire, la production de lait ainsi que la date de retour à un seuil de CCS inférieur à 200 000 cellules/mL n'étaient pas différents entre le groupe vacciné et le groupe témoin malgré une légère modification de la réponse immunitaire [213].

Il a été montré *in vitro* que les anticorps développés à la suite de l'immunisation permettent une diminution de l'internalisation de la bactérie [213]. Néanmoins cette production n'est pas suffisante et ne permet pas la bactéricidie. L'activation spécifique de cellules immunitaires tels que les lymphocytes T CD8+ semble nécessaire.

1.1.2.3.3. Vaccination contre les broncho-pneumonies infectieuses

A l'origine des bronchopneumonies infectieuses, de nombreux pathogènes peuvent être impliqués. Cela complexifie d'autant la vaccination car l'animal doit être protégé simultanément vis-à-vis de plusieurs virus et/ou bactéries. Des schémas vaccinaux contre ces agents pathogènes sont réalisés pour protéger les jeunes bovins d'élevage (destinés au renouvellement du troupeau) et dans le cadre particulier des élevages d'engraissement. Cela concerne surtout les pathogènes viraux, avec un taux de vaccination supérieur à 80% aux Etats-Unis. En revanche, la vaccination vis-à-vis des bactéries concerne une plus faible proportion d'animaux ; elle est comprise entre 53 et 62% [39]. Néanmoins, la vaccination contre les agents viraux protège en partie vis-à-vis des infections bactériennes car les virus prédisposent à ces infections secondaires.

1.1.2.3.3.1. Vis-à-vis des virus

Le BRSV est fréquemment identifié chez les bovins dans les ateliers d'engraissement [214] et les animaux développent souvent des symptômes moins de 7 jours après leur arrivée [215]. La vaccination est utile mais l'utilisation d'une souche n'induit pas de protection vis-à-vis de la totalité des isolats [216]. Les animaux vaccinés vis-à-vis de ce virus sont 2 fois moins traités pour des bronchopneumonies infectieuses que les animaux non vaccinés [217]. Une équipe de recherche a comparé un vaccin quadrivalent (*BPI3*, *BRSV*, *BVDV* et *BoHV-1*) et un vaccin trivalent (*BPI3*, *BVDV* et *BoHV*) et les résultats montrent une diminution de la morbidité et de la mortalité respiratoire [218]. La vaccination contre ce virus est donc nécessaire mais le choix entre un vaccin vivant atténué et un vaccin tué se pose. La majorité des vaccins actuellement commercialisés nécessite 2 injections pour permettre une protection [218]. La durée de protection semble être assez similaire quel que soit le type de vaccin, vivant atténué ou tué, qui est employé. Une étude comparant une immunisation avec des vaccins tués (*BRSV*, *BPI3* et *M. haemolytica*) et des vivants atténués (*BVDV* et *BRSV*) démontre que le vaccin inactivé produit de meilleurs résultats que le vaccin vivant atténué [219].

En raison d'un impact économique important dans les ateliers d'engraissement à cause des infections persistantes, le *BVDV* contribue à l'apparition ou à l'aggravation des BPI, probablement par l'immunodépression transitoire et le contexte favorable aux autres infections qu'il provoque).

Le *BoHV-1* est la cause de la rhinotrachéite bovine et engendre des symptômes au niveau respiratoire et oculaire. Il est quelquefois présent chez l'animal sous forme latente et une

augmentation du stress entraînant une production de cortisol permet la réactivation du virus [216]. Ce dernier est très contagieux dans les ateliers d'élevage du fait de la proximité entre les animaux car il diffuse par les sécrétions respiratoires. Il a été établi que la réponse humorale était efficace contre ce type de virus mais qu'elle n'empêche pas la réactivation du virus en cas de latence, ni l'infection et l'établissement de la latence. Le défi de la vaccination contre ce virus est qu'elle protège vis-à-vis de l'infection et de l'entrée en latence de celui-ci. Une étude a mis en évidence qu'une vaccination avec un vaccin multivalent contenant des souches de *BVDV*, *BPI3*, *BRSV* et *BoHV-1* à des animaux non vaccinés permettait un gain de poids plus grand, une réduction des signes cliniques et de la diffusion virale s'il est administré 72h avant le contact avec le virus [220]. Une autre étude a déterminé qu'une administration d'un vaccin vivant atténué vis-à-vis du *BoHV-1* entraîne une diminution de l'incidence des maladies respiratoires, de 17% à 1%. Vis-à-vis du *BoHV-1*, il existe des vaccins tués et des vaccins vivants atténués. Les vaccins tués sont onéreux, ils ont une faible durée d'immunité et nécessitent donc 2 doses à un mois d'intervalle, ce qui est peu compatible avec le fonctionnement des ateliers d'élevage alors que les vaccins vivants atténués ne nécessitent qu'une seule administration car l'immunité induite est plus forte. Concernant les vivants atténués, ils peuvent être injectés au niveau systémique par voie parentérale ou au niveau local par voie nasale. La vaccination par voie systémique entraîne une protection rapide avec une seule dose [216]. Celle par voie nasale a l'avantage d'induire une réponse locale rapide [220]. Des études comparant les 2 types ont eu lieu et il n'y avait pas de différence sur le gain moyen quotidien ou la morbidité [221]. Dans le cas de ce virus, il y a peu de diversité entre les isolats et l'immunisation avec une souche permet d'avoir une immunité croisée avec les autres souches.

Le *virus parainfluenza 3* ou *BPIV3* participe à l'apparition des troubles respiratoires chez le veau. Les signes développés à la suite de l'infection par ce virus sont souvent limités mais il peut être associé à d'autres pathogènes, souvent bactériens, pour provoquer des bronchopneumonies infectieuses [222]. Néanmoins beaucoup d'animaux arrivant en atelier d'élevage ont été en contact avec le *BPIV3* et possèdent donc déjà une immunité vis-à-vis de ce virus ; la vaccination a donc peu de plus-value [216]. De plus, il est difficile d'étudier le rôle seul d'un vaccin *BPIV3* car il est souvent inclus dans des vaccins multivalents.

1.1.2.3.3.2. Vis-à-vis des bactéries

La pasteurellose est causée principalement par deux types de bactéries, *Mannheimia haemolytica* et *Pasteurella multocida*. Celles-ci sont retrouvées dans de nombreux cas de maladies respiratoires notamment après une infection virale. *P. multocida* cause des infections respiratoires chez les jeunes veaux alors que *M. haemolytica* infecte des catégories plus diverses d'animaux, mais la première implique la participation d'autres agents, notamment viraux, pour provoquer la maladie respiratoire [223]. Le sérotype S1 de *M. haemolytica* est la forme pathogène la plus répandue dans les cas fatals de maladies respiratoires [224]. Les anticorps vis-à-vis de *M. haemolytica* et leurs interactions auraient un rôle dans la pathogenèse des bronchopneumonies infectieuses [225]. Cette dernière produit une leucotoxine jouant un fort rôle dans la pathogenèse. Pour prévenir cette infection, il est important de stimuler la production d'anticorps vis-à-vis de la leucotoxine, mais aussi vis-à-vis des antigènes de la capsule bactérienne. La composition des vaccins actuellement commercialisés est variable et ils contiennent la bactérie seule, la bactérie et la leucotoxine, ou bien des extraits bactériens [226]. Néanmoins tous ces vaccins ont en commun l'utilisation du sérotype A1. Il a été démontré que les vaccins contenant la leucotoxine entraînaient une meilleure protection croisée entre les souches [227]. Le vaccin utilisant une souche de *M. haemolytica* vivante atténuée a été montré efficace pour prévenir l'apparition des troubles [228], alors que celui utilisant des antigènes de *M. haemolytica* a été montré peu performant [229].

Hemophilus somnus est une bactérie impliquée également dans les bronchopneumonies infectieuses. Elle peut causer des infections systémiques ayant un fort taux de mortalité [230]. Une seule dose de vaccin vis-à-vis de cette bactérie est plus délétère par rapport à des animaux non vaccinés ou des animaux complètement vaccinés [231]. Il y a peu de consensus sur l'utilisation de la vaccination envers cette bactérie.

Mycoplasma bovis est une bactérie de plus en plus présente dans les ateliers d'engraissement. En effet une étude a établi que 86% des poumons d'animaux issus d'ateliers d'engraissement étaient positifs aux mycoplasmes et notamment à cette bactérie [232]. De plus, il a été montré que plus de 90% des animaux ayant des pneumonies résistantes aux antibiotiques hébergent *M. bovis* au niveau pulmonaire [233]. En effet cette espèce est résistante aux antibiotiques en raison de l'absence de paroi cellulaire [234]. La vaccination est donc une des seules méthodes empêchant le développement de cette bactérie. Des tests de vaccinations ont été réalisés et ont permis une réduction des lésions respiratoire causée par *M. bovis* seule ou en co-infection avec *BRSV* [235]. Un autre test utilisant une préparation issue de cellules infectées par *M. bovis*

saponisées permettaient de réduire les infections pulmonaires chez le veau dues à cette bactérie [234].

1.1.2.3.4. Limites / Perspectives

Le développement de la vaccination a permis de lutter efficacement vis-à-vis de pathologies touchant l'espèce bovine. Cet élan est dû à une nécessité de diminuer l'utilisation des antibiotiques. Néanmoins, la vaccination n'est pas parfaite et la prévalence de nombreuses maladies ne peut être diminué.

Un frein majeur au développement de la vaccination concerne la variabilité antigénique relativement élevée des virus [216], ce qui peut permettre l'échappement du virus à la réponse immunitaire malgré un protocole vaccinal adapté et parfaitement réalisé.

Le second frein concerne la réactivité croisée du vaccin. Il a été montré *in vitro* que le vaccin inactivé induisait une réponse humorale avec une réactivité croisée similaire à ceux des vaccins vivants modifiés [236]. Concernant la réponse cellulaire induite par les vaccins vivants atténués, elle est similaire quel que soit le biotype utilisé [237]. Néanmoins il a été établi que l'animal avait une meilleure réponse lors d'une infection si le virus correspondait à celui utilisé pour la vaccination [238].

Le troisième frein concerne la durée de l'immunité qui varie selon le type de vaccin utilisé. Le problème de l'utilisation de vaccins vivants modifiés avec le *BVDV* est l'induction d'une immunodépression temporaire, mais elle permet d'avoir une longue durée de protection. Il a été démontré que le taux d'anticorps ne diminue que 4 mois et demi après l'immunisation [236]. Cependant, une étude a mis en évidence que la vaccination avec un vaccin vivant atténué commercial pouvait entraîner une immunosuppression prolongée [239]. De plus il a été montré que le virus du fait de son fort taux de mutation pouvait redevenir virulent [240]. En revanche l'utilisation de vaccins tués ne cause pas d'immunodépression mais leur induction de réponse immunitaire est moins forte. Ce dernier type de vaccin nécessite plusieurs injections qui sont quelquefois non réalisées dans le cadre pratique. Il existe de nombreux vaccins contre ce virus avec chacune leur spécificité et il n'existe pas de consensus sur le type de vaccin à utiliser.

Enfin, le dernier frein concerne la vaccination vis-à-vis des bactéries. La majorité des vaccins ont pour objectif la production d'anticorps. Or cette réponse n'est souvent pas suffisante et il est donc nécessaire de développer des vaccins activant la réponse cellulaire qui est mobilisée et efficace contre les bactéries.

L'objectif des développements actuels est de comprendre les mécanismes de protection vis-à-vis de ces maladies. Les méthodes dites omiques ont un fort intérêt dans les études pour la compréhension des mécanismes.

1.1.3. Les enjeux autour de la santé des bovins

L'utilisation intensive des antibiotiques en médecine humaine et vétérinaire est devenue un véritable problème de santé publique. Aussi, tous les secteurs sont fortement incités à réduire leur utilisation comme cela est déjà le cas depuis plusieurs années. Le développement de méthodes alternatives telles que la vaccination est également promue pour pallier la réduction des traitements antibiotiques et pour conserver un statut sanitaire de bon niveau au sein des troupeaux.

1.1.3.1. Les enjeux vis-à-vis des mammites

L'inflammation mammaire d'origine infectieuse est l'affection la plus communément observée dans les troupeaux de vaches laitières. Ces infections sont causées majoritairement par trois espèces bactériennes que sont *S. aureus*, *S. uberis* et *E. coli*, avec des proportions qui varient selon les régions et les types d'élevage laitier. Le traitement contre les mammites cliniques repose sur l'utilisation d'antibiotiques, dont les bêta-lactamines, des céphalosporines, parmi les plus utilisés. Ces molécules sont parfois utilisées de manière préventive pour empêcher l'apparition de mammites. Les bénéfices de cette utilisation sont une diminution de l'incidence des infections en début de lactation, avec une diminution de la morbidité et *in fine* une augmentation de la productivité des animaux. Néanmoins, cette utilisation intensive fait craindre la sélection de bactéries résistantes chez des espèces bactériennes touchant aussi l'homme. En outre, le lait des animaux traités contient des résidus d'antibiotiques, ce qui le rend non commercialisable [241].

Une étude de la résistance aux antibiotiques entre 2006 et 2016 au sein des élevages français a été réalisée. Elle a mise en évidence que plus de 32% des isolats de *S. uberis* étaient résistants à au moins 1 ou 2 antibiotiques, et que 14,5% étaient multi-résistants. Concernant *E. coli*, les chiffres sont de 23% et de 6% alors que pour *S. aureus*, ils sont de 34% et de 3,6%, respectivement [242].

Les conséquences économiques des mammites chez la vache laitière sont importantes à cause de plusieurs facteurs tels que la réforme précoce, la diminution de la production et de la qualité du lait, ainsi que le coût de traitement et de prévention. Une étude a montré que le taux de

mortalité due aux mammites était de 0,22% en France, ce qui est faible [243]. Par contre, les mammites sont considérées comme la 1^{ère} cause de réforme et sont donc responsables de la réduction de la longévité des animaux. La perte de production de lait due à une mammite subclinique tous germes confondus, est estimée entre 0,3 et 0,6 kg par jour selon le type de vache. Les mammites cliniques causent des pertes beaucoup plus marquées, équivalentes à 5% de la production de l'animal [30]. De plus, le lait est soumis à des conditions réglementaires concernant le nombre de cellules somatiques et la présence de résidus d'antibiotiques, à l'origine de la non-commercialisation du lait au cours d'un épisode de mammite clinique, et une perte financière conséquente. D'un point de vue économique, l'impact global des mammites est estimé entre 80 et 93 euros par vache et par an [244].

1.1.3.2. Les enjeux vis-à-vis des bronchopneumonies infectieuses

Les bronchopneumonies infectieuses touchent majoritairement les veaux et correspondent à une des deux principales causes de pertes économiques en élevage bovin. Celles-ci ont un impact économique fort à cause des coûts de traitement, mais aussi des répercussions sur la santé des animaux avec une augmentation de la mortalité, une réduction de la croissance et du gain de poids. Ces maladies peuvent causer jusqu'à 75% de morbidité [245], et entre 5 et 7% de mortalité dans les ateliers d'engraissement [246]. Elles affectent le gain de poids quotidien, le poids final, et la qualité de la carcasse [247]–[249]. Cinquante pour cent des animaux infectés par une BPI ne présentent pas de symptômes conduisant à une sous-estimation de la prévalence générale dans les élevages [250]. Le coût économique global dans ces ateliers est estimé entre 800 et 900 millions de dollars à cause des pertes directes (mort) et indirectes comme les frais de traitement [251] ; celui-ci serait sous-estimé. En effet, de récentes études ont montré que les infections respiratoires entraînent un allongement de la durée d'engraissement en atelier pour atteindre le poids requis, et une diminution de la qualité de la carcasse [252], [253]. Cette dernière étude a estimé l'effet d'un ou plusieurs traitements sur le bénéfice final, avec une baisse de 38\$ si l'animal a reçu un seul traitement, 167\$ pour deux, et 230\$ pour trois traitements. Une étude similaire a été réalisée au sein des ateliers d'engraissement en Australie. Il a été montré que les BPI cliniques coutaient 214\$ par animal, et jusqu'à 385\$ dans le cas où trois traitements ont été appliqués. Les pertes dues aux BPI subcliniques sont estimées à 67\$ par animal [254].

Les BPI au sein des ateliers d'engraissement ont donc un fort impact économique. Pour lutter contre ces maladies multifactorielles, plusieurs démarches sont possibles. Lorsque l'animal est atteint, il n'y a pas d'autre alternative que le recours au traitement avec l'utilisation d'un antibiotique. La première étude sur la fréquence et la nature de la résistance aux antibiotiques des bactéries pathogènes isolées de l'appareil respiratoire des bovins a été publiée en 1976 [255]. Les résistances acquises par les bactéries causant des BPI touchent de nombreuses familles d'antibiotiques (tétracyclines, fluoroquinolones, bêta-lactamines, macrolides, sulfonamides, ...). Il existe plus de 34 protocoles différents pour l'administration d'antibiotiques en cas d'infection pulmonaire. Une méta-analyse a été réalisée sur l'ensemble de la littérature pour définir les meilleures combinaisons via des méthodes statistiques. Il a été montré que la classe des macrolides avaient les meilleurs résultats vis-à-vis des pathogènes causant des BPI [256].

Une alternative aux traitements avec des antibiotiques est l'utilisation de méthodes préventives, qui sont recommandées, mais dont les résultats bien que prometteurs ne sont pas encore à la hauteur des attentes.

Trois grands ensembles de mesures préventives peuvent être envisagés. Ils concernent l'amélioration des méthodes d'élevage au sens large, la réduction de la sensibilité des animaux vis-à-vis des maladies par le renforcement de leur immunité vis-à-vis des agents pathogènes les plus prévalents, par la vaccination par exemple, et la sélection génétique. Ils seront vus dans le chapitre suivant.

1.2. Sciences animales et biostatistiques

1.2.1. *Historique des biostatistiques et de l'analyse de données*

1.2.1.1. Histoire des statistiques

La statistique est une discipline dont l'existence remonte aux premières civilisations. A cette époque, cela consistait à la collecte de données tels que des recensements animaliers ou humains. Cette discipline n'a que peu évolué en Europe entre le monde antique et le XVI^{ème} siècle. Durant cette même période, la civilisation des Incas, entre le XV^{ème} et le XVI^{ème} siècle de notre ère possédait des connaissances avancées en mathématiques, et en statistiques notamment [257].

A partir du XVII^{ème} siècle, des développements sur la statistique et la probabilité ont lieu via la publication de traités indépendants par Blaise Pascal (1665) et Pierre de Fermat (1679) sur la probabilité des jeux de hasard [258], [259]. Ces travaux sont poursuivis par Christiaan Huygens avec une introduction de la notion d'espérance mathématique. Celui-ci va publier le « premier livre » liés aux probabilités en 1657 [260]. Enfin à la fin de ce siècle, les travaux de Jakob Bernoulli ont été publiés avec 5 documents traitant des séries infinies. Il va écrire un ouvrage mathématique « *ars conjectandi* » en 1692 mais qui ne sera publié que 8 ans après sa mort par son neveu Nicolas Bernoulli [261]. Ce livre va se baser sur les travaux précédemment cités pour consolider la théorie des probabilités et introduit ce que l'on nomme aujourd'hui la loi des grands nombres et *in fine* la loi de Bernoulli. La famille Bernoulli est composée de nombreux mathématiciens qui vont avoir un impact majeur dans le développement des mathématiques entre le XVII^{ème} et le XVIII^{ème} siècle.

A partir du XVIII^{ème} siècle, le développement des domaines des statistiques et des probabilités évolue. Les travaux réalisés par Jakob Bernoulli sont complétés par divers mathématiciens (Pierre Rémond de Montmort, Nicolas Bernoulli, Abraham de Moivre, Georges Louis Leclerc de Buffon et Daniel Bernoulli). L'ensemble de ces travaux permettent de définir la loi des grands nombres ainsi que d'améliorer la notion de récurrence. Les résultats obtenus et notamment ceux d'Abraham de Moivre concernant le théorème central limite et le cas spécial de la loi de Bernoulli.

En parallèle de ces études, le mathématicien Antoine Deparcieux écrit en 1746 un essai intitulé « *Essais sur les probabilités de la durée de la vie humaine, d'où l'on déduit la manière de déterminer les rentes viagères tant simples que tantines, précédé d'une courte explication sur*

les rentes à terme, ou annuités, et accompagné d'un grand nombre de tables » [262]. Cet ouvrage contenait les premières données statistiques correspondant à des tables de mortalité. Ces informations furent utilisées par les compagnies d'assurance-vie.

Durant ce même siècle, Thomas Bayes élaborera le théorème de Bayes basé sur des probabilités inversées qui fut publié à titre posthume en 1763. Ce théorème est très utilisé aujourd'hui dans le domaine de l'intelligence artificielle. En effet il permet de calculer des probabilités conditionnelles, c'est-à-dire « la probabilité d'un événement sachant qu'un autre événement a eu lieu [263], [264].

Mais ce ne fut qu'à partir du XIX^{ème} siècle que la discipline des statistiques explose avec l'émergence de recherches dans de nombreux domaines (Démographie, Physique, Informatique ...). En parallèle de l'utilisation des statistiques pour l'étude historique de la démographie, un développement des représentations visuelles associées a eu lieu avec notamment l'utilisation de diagrammes circulaires ou « pie chart » développés par William Playfair en 1801 [265]. L'une des pionnières dans ce domaine fut Florence Nightingale qui utilisa une version évoluée de ces représentations pour étudier l'effet des saisons sur la mortalité au sein d'un hôpital militaire [266]. Ce fut l'une des premières études utilisant les statistiques dans le domaine de la santé. De nombreuses lois utilisées aujourd'hui en statistiques ont été développées durant ce siècle tels que le théorème central limite publié en 1812 par Pierre Simon de Laplace. Ce théorème indique qu'une somme de variables aléatoires suit une loi normale. Celle-ci est une des lois de la probabilité les plus utilisées dans le domaine biologique pour la modélisation de phénomènes naturels. La première idée de la loi normale est basée sur le travail d'Abraham de Moivre lors d'une étude d'un système pile-face ayant eu lieu en 1733. Ces travaux furent repris par Pierre Simon de Laplace et Carl Friedrich Gauss pour donner le théorème cité précédemment [267], [268]. Le nom de loi normale ne sera donné qu'à la fin du XIX^{ème} siècle par Henri Poincaré. Pierre-Simon Laplace travailla aussi sur les probabilités inversées et notamment sur le théorème de Bayes. La régression linéaire fut développée au XVIII^{ème} siècle par Ruder Josip Boskovic [269].

Dans le cadre d'études dans le domaine de l'astronomie, Adrien-Marie Legendre en 1805 et Carl Friedrich Gauss ont développé indépendamment la méthode des moindres carrés permettant l'étude comparative entre jeux de données expérimentales [270]. A la fin de ce même siècle, Karl Pearson développa le coefficient de Pearson permettant l'étude de la corrélation entre 2 jeux de données, ainsi que le test du χ^2 [271]. Durant ce siècle, les statistiques se sont considérablement développées via l'impulsion de Pearson. Les statistiques

sont considérées comme explicatives mais aussi descriptives via le développement de représentations statistiques. Dans les années 1880, John Venn publia un ouvrage intitulé «*On the Diagrammatic and Mechanical Representation of Propositions and Reasonings* » dans lequel les diagrammes de Venn sont introduits. A cette époque, ils étaient nommés Eulerian Circles avant d'être renommés en 1918 en l'honneur de John Venn [272].

Francis Galton développa dès la fin du XIX^{ème} siècle les premières méthodes statistiques permettant la comparaison de données biologiques telle que la régression ou encore la corrélation. De plus, il est le premier à introduire la notion d'analyses factorielles qui seront développées au XX^{ème} siècle [273]. Il est aujourd'hui considéré comme le « père des biostatistiques ».

Au cours du XX^{ème} siècle, le développement des statistiques fit un bond majeur via notamment les travaux de Karl Pearson, Ronald Aylmer Fisher, William Gosset et George Snedecor. Au début de ce siècle, Charles Spearman développa une nouvelle méthode de corrélation connu aujourd'hui sous le nom de corrélation de Spearman [274].

La loi de Student fut écrite par William Gosset en 1908. Cette loi fut nommée Student car il l'a publiée sous ce pseudonyme, car il lui était interdit de publier en son nom [275].

Les tests statistiques permettant une recherche appliquée ont été introduits au XVIII^{ème} siècle mais ce n'est qu'au début du XX^{ème} siècle qu'ils furent utilisés par Karl Pearson et améliorés par la suite par de nombreux mathématiciens.

Ronald Aylmer Fisher publia en 1922 dans le journal de la Royal Society de Londres l'article «*Sur les fondements mathématiques de la statistique théorique* » où il introduit une dizaine de notions statistiques fondamentales encore utilisées aujourd'hui tels que l'estimateur du maximum de vraisemblance ou encore l'analyse de la variance. C'est lors d'une rencontre avec William Gosset qu'il développa cette dernière méthode [276].

En 1925, il publia un autre ouvrage intitulé «*Statistical Methods for Research Workers* » où il utilise la distribution de Student, du χ^2 , et celle qui sera appelée plus tard comme celle de Fisher-Snedecor [277]. En plus de ces travaux sur les théorèmes, il publia un ouvrage en 1935 «*The Design of Experiments* » sur les plans d'expérience en Agriculture [278]. Le fils de Karl Pearson, Egon Pearson et Jerzy Neyman ont beaucoup développé la théorie des tests d'hypothèses avec notamment la définition des risques d'erreurs alpha et bêta et les intervalles de confiance encore utilisés aujourd'hui [279]. Jerzy Neyman a élaboré en 1934 la théorie d'échantillonnage très utilisée à l'heure actuelle pour l'exploitation des sondages [280].

Lors de la seconde guerre mondiale, Alan Turing et Bletchley Park ont déchiffré le code Enigma allemand grâce aux statistiques bayésiennes [281]. Lors des années 1950, deux chercheurs Richard Doll et Bradford Hill ont établi le lien entre le cancer du poumon et la consommation de tabac par la cigarette permettant une prise ne compte du tabagisme dans les actions préventives de santé publique [282]. En 1958, l'estimateur de Kaplan et Meier a permis de déterminer les meilleurs traitements à utiliser permettant de sauver des millions de vie vis-à-vis de maladies telles que le cancer [283]. Durant cette même période, Frank Wilcoxon (1945), Henry Mann (1947) et Donald Ransom Whitney (1947) ont développés le Test de Wilcoxon-Mann-Whitney permettant la comparaison d'échantillons ne suivant pas une loi normale [284], [285].

En 1977, John Tukey développa les boites à moustaches ou box-plot correspondant à la représentation de nombreuses informations sur un seul graphique [286]. Pour continuer dans le développement de la visualisation de données, Edward Tufte publia en 1982 l'ouvrage « *The visual Display of Quantative Information* » présentant de nombreuses modalités de représentation graphique [287].

En 1993, le langage R qui permet de réaliser des analyses statistiques plus poussées apparait. Celui-ci est encore un logiciel majeur dans l'analyse statistique dans le domaine de la biologie [288].

L'ensemble de ces travaux ont permis le développement de tests statistiques très utilisés dans le domaine de la biologie tels que l'analyse de variance (Fisher), le test de Student (Gosset), le test du χ^2 (Pearson) ou encore le Test de Wilcoxon-Mann-Whitney.

1.2.1.2. Historique de l'analyse de données

1.2.1.2.1. Analyses factorielles

Les analyses factorielles qui seront discutées ci-dessous nécessitent des jeux de données bien spécifiques organisés sous la forme de tableaux de contingence. Une représentation des données sous cette forme fut pour la première fois introduite par Karl Pearson en 1904 [289].

Francis Galton est le premier à introduire la notion d'analyses factorielles qui sera développée au cours du XX^{ème} siècle. En fonction du type de données et de la disposition des données, différentes méthodes sont disponibles.

La première méthode développée fut l'analyse en composantes principales ou ACP d'après les travaux de Karl Pearson et ceux de Francis Galton qui furent publiés en 1901 [290]. Cette méthode a aussi pour nom transformée de Karhunen-Loève ou transformée de Hotelling, car

c'est Harold Hotelling qui a redéveloppé cette méthode dans les années 1930 [291]. Il développa aussi l'analyse canonique des corrélations dérivée de l'ACP, permettant la comparaison de variables quantitatives appliquée aux mêmes individus. Dans les années 1960, Jean-Paul Benzécri et Brigitte Escofier proposèrent l'analyse factorielle des correspondances en utilisant les travaux de Herman Otto Hirschfeld permettant une analyse factorielle de données qualitatives [292]. Néanmoins, les fondements de cette méthode furent publiés par Chikio Hayashi en 1954 [293]. L'analyse factorielle des correspondances multiples utilisant un tableau de données qualitatives a commencé son développement entre les années 1940 et 1960 avec les travaux de Louis Guttman, Cyril Burt et Chikio Hayashi. Avant cette dénomination, cette méthode fut connue sous le nom de « *Dual Scaling* (Shizuhiko Nishisato, 1952) » ou « *Homogeneity analysis* (Albert Gifi, 1990) » [294]. L'analyse factorielle multiple (ou AFM), est une extension de l'ACP et de l'AFC. Elle fut développée dans les années 1980 par Brigitte Escofier et Jérôme Pagès [295]. Cette méthode permet l'analyse de données catégorisées, qu'elles soient qualitatives ou quantitatives. Brigitte Escofier et Gilbert Saporta développèrent en 1979, l'analyse factorielle de données mixtes permettant l'analyse de données qualitatives ou quantitatives sans regroupement nécessaire [296]. L'ensemble de ces méthodes a vu leur utilisation s'accroître considérablement avec l'avènement de l'informatique et des langages spécialisés tels que le langage R.

1.2.1.2.2. Classification

Carl von Linné entre les années 1735 et 1758 fut le premier à classifier des données avec sa recherche sur la taxonomie [297].

Les algorithmes de classification se divisent en plusieurs catégories en fonction des modèles statistiques utilisés. Les méthodes les plus utilisés aujourd'hui sont les « *connectivity models* (ou Classification hiérarchique) », les « *centroid models* (ou K-means), les « *density models* (ou DBSCAN) » et les « *distribution models* (ou Algorithme espérance-maximisation) ». Cette dernière méthode fut développée en 1977 par Dempster [298]. La méthode DBSCAN fut développée en 1996 par 4 chercheurs (Martin Ester, Hans-Peter Kriegel, Jörg Sander et Xiaowei Xu) [299]. La méthode des K-means fut développée par Stuart P. Lloyd en 1957 mais fut citée pour la première fois en 1967 par James MacQueen [300]. Une méthode similaire fut publiée en 1965 par E. W. Forgy [301] (**Table 3**). Les méthodes énoncées précédemment sont les méthodes les plus utilisées mais il en existe d'autres basées notamment sur le clustering mixte tels que la Classification hiérarchique sur composantes principales

(HCPC) utilisée dans nos projets qui combinent les k-means et la classification hiérarchique ascendante [302].

Ces méthodes sont aujourd’hui très utilisées dans l’analyse de données à haut débit, et notamment par les méthodes d’apprentissage utilisées en intelligence artificielle.

Table 3 : Comparaison des méthodes de classifications les plus utilisées

Méthode	Catégorie	Avantage	Inconvénient
K-means	Méthode de partitionnement	Facilité d’utilisation et utilisable sur de grands jeux de données	Nombre de classe fixé au départ
Classification hiérarchique ascendante (CAH)	Méthode hiérarchique	Permet de déterminer le nombre optimal de classes	Temps de calcul élevé en cas de grands jeux de données
Classification hiérarchique sur composantes principales (HCPC)	Méthode de clustering mixte	Combine les deux dernières méthodes	
DBSCAN	Méthode basée sur la densité	Algorithme simple	Impossible de gérer des clusters avec des densités différentes

1.2.1.2.3. Le big data et l’intelligence artificielle

Le terme de science des données ou « data science » fut créé lors d’un colloque de statistiques en 1992. Quelques années plus tard en 1997, le terme de *big data* fut proposé ; il correspond à de grands ensembles de données quelles que soient leurs natures (**Figure 3**) [303].

Cette discipline répond à l’augmentation des bases de données disponibles et à la nécessité de développer de nouvelles méthodes pour leur analyse. Une majeure partie des méthodes utilisées pour l’analyse de ces données est basée sur l’intelligence artificielle.

L’intelligence artificielle ou IA englobe plusieurs méthodes très utilisées aujourd’hui tels que la classification, les réseaux de neurones ou encore *le deep learning*. Toutes ces notions sont très fréquemment utilisées aujourd’hui, notamment avec l’avènement des réseaux sociaux.

Les prémices de l’intelligence artificielle sont apparues dans les années 1950 via les travaux d’Alan Turing et le développement de son test sur la conscience d’une machine [304]. Durant les années 1980, le développement sur l’IA est important avec l’amélioration des dispositifs

informatiques ainsi qu'avec le développement des algorithmes d'apprentissage qui sont les ancêtres des réseaux de neurones utilisés aujourd'hui [305]. Durant les années 2000, l'augmentation massive de données informatiques et de la puissance de calcul entraînent de nouvelles méthodes d'apprentissage tels que l'apprentissage profond (ou *deep learning*) [306]. Parmi les méthodes de l'intelligence artificielle, deux méthodes d'apprentissage s'opposent selon que l'apprentissage est supervisé ou non. La première utilise un jeu de données déjà classifié et attribue la nouvelle variable à l'une des catégories, alors que la seconde utilise les variables pour regrouper les individus ayant de fortes similarités. Parmi les méthodes d'apprentissage non supervisées, les méthodes discutées précédemment tels que les k-means ou encore la classification hiérarchique sont utilisées. Le deep learning, lui, fait partie des méthodes d'apprentissage supervisées, et est très utilisé par exemple pour la reconnaissance faciale présente sur les téléphones portables de nouvelles générations.

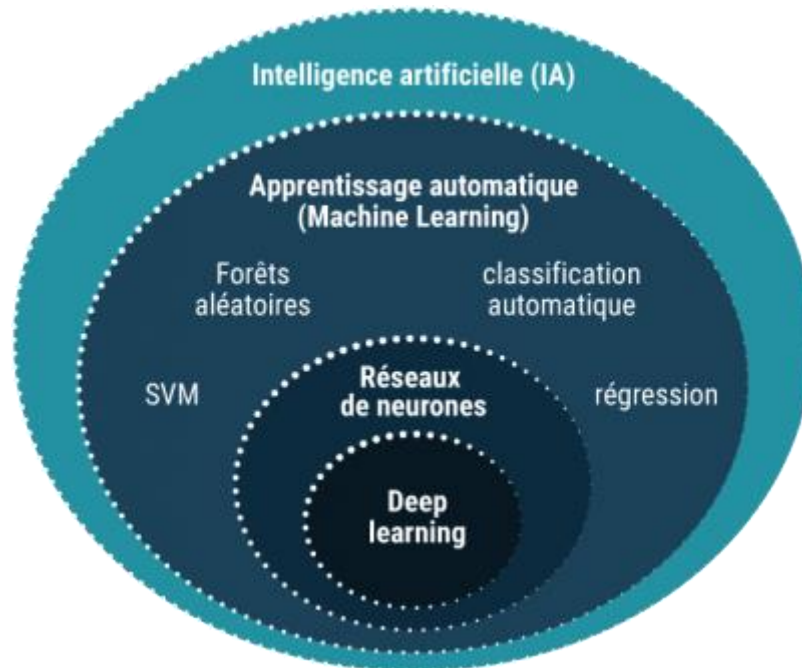


Figure 3 : Représentation schématique des relations entre les différents termes liés à l'intelligence artificielle (Datavaloris)

1.2.2. Les champs d'application

L'utilisation des méthodes biostatistiques est devenue nécessaire aujourd'hui dans l'étude de la biologie via l'utilisation de tests statistiques qui reposent sur les lois énoncées dans la partie précédente. Concernant, l'analyse de données à haut débit, elle devient de plus en plus présente en particulier à cause du développement des technologies dites *omiques*.

1.2.2.1. Les différents types de biostatistiques

La discipline des biostatistiques est retrouvée dans de nombreux domaines tels que la santé publique, l'épidémiologie, les essais cliniques ou encore la génétique. Cette liste n'est pas exhaustive mais il s'agit des domaines les plus concernés. Au sein cette discipline, deux types de biostatistiques existent avec l'analyse descriptive (graphique, résumé des données) et l'analyse explicative et le recours aux tests statistiques.

L'utilisation de représentations graphiques pour illustrer les données est très répandue dans le domaine de la biologie. De nombreux types de représentation sont utilisés dont les plus connues sont l'histogramme, les graphiques en barre (« barplots »), les boîtes à moustache (« boxplot »), etc. Ces types de graphiques sont plus compréhensible que des tables de données et peuvent apporter des informations supplémentaires sur la dynamique des données, avec l'affichage de la moyenne et des quartiles notamment. Les statistiques descriptives ont pour rôle d'étudier la distribution des données en utilisant des variables informatives tels que la moyenne, la médiane, la variance ou encore l'écart type. Les informations visuelles ou statistiques obtenues peuvent être résumées sous la forme d'une boîte à moustaches. Les statistiques descriptives constituent la première étape de l'analyse des données biologiques.

La seconde étape ou statistique explicative consiste à expliquer des variables. Cette discipline contient 2 ensembles de méthodes en fonction des données et de la question posée. Il existe des méthodes dites univariées qui sont utilisées lorsqu'une seule variable est étudiée ou multivariées dans le cas où il existe un ensemble de variables à expliquer ou explicatives. Les méthodes statistiques utilisées dans ces 2 catégories sont différentes.

Lors de l'analyse d'une seule variable, les tests statistiques sont relativement simples et nécessitent peu de conditions comme le test de Mann-Whitney par exemple. Il en va de même pour les représentations qui restent simples (*barplots*, *boxplots*).

En revanche, pour l'analyse d'un ensemble de variables, les méthodes graphiques et explicatives se complexifient. Les méthodes graphiques sont des représentations en composantes principales (ACP, ACM, AFM ...) ou des méthodes de partitionnement de données (Classification hiérarchique ascendante, K-means). Concernant les méthodes explicatives, ce sont celles qui sont très utilisées en science des données telles que l'analyse de régression linéaire ou logistique, et plus communément l'analyse de variance. Ces dernières font majoritairement partie des méthodes d'apprentissages non supervisées. Ces ensembles de méthodes nécessitent des compétences en statistiques et en informatique plus poussées car une

majorité de ces méthodes ne sont réalisables qu'avec des outils informatiques tels que ceux développés dans l'ensemble R.

Cependant, les méthodes utilisées en biologie varient selon les disciplines. Par exemple, dans le cadre particulier de la génomique, les méthodes sont la plupart du temps multivariées alors que pour les essais cliniques, elles peuvent être univariées ou multivariées selon le contexte.

1.2.2.2. Les champs d'application

Les exemples d'application des biostatistiques sont nombreux et variés, et les questions de recherche traitées le sont tout autant :

- L'une des utilisations concerne la détermination de la distribution d'une variable dans une population donnée afin de déterminer un intervalle des valeurs usuelles (dit intervalle de normalité) permettant de prédire en-deçà ou au-delà de quelles, le résultat de sa mesure peut être considéré anormal.
- La seconde est la comparaison et l'étude de la différence d'une variable entre conditions. Elles peuvent être réalisées entre deux ou un plus grand nombre de conditions.
- L'étude de la corrélation est très utilisée pour déterminer s'il existe un lien entre deux variables, et si celui-ci est robuste ou non.

L'ensemble de ces trois conditions d'utilisation peut s'appliquer à toutes les disciplines et correspond aux analyses les plus fréquemment réalisées.

Les domaines d'application sont très nombreux. Dans le cadre de la pharmacologie, les biostatistiques peuvent être utilisées pour déterminer l'action, tester le dosage ou déterminer le potentiel d'un médicament. C'est dans ce cadre que sont réalisés les essais cliniques avec l'utilisation d'un groupe placebo et d'un groupe traité.

L'utilisation de ces méthodes dans l'étude phylogénétique est un cas à part. En effet, la phylogénie utilise des algorithmes spécialement créés pour cette occasion et leur utilisation nécessite des compétences spécifiques que ce soit en statistiques ou en langages informatiques. De même pour la génétique, une étape préalable consiste à l'alignement de séquences génomiques vis-à-vis d'un génome de référence, puis à la détermination fonctionnelle des gènes différemment exprimés, ou encore l'étude d'un variant génomique au sein d'une population (génétique des populations). De même que pour la phylogénie, les algorithmes développés sont spécifiques de ces applications.

Pour les études épidémiologiques, l'utilisation des biostatistiques est indispensable et permet via des modèles statistiques complexes de déterminer l'étiologie, la progression ou encore

l'effet d'un traitement vis-à-vis d'une maladie ou affection donnée. Via d'autres méthodes, il est aussi possible de déterminer la causalité d'une ou plusieurs variables sur l'affection étudiée. Cela a notamment été le cas dans le cadre de l'infection via la COVID-19 avec la détermination de modèles prévisionnels.

Les biostatistiques interviennent dans d'autres champs thématiques que sont l'agriculture, les biotechnologies, ou encore l'infectiologie. Cette dernière concerne plus particulièrement les questions abordées au cours du Doctorat et sera détaillée dans la partie suivante.

L'avènement des biotechnologies à haut débit, et notamment la collecte de données omiques en grand nombre a entraîné l'émergence de méthodes statistiques poussées pour l'analyse de celles-ci. C'est une des raisons de l'utilisation de méthodes liées à l'intelligence artificielle dans le domaine de la biologie. Ces données représentent une quantité d'informations très grande jamais atteinte jusque-là. Durant mon travail de thèse, l'utilisation de diverses méthodes notamment pour l'analyse de données transcriptomiques ou de profilage cytokinique de la réponse sera détaillée dans les paragraphes suivants.

1.2.3. Démarche expérimentale et méthodes statistiques appliquées à la santé des bovins

1.2.3.1. Études épidémiologiques, dites « de terrain »

Les études épidémiologiques dites de terrain ont pour rôle l'étude d'une ou plusieurs variables dans une population donnée tels qu'un élevage, une filière d'élevage ou une aire géographique (**Figure 4**). Le but de ces études est de déterminer par exemple la présence d'un agent pathogène et sa circulation au sein de l'effectif, ou bien d'étudier l'effet d'un traitement prophylactique à grande échelle. L'intérêt de ces études est qu'elles correspondent aux conditions réelles de vie des animaux, et permettent de répondre à des enjeux appliqués au plus près des conditions d'apparition ou d'application des mesures de gestion. Dans le cas de l'espèce bovine, elles permettent d'étudier la circulation d'agents pathogènes à l'origine des maladies respiratoires par exemple. Dans ce cadre, les essais sont dits randomisés contrôlés. En effet, les animaux sont répartis aléatoirement dans divers groupes d'étude. Néanmoins, les groupes doivent être similaires pour des facteurs comme l'âge, la génétique ou divers critères de santé. Ces études sont réalisées au sein d'élevages commerciaux le plus souvent. L'accès à ces ressources animales peut donc être rendu difficile par l'adhésion plus ou moins grande des éleveurs à la question traitée. En outre, une grande variabilité des conditions entre élevages requiert de grands effectifs pour s'affranchir de cette variabilité et faire ressortir les effets étudiés. A

l'inverse, un avantage majeur de ces études est la disponibilité d'un grand nombre d'animaux par condition lorsque le recrutement et l'inclusion le permettent. Celui-ci permet de réaliser des analyses statistiques avec des conclusions souvent plus robustes dont le recours aux méthodes paramétriques, si la loi normale est respectée. La taille de l'effectif étudié permet généralement d'avoir des échantillons dont la distribution suit la loi normale et donc d'utiliser des tests statistiques tels que l'ANOVA (si plusieurs variables) ou le test de Student lors de l'étude d'une variable unique. L'utilisation de méthodes d'analyses multivariées est de plus en plus fréquente afin d'étudier les liens entre les variables (par exemple la réponse de l'hôte via la production de cytokines, celle des anticorps, et la quantité de pathogène excrétée) via des analyses de corrélation. Elles permettent aussi de regrouper les animaux en fonction de leur profil et de déterminer ou non la présence d'individus divergents. L'utilisation des méthodes en composantes principales permettent de résumer l'information contenue dans de nombreux paramètres. L'avantage de l'utilisation de ces méthodes est la possibilité de les coupler avec des méthodes de classification tels que la méthode HCPC par exemple, que nous avons utilisée dans un de nos projets. Via ces méthodes, l'analyse ne consiste plus à analyser chaque paramètre indépendamment, mais d'étudier un sous-ensemble de paramètres informatifs qui permettent de décrire des profils différents et ainsi de créer des catégories. L'analyse ainsi produite est plus globale.

L'objectif de ces méthodes analytiques est essentiellement descriptif, même si dans certains cas, les mécanismes (immunitaires dans notre cas) peuvent être envisagés en fonction d'une ou plusieurs variables (comme la vaccination, le fond génétique, ou l'âge par exemple).

1.2.3.2. Les études individu-centrées

Les études centrées sur l'individu ont pour objectif d'étudier des mécanismes ou d'évaluer l'effet de certaines interventions comme un traitement par exemple sur un nombre souvent beaucoup plus réduit d'animaux (**Figure 4**). Dans ce cadre, les paramètres peuvent être très contrôlés et la quantité de données collectées est souvent plus élevée car la fréquence de prélèvements est souvent beaucoup plus grande. En plus d'avoir la possibilité de répéter les prélèvements, la nature et la diversité des informations collectées sont souvent beaucoup plus grandes. Ces différences entre protocoles s'appliquent aussi à l'espèce bovine avec des prélèvements de sang, de lait, la réalisation possible de biopsies tissulaire qui offrent en outre l'avantage de questionner un même phénomène biologique de diverses façons. Par exemple dans le cadre de l'évaluation des effets d'une vaccination, outre les signes cliniques systémiques

et locaux qui sont enregistrés, divers prélèvements de sang pour le dosage cytokinique ou l'analyse de l'activité transcriptionnelle peuvent être réalisés. Dans ces études, l'utilisation de méthodes de collecte de données à haut débit (ensemble des techniques omiques) est de plus en plus fréquente grâce aux évolutions techniques et à une diminution du coût qui les rendent accessibles à ce type d'essai. Le recours à ces techniques est responsable du changement d'échelle de la quantité de données à exploiter et modifie en conséquence les méthodes statistiques utilisées dans ces travaux.

Le nombre d'animaux inclus dans ces études est généralement faible (quelques unités à quelques dizaines maximum) notamment dans l'espèce bovine à cause du coût de fourniture et d'entretien des animaux. C'est une limite de ce type d'étude dont le coût est souvent plus onéreux (fourniture, hébergement, personnel avec des compétences spécifiques, analyses) que celui des espèces de laboratoire.

Néanmoins, l'information apportée par ce type d'études est nécessaire si l'on veut s'intéresser aux mécanismes physiopathologiques ou thérapeutiques, en particulier dans le cadre général de l'infectiologie. Lors de l'analyse de variables individuelles, l'utilisation de tests paramétriques est moins appropriée du fait du faible nombre d'animaux et de la distribution dissymétrique des valeurs. L'utilisation de tests non paramétriques (Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman) est nécessaire. Néanmoins, les analyses de ce type sont remplacées de plus en plus par les analyses multivariées car la majorité du temps, les variables ne sont pas strictement indépendantes. Le recours à ces tests statistiques est nécessaire pour l'analyse des données de grande dimension dites omiques car ces données représentent des milliers de variables propres à un individu.

L'étude des mécanismes en infectiologie repose sur l'analyse de plusieurs compartiments ou strates biologiques (métabolites, protéines, ARN, ADN, phénotype) de l'expression des gènes et de leurs conséquences biologiques. Des méthodes de corrélation ont été développées afin d'étudier les liens entre ces différents jeux de données et afin de comprendre les liens mécanistiques entre les différents aspects d'une même question biologique.






	Etudes épidémiologiques	Etudes individus centrées
	Etude d'une variable dans une population ou territoire donnée	Etude d'un mécanisme d'une intervention ou d'un traitement
	Effectifs > 30	Effectifs < 30
	Conditions de vies réelles des animaux	Ensemble des paramètres contrôlés
	Accès aux prélèvements plus restreints	Fréquence et types de prélèvements plus élevés
	Cout faible (Analyses)	Cout élevé (Animaux, Entretien, Analyses)

Figure 4 : Les différences entre les études épidémiologiques et les études individus centrées

1.3. Méthodes statistiques appliquées aux données de grandes dimensions

1.3.1. *De la biologie moléculaire à la biologie multi-niveaux*

1.3.1.1. Développement de la biologie moléculaire

Dès l'année 1869, les travaux de Friedrich Miescher ont permis de découvrir que les chromosomes étaient formés de *nucléine* d'où vient plus tard les noms respectivement d'acide ribonucléique (ARN) et d'acide désoxyribonucléique (ADN) [307]. Ceux-ci sont des enchainements ordonnés de sucres et de résidus phosphate.

Les fondements de la biologie moléculaire reposent sur la découverte de la double hélice d'ADN par Watson et Crick en 1953 [308]. Elle est le résultat de différents travaux commencés plusieurs années auparavant. En effet, William Astbury fut le premier à étudier l'ADN en utilisant la radiocristallographie qui fut démocratisée en 1914 et découvre que l'ADN se présente sous la forme d'un long filament constitué d'une chaîne de constituants répétitifs espacés de 0,34 nm [309]. En 1949, deux biochimistes (James Norman Davidson et Erwin Chargaff) ont pu déterminer que les bases étaient présentes en quantités identiques deux à deux. En effet, la quantité de thymine était identique à celle de l'adénine ; et il en est de même pour la cytosine avec la guanine [310]. Dans les années 50, les premières images de l'ADN sont obtenues par Maurice Wilkins et Rosalind Franklin par diffraction aux rayons X [311]. Toutes ces études ont permis à Watson et Crick d'élaborer le modèle de la double hélice d'ADN. De plus, Crick émet l'hypothèse que le code génétique est composé de quatre bases [310].

L'ensemble de ces travaux a permis le développement des principes fondamentaux de la biologie moléculaire par Crick qui indique que l'ADN est le support de l'information génétique exprimé via les protéines et autres molécules biologiques.

Dans les années 1930, il fut démontré que la production protéique se déroulait dans le cytoplasme de la cellule qui était riche en ARN alors que l'ADN se situait dans le noyau. Plus tard, une étude a révélé que la quantité d'ARN et la synthèse de protéines étaient corrélées. Ces études furent confirmées lorsque François Jacob et Sydney Brenner ont montré que l'ARN se liait aux ribosomes pour la synthèse protéique. Mais ce n'est qu'en 1970 que tous les détails concernant les mécanismes complets de la transcription et de la traduction ont été révélés. Toutes ces études ont créé un fort intérêt pour la biologie moléculaire, et notamment l'étude du

génomique et de ses produits, entraînant le développement de technologies permettant l'étude de toutes ces composés moléculaires (ARN, ADN et protéines) comme les blots.

1.3.1.2. Le développement des technologies à haut débit

Afin d'étudier les mécanismes biologiques dans divers contextes, la connaissance et l'analyse de l'ARN et de sa traduction, ainsi que celle des protéines furent décisifs, et ont entraîné le développement de méthodes permettant de connaître précisément la quantité et la nature des produits de l'expression des gènes. En 1990, les technologies à haut débit ont émergé permettant une analyse plus globale et plus détaillée des processus biologiques.

L'une des premières techniques à être développée correspond à l'étude de l'ADN, et globalement désignée sous le terme de génomique. Cette technologie a d'abord pour but de déterminer la séquence (par différentes techniques de séquençage), puis de caractériser ces séquences qui peuvent correspondre à des gènes, des régions régulatrices ou des régions non codantes dont le rôle n'est pas toujours précisément déterminé à la suite d'un travail long et complexe d'annotation. Ce domaine s'intéresse aussi à l'effet de la variabilité de séquence sur les caractères phénotypiques et vice versa. Les études de séquençage ont débuté avant les années 90 avec en moyenne 1200 bases analysées par essai, puis entre les années 1990 et 2000, avec environ 77000 bases et une augmentation considérable à partir des années 2000, avec plus de 200 milliards de bases analysées par essai. Après les années 2010, il devient possible d'avoir accès à la séquence d'un génome procaryote, mais aussi eucaryote tout entier. En parallèle de cette augmentation du nombre de bases séquencées, le prix a fortement diminué passant de 12 bases pour un dollar à 6,5 millions de bases pour un dollar dans les années 2010. Aujourd'hui, il est possible de connaître la séquence d'un génome pour moins de 1000 dollars. Les premiers génomes entièrement séquencés furent ceux de virus puis de bactéries en 1995. Le séquençage complet du génome humain fut terminé en 2001 après 15 ans de travaux grâce au projet *Génome Humain*. Dans les années qui ont suivi, les génomes d'autres espèces ont été entièrement séquencés comme le porc (2003), la vache (2009), la brebis (2010) ou encore la chèvre (2016). Depuis, des milliers de génomes ont été entièrement séquencés. Ces données sont accompagnées par l'annotation des gènes qui est un processus souvent long et itératif. Le problème actuel réside dans l'homogénéisation des annotations entre les différentes espèces. Le séquençage a permis le développement de l'étude des variants de séquence de l'ADN et de leur corrélation avec la variabilité phénotypique entre les espèces animales, et entre les individus d'une même espèce.

La transcriptomique correspond à l'étude des ARNm et donc de l'expression des gènes codant les protéines. Ceux-ci représentent à peine 5% du génome. L'expression des gènes varie selon le tissu, le type cellulaire, le contexte physiologique. De ce fait, de nombreuses bases de données ont été développées pour héberger ces informations dont la masse croissante devient considérable. De nombreuses méthodes ont été mises au point pour étudier le transcriptome, comme par exemple la PCR quantitative (ou Q-PCR) permettant de déterminer l'expression d'un seul gène, les puces à ADN (ou microarrays) permettant l'étude de milliers de gènes en parallèle, puis les méthodes fondées sur le séquençage comme le RNA-seq permettant l'étude de la quasi-totalité des gènes exprimés dans un échantillon. Ces technologies seront détaillées dans les parties suivantes.

Il existe d'autres méthodes omiques pour les autres composés biologiques tels que la protéomique pour les protéines, la métabolomique pour les métabolites de nature lipidique, oligosaccharidique ou minérale.

La protéomique est un champ qui s'est beaucoup développé depuis 1997 et dont l'objectif est de connaître la nature et l'abondance des protéines d'une cellule ou d'un tissu. Elle permet d'identifier et de quantifier l'ensemble des protéines visant l'exhaustivité dans un contexte biologique donné. Le protéome ainsi établi correspond à des milliers de protéines comme résultat de l'expression des gènes codants et de l'épissage alternatif ainsi que les modifications post-traductionnelles. Cette étude des protéines à haut débit a pour objectif la compréhension des processus biologiques, la recherche de biomarqueurs ou de cibles thérapeutiques. De nouvelles approches résultent de la connaissance de la composition protéique comme l'interactomique. Ce champ est un domaine omique parmi les plus récents et permet l'étude des liens entre les molécules. L'interaction la plus évidente se fait entre protéines mais il existe aussi des interactions entre ADN et protéines. Ces interdépendances peuvent être illustrées à l'aide de réseaux.

La métabolomique correspond à la découverte, l'identification systématique, ou encore la classification des molécules dont le poids moléculaire est inférieur à 1500 Da dans une cellule ou un échantillon tissulaire. Les métabolites correspondent à des macromolécules de la famille des acides aminés, lipides, protéines, vitamines, produits du métabolisme, ...

Les études liées à la génomique ont fortement augmenté entre le début des années 2000 (30000) et l'année 2015 (70000), puis une diminution s'est produite entre 2015 et 2018 au bénéfice des méthodes suivantes (transcriptomique, protéomique, métabolomique) dont le nombre d'études a évolué moins vite au début des années 2000, mais qui est toujours en augmentation

aujourd'hui [312]. L'étude globale de toutes ces données à haut débit a permis le développement du champ de la biologie des systèmes. C'est un domaine qui étudie les organismes comme des systèmes à part entière en cherchant à intégrer différents niveaux biologiques tels que l'ADN, l'ARN, les protéines et les métabolites. Le développement de ces nouvelles méthodes est intrinsèquement lié aux développements de la bioinformatique et de la biostatistique pour l'étude de ces données souvent multi-dimensionnelles et de grande échelle. La bioinformatique est un champ multidisciplinaire créé en 1970 permettant de résoudre des problèmes biologiques via l'utilisation de l'informatique. Elle a de nombreux domaines d'application tels que l'étude des séquences, l'analyse structurale pour l'étude en 3D des protéines ainsi que l'étude des réseaux très utilisée pour l'exploitation des données, comme indiqué précédemment. Dans le cadre de cette thèse, l'analyse des séquences et l'étude via des réseaux sont les applications majeures auxquelles nous avons eu recours. La bioinformatique est indissociable de la biostatistique car de nombreux algorithmes ou outils bioinformatiques utilisent les principes et les lois statistiques.

1.3.1.3. Le développement de la biologie computationnelle

La biologie computationnelle est une biologie intégrant le développement et l'application de méthodes d'analyse pour l'étude des données biologiques. Cette discipline est au carrefour de la bioinformatique, de la biostatistique et de la biologie des systèmes.

Les débuts de ce domaine de la biologie datent de l'origine de l'informatique. En effet, Alan Turing fut le premier à utiliser l'informatique pour implémenter un modèle de morphogenèse au début des années 1950. Dans les années 1960, l'utilisation de l'informatique pour l'étude des structures protéiques a permis le développement de la biologie computationnelle. Durant les mêmes années, la phylogénie, science étudiant l'évolution des organismes a elle-aussi utilisé l'informatique pour la classification des organismes en utilisant des algorithmes de similarité. Le développement des technologies permettant l'étude des séquences protéiques et nucléotidiques a permis le développement de nouvelles méthodes au sein de la biologie computationnelle. Au début des années 1980, le développement de cette discipline fut important avec l'arrivée de l'intelligence artificielle. Celle-ci a permis le développement de la codification des connaissances biologiques sous une forme utilisable par l'informatique. Cette biologie a été divisée en plusieurs champs thématiques tels que la phylogénie, l'étude protéique et surtout l'étude génomique incluant la transcriptomique.

Elle est devenue depuis la fin des années 90 une part importante du développement et de l'émergence des nouvelles technologies dans le domaine de la biologie. L'analyse fonctionnelle est maintenant très en vogue car elle permet d'accéder à la connaissance des mécanismes biologiques en exploitant des données qui sont par ailleurs peu accessibles au raisonnement humain. L'étude de la relation entre les différents jeux de données omiques a entraîné un fort développement de la biologie computationnelle. Une majeure partie de ces outils bioinformatiques sont gratuits et sont disponibles via l'ensemble R. En 2004, une base de données recensant ces outils a été développée ; Baptisée *Bioconductor*, elle permet un accès facilité à l'ensemble de ces ressources [313]. Les logiciels développés aujourd'hui sont pour la plupart développés via des méthodes d'intelligence artificielle (IA) pour l'utilisation d'algorithmes plus complexes et demandant des temps et des capacités de calcul toujours plus élevés ; ils permettent ainsi des études de corrélation de plus en plus poussées et un accès croissant aux informations et à la compréhension des phénomènes biologiques.

1.3.2. Les différents types de données omiques collectées à haut-débit

1.3.2.1. Les données transcriptomiques

La transcriptomique est l'étude des produits issus de l'analyse du transcriptome, c'est-à-dire de l'expression des gènes sous la forme des ARN. Les ARN sont produits par transcription de l'ADN et sont des molécules constituées de 4 bases (– uracile remplaçant la thymine).

Les ARN ont de nombreuses fonctions liées à leur structure. Ils se divisent en quatre familles

- ARNm (ARN messager) : molécule permettant la synthèse des protéines.
- ARNr (ARN ribosomique) : la composition et l'activité des ribosomes
- ARNt (ARN transfert) : incorporation des acides aminés dans les protéines
- siRNA (ARN interférent) ou miRNA (microRNA): régule l'expression des gènes en ciblant les ARNm pour la dégradation ou en inhibant la traduction de la protéine cible.

Les ARN messagers correspondant au support d'information utilisé par les analyses transcriptomiques. En effet les ARNm correspondent à l'information utilisée pour la traduction et la synthèse des protéines, et leur étude permet donc l'étude dynamique des processus biologiques.

Les dot blots sont un ensemble de méthodes qui ont été parmi les premières à permettre l'analyse transcriptionnelle. Le Southern blot a été développé en 1975 pour l'étude de l'ADN, le northern blot en 1977 pour l'ARN et enfin le western blot en 1981 pour les protéines. Une

autre méthode plus tardive fut développée pour l'étude des lipides connu sous le nom de Far eastern blot en 1994.

Le northern blot est une méthode permettant la détection de fragments d'ARN spécifiques au sein d'un ensemble d'ARN. Cette technologie permet d'analyser un échantillon ARN issu d'un tissu ou type cellulaire particulier afin de mesurer l'expression des transcrits. Cette technologie permet d'étudier la taille, l'abondance relative, la dégradation et l'épissage de l'ARN dans un échantillon donné. Les différentes étapes sont d'abord la purification de l'ARN des échantillons, puis une séparation. Une électrophorèse est réalisée afin de séparer les ARN en fonction de leur taille. Après transfert sur une membrane, une hybridation avec une sonde spécifique de l'ARN à étudier est produite afin de détecter celui-ci. Les produits utilisés par cette méthode sont parfois dangereux pour la santé, parce que radioactifs, ce qui en fait un inconvénient majeur. Cette technologie a maintenant été supplantée par la qPCR depuis les années 90. Cependant le principe de cette technologie a été à l'origine du développement des puces avec sondes nucléotidiques (ou microarrays pour ceux dont la densité est la plus élevée) qui ont constitué un tournant dans le changement d'échelle de l'analyse à haut débit du transcriptome.

Les puces à ADN (microarrays) correspondent à une technologie développée en 1991 permettant l'étude quantitative de plusieurs milliers de gènes en parallèle dans un échantillon donné. Cette méthode est basée sur le phénomène d'hybridation, selon lequel deux fragments d'acides nucléiques complémentaires peuvent s'associer, mais aussi se dissocier. Pour cela, des sondes correspondant à des séquences nucléotidiques spécifiques d'un gène sont fixés sur un support, le plus souvent une lame de verre. Celui-ci va être mis en contact avec le mélange d'ARN purifiés, puis une fois hybridé la quantification de chaque type d'ARN comme produit de l'expression d'un gène est réalisé par l'utilisation de la fluorescence qui a été introduite au moment de la transcription inverse. Une étape d'amplification peut être réalisée au préalable si la quantité d'ARN n'est pas suffisante. Ces puces peuvent être utilisées avec hybridation d'un seul échantillon ou une hybridation compétitive de deux échantillons permettant la comparaison du niveau d'expression entre les deux conditions. Les sondes peuvent être des oligonucléotides comme indiqué précédemment ; ils peuvent aussi permettre la détection du polymorphisme de séquence et la présence de variants ou d'épissage alternatif. L'analyse de ces données peut être réalisée à l'aide de différentes suites logicielles spécialisées comme Limma, certaines étant disponibles dans l'ensemble R. Cette méthode peut être considérée comme la première approche réussie d'analyse à haut débit.

La PCR ou réaction de polymérisation en chaîne est une technique développée en 1985 permettant l'amplification de fragments d'ADN jusqu'à ce que leur concentration devienne détectable. La PCR quantitative ou Q-PCR est une évolution de cette PCR, inventée en 1992, permettant l'étude quantitative directe de la quantité de cible, et donc de l'expression génique. Elle est basée sur le même principe d'amplification de fragments associée à une quantification de la quantité d'ADN accumulé en continu. Pour cela, la technologie utilise la fluorescence sur une ou plusieurs sondes capables de se fixer spécifiquement sur le fragment d'ADN à quantifier. La quantification est obtenue via le Ct ou cycle seuil qui est le cycle à partir duquel la quantité d'ADN devient détectable car supérieure à un seuil établi *a posteriori* par l'utilisateur ou par un logiciel d'analyse. La quantification de séquences d'ARN est possible après une étape de transcription inverse. Néanmoins, cette méthode a plusieurs limites, liées au fait que l'amplification est exponentielle et peut connaître des dérives. Par ailleurs, la quantité de transcrits ne correspond pas toujours à la quantité de protéine biologiquement active ; ce phénomène est bien décrit pour les gènes codant des cytokines. Des technologies moyen débit ont été développées tels que le système Fluidigm. Ils utilisent ce principe via des puces permettant l'analyse d'un nombre préalablement fixé de gènes et d'échantillons avec des combinaisons 48x48 ou 96x96. Néanmoins, cette technologie est de moins en moins utilisée du fait qu'il s'agit d'une méthode de débit relativement faible et qu'elle est surpassée par le séquençage de l'ARN en termes d'informations apportées et rapport informations / coût.

La PCR digitale utilise le même système de base que la PCR quantitative classique. Néanmoins, l'échantillon est divisé en de nombreux fragments et chacun de ces échantillons va subir une amplification PCR en parallèle au sein de chambres. La précision est améliorée avec une quantification absolue en nombre de copies et non plus en nombre de cycles. Elle possède une sensibilité plus grande que la PCR quantitative classique. Cette méthode ne nécessite pas de courbe étalon. La limite de cette méthode concerne le coût.

Nanostring est une technologie basée sur la quantification d'ARN. Elle permet de quantifier des centaines de transcrits au sein d'une seule réaction et ne nécessite pas d'amplification via une étape de PCR. L'absence de cette dernière permet une réduction des biais. Comparée aux méthodes à haut débit, cette méthode peut être réalisée sur des quantités d'ARN moindre et sur des ARN de qualité plus faible. Néanmoins, une des limites concernent le nombre de gènes pouvant être quantifiés.

Les méthodes d'acquisition de données transcriptomiques suivantes sont très utilisées actuellement et reposent sur la mesure de l'ARNm comme information. Elles sont issues du

développement des méthodes de séquençage. Ces analyses sont dites sans *a priori* car l'ensemble des séquences présentes dans un échantillon, au contraire des méthodes décrites précédemment, est séquençé et donc disponible pour l'analyse. Les technologies permettant ces analyses ont fortement évolué durant les dix dernières années.

La technologie de séquençage de l'ARN ou RNA-seq pour (RNA sequencing) développée dans les années 2000 fut l'une des premières méthodes utilisant le séquençage haut débit afin de quantifier et d'étudier le transcriptome. Une fois les ARNm purifiés, ils sont rétro-transcrits, puis fractionnés en fragments très courts avant d'être séquençés. Comparée aux puces à ADN, cette méthode a une sensibilité plus élevée (10 à 100 fois plus grande), ainsi qu'une résolution plus grande à l'échelle de la base. Cette technologie permet en outre l'étude de l'épissage et des différents transcrits produits par la transcription d'un gène donné. Néanmoins cette méthode ne permet que l'étude des ARNm codants matures qui sont poly-adénylés et épissés. La profondeur de séquençage est une notion fondamentale correspondant au nombre de fragments lus (reads) à chaque position. Plus le nombre de reads est grand, plus la sensibilité et la robustesse de la mesure sont grandes. Le coût à l'échantillon est de l'ordre de 2-300 euros, ce qui est un prix très bas par rapport à la quantité de données recueillies. Cette méthode présente des limites pour l'analyse fine des processus biologiques. En effet, appliquée à un tissu, l'expression mesurée est globale pour tous les types cellulaires présents dans le tissu ; il n'est donc pas possible de déterminer *a priori* quel type cellulaire est concerné par le changement d'expression. Pour pallier ce problème, une nouvelle technique dérivée a été développée, qui mesure l'expression à l'échelle de la cellule isolée : scRNA seq (pour single cell RNA-seq). Cette technique permet l'analyse du transcriptome à l'échelle de chaque cellule. Elle permet également d'étudier des différences entre cellules avec une sensibilité plus grande qu'auparavant. L'avantage de cette technique est l'étude de cellules rares ou de changements de la composition cellulaire ou du transcriptome dans des types cellulaires spécifiques. Néanmoins, cette technologie est plus lourde à mettre en œuvre à cause de la difficulté à séquençer une quantité infime d'ARN par cellule. Les ARN de chaque cellule sont identifiés grâce à l'ajout d'une séquence spécifique qui joue le rôle de code barre aux transcrits qu'elle contient. Les cellules doivent être préalablement isolées par tri ou séparation physique par cytométrie en flux ou dilution limitante. La méthode la plus utilisée pour la préparation des cellules est une technique microfluidique développée par la société 10xGenomics. Chaque cellule est isolée dans une gouttelette lipidique contenant les enzymes permettant l'amplification ainsi qu'une bille contenant le code barre unique, comme indiqué plus haut. Après la lyse des cellules, la bille va capturer les ARN messagers puis la totalité de l'ARNm de la cellule est transformée en ADNc avec insertion du

code-barre. La préparation des bibliothèques, le séquençage ainsi que l'analyse bioinformatique sont très similaires à ceux utilisés pour le RNA seq. En revanche, la partie d'analyse biostatistique est plus complexe et nécessite des outils dédiés telles que la suite logicielle *Seurat* [314].

De nombreuses suites logicielles ont été développées pour l'analyse bioinformatique et biostatistique de ces essais. Pour obtenir une sensibilité plus élevée, cette technologie peut être couplée à des méthodes de tri cellulaire telle que la cytométrie en flux permettant de sélectionner certaines catégories de cellules via des marqueurs de surface. La principale limite de cette technique est l'absence de spatialisation de l'analyse car au moment de la dissociation du tissu, les cellules sont mélangées les unes aux autres. Une technique permettant d'accéder à l'information sur la localisation des cellules et les changements régionaux de l'expression génique a été développée : il s'agit de la transcriptomique spatiale.

La transcriptomique spatiale dont le développement est très récent puisqu'il a débuté en 2016 englobe plusieurs méthodes dont la différence majeure est la sensibilité obtenue. En 2016, la résolution était encore faible avec des spots espacés de 200 μm . En 2019, la résolution est quasiment à l'échelle cellulaire (10 μm) et prend le nom de slide-seq. La technologie est basée sur un « palet » d'une taille de 3 millimètres et contenant des milliers de billes de 10 μm . Par palet, il y a environ 70000 spots analysés. Chacune de ces billes est identifiée par un oligonucléotide spécifique permettant d'enregistrer sa localisation. Une coupe de tissu d'une taille maximum de 10 μm d'épaisseur va être mis en contact du palet, les ARNm vont alors être capturés par les billes. Lors de l'analyse, il est possible de rapprocher la structure du tissu et le profil d'expression qui a été obtenu pour chaque spot, avec une résolution spatiale de l'expression génique jamais égalée. Cette méthode a d'abord été appliquée au cerveau de souris permettant de décrire les types cellulaires de chaque région morphologique et l'expression de gènes spécifiques [315]. Cette technique permet d'étudier l'expression de gènes ainsi que leur localisation ; elle est particulièrement utilisée dans les domaines d'étude du cancer et des processus infectieux.

D'autres méthodes existent telles que MERFISH et SeqFISH [316]. La différence majeure concerne la méthode d'encodage qui est basée sur des billes avec codes-barres pour le slide-seq et sur de l'hybridation *in situ* pour les méthodes MERFISH et seqFISH [317]. Du fait de son développement très récent, cet ensemble de techniques a encore des limites comme l'identification précise des types cellulaires qui n'est pas très précise, et l'analyse qui requiert des compétences bioinformatiques et biostatistiques plus poussées que les techniques d'analyse du transcriptome précédemment présentées. De plus, pour contourner le problème de

l'identification cellulaire, des équipes ont eu l'idée d'associer les approches de transcriptomique spatiale et RNA seq sur cellules isolées [318]. Cette association augmente encore la complexité de l'analyse des données, de l'étude et de la compréhension des phénomènes biologiques ainsi décrits.

Le développement des méthodes permettant l'analyse du transcriptome est de plus en plus rapide et les techniques sont de plus en plus sophistiquées. Les protocoles de préparation d'échantillons se complexifient, mais c'est la partie analytique qui nécessite de plus en plus de compétences spécifiques par des biologistes et des équipes spécialisées pour le traitement des données.

1.3.2.2. Analyse des médiateurs immunitaires, les cytokines

1.3.2.2.1. Qu'est-ce qu'une cytokine ?

Les cytokines correspondent à un ensemble de protéines jouant un rôle dans la signalisation entre cellules, de la réponse immunitaire notamment. Dans le cadre de l'étude de la réponse immunitaire, ces protéines jouent un rôle essentiel et nous permettent d'étudier certaines populations cellulaires responsables de leur sécrétion dans un contexte donné. Elles sont produites par divers types cellulaires en réponse à un signal activateur comme une infection par exemple. La reconnaissance de ces molécules par les cellules permet une modification de l'activité cellulaire, comme la prolifération, la mort apoptotique, ou encore la production de nouvelles molécules, y compris des cytokines. Les cytokines considérées jouent donc un rôle majeur lors d'infections ou à la suite d'une réponse auto-immune, mais elles jouent aussi des rôles dans d'autres processus biologiques tels que l'hématopoïèse ou la reproduction.

Les cytokines sont divisées en plusieurs familles que sont les Interférons, les Interleukines, les chimiokines et un autre regroupement contenant des molécules tels que le facteur de nécrose tumorale (**Table 4**).

Les Interleukines correspondent au principal ensemble de cytokines. Il en existe une quarantaine intervenant dans l'ensemble des fonctions du système immunitaire. Il existe des sous-familles d'interleukines en fonction de leur ressemblance fonctionnelles et structurales. Le rôle principal est de servir de médiateurs entre cellules immunitaires mais aussi non immunitaires.

Les Interférons sont des protéines appartenant à la famille des cytokines qui sont produites par les cellules immunitaires et certaines cellules non immunitaires tels que les cellules épithéliales. Il existe 3 types d'interférons ayant toutes un rôle majeur vis-à-vis des infections virales.

L'interferon de type II est produit par les lymphocytes (Innate lymphoid cells (ILC) ou lymphocytes T) alors que les interférons de type I et III sont produits par des cellules infectées par des virus. Ces molécules permettent l'activation des cellules immunitaires innées de type macrophage ou cellules NK.

Les facteurs de nécrose tumorale (TNF) appartiennent à la famille des facteurs de croissance. En effet, ils permettent la croissance des lymphocytes et des fibroblastes. Le membre principal de cette famille est le TNF- α qui est produit en majorité par les macrophages. Cette molécule est majoritairement retrouvée dans des contextes inflammatoires, et permet le recrutement et l'activation des lymphocytes et des macrophages.

Etant donné leur rôle central dans de nombreux processus biologiques, de nombreuses méthodes ont été développées pour détecter et connaître précisément la concentration de ces molécules. Par exemple, le dosage des cytokines IL-1 α , IL-1 β , IL-6 ou TNF α permet de déterminer la présence d'une réponse inflammatoire, par exemple à la suite d'une infection.

Les chimiokines correspondent à des cytokines ayant un rôle de chimioattractant. On dénombre une cinquantaine de chimiokines réparties en plusieurs familles en fonction de la présence ou non d'un acide aminé entre les 2 cystéines présent en position N-terminale. Ces molécules vont se fixer sur des récepteurs couplés aux protéines G présents à la surface des cellules immunitaires notamment et permettre leur recrutement.

Table 4 : Caractéristiques des cytokines et des chimiokines les plus connues

	Famille	Production	Rôle
IL-1 α	Interleukine	Nombreux types cellulaires	Méiateur inflammatoire
IL-1 β	Interleukine	Nombreux types cellulaires	Méiateur inflammatoire
IL-1RA	Interleukine	Leukocytes, cellules épithéliales	Activité anti-inflammatoire
IL-6	Interleukine	Leukocytes et cellules stromales	Méiateur inflammatoire
IL-2	Interleukine	Lymphocytes T	Stimule les lymphocytes
IL-4	Interleukine	Lymphocytes T, cellules stromales et mastocytes	Favorise la croissance des lymphocytes
IL-10	Interleukine	Lymphocytes T activés	Activité anti-inflammatoire
IL-12	Interleukine	Macrophages, cellules dendritiques	Activation cellules NK, lymphocytes T CD4 Th1
IL-17A	Interleukine	Lymphocytes T CD4 Th17	Activation des neutrophiles via les cellules épithéliales
IFN α	Interferon	Leukocytes	Activité anti-virale
IFN β	Interferon	Fibroblastes, cellules dendritiques	Activité anti-virale
IFN γ	Interferon	Lymphocytes T CD4 Th1, ILC et NK	Activité antivirale et anti-bactérienne
CXCL8	Chimiokine	Nombreux types cellulaires	Activité chimioattractante
CXCL10	Chimiokine	Monocytes, cellules endothéliales	Activité chimioattractante des leukocytes
CCL2 (MCP-1)	Chimiokine	Cellules myéloïdes	Activité chimioattractante des leukocytes
CCL3 (MIP1 α)	Chimiokine	Cellules polynucléaires	Activité chimioattractante des cellules polynucléaires
CCL4 (MIP1 β)	Chimiokine	Nombreux types cellulaires	Activité chimioattractante des leukocytes
TNF α	Facteurs de croissance	Leukocytes et fibroblastes	Méiateur inflammatoire
TGF β	Facteurs de croissance	Nombreux types cellulaires	Inhibe la croissance

1.3.2.2.2. Méthodes de dosage des cytokines

Le dosage des cytokines peut être réalisé par la méthode immuno-enzymatique ou ELISA (pour *Enzyme linked immunosorbent assay*). Cette méthode peut être utilisée pour des dosages quantitatifs avec l'utilisation d'une référence externe comme une gamme étalon, qualitatifs ou semi-quantitatifs via la mesure de l'expression par rapport à un témoin positif. Cette technique peut être utilisée sur de nombreux types d'échantillons biologiques dont le plus utilisé est le plasma, mais elle peut également être réalisée sur des surnageants de culture après activation des cellules ou sur des broyats de tissu. Cette technique est très spécifique et très sensible car elle permet de détecter, voire quantifier de faibles quantités de protéines, de l'ordre du picogramme. Le principe de l'ELISA est basé sur la colorimétrie via une réaction enzymatique. L'enzyme (peroxydase) a été couplé à l'anticorps utilisé pour la détection de la molécule à quantifier, qui une fois en présence de son substrat va entraîner une réaction colorée. L'ELISA peut être réalisé de multiples façons en fonction de l'analyte recherché. Dans le cas du dosage d'une cytokine, l'ELISA est de type sandwich ; le principe consiste à fixer un anticorps spécifique de la molécule sur le support comme une plaque en polystyrène, puis après avoir déposé l'échantillon contenant la molécule à capturer, l'ajout d'un deuxième anticorps couplé à la peroxydase permet la quantification de la molécule avec l'utilisation d'une gamme étalon en parallèle. Les avantages de cette méthode sont la grande spécificité car si des anticorps monoclonaux sont utilisés, la réaction est très spécifique de la molécule, et elle peut être détectée dans un échantillon de composition complexe.

La technique ELISA est facile à mettre en œuvre. Elle est simple et rapide à réaliser. Néanmoins, elle présente certains inconvénients comme le faible débit et le fait qu'un seul analyte est dosé à la fois ; le volume nécessaire pour le dosage peut être une contrainte dans le cas où de multiples dosages doivent être réalisés pour quantifier divers analytes.

Plusieurs techniques alternatives ont été développées pour résoudre ces difficultés. C'est le cas notamment des méthodes Mésoscale et Luminex par exemple qui permettent l'analyse de plusieurs analytes en parallèle (multiplexage) et une augmentation sensible du débit analytique.

La technique Mésoscale est dérivée de la technique ELISA, mais elle permet de quantifier jusqu'à 10 molécules en simultanée. De plus, à la différence de l'ELISA, la révélation est basée sur l'électrochimiluminescence permettant une forte sensibilité allant jusqu'au picogramme. Les analytes vont être reconnus par des anticorps spécifiques fixés sur le fond du puits. Tout comme un ELISA, la reconnaissance va se faire via l'utilisation d'un anticorps secondaire

présentant un tag Sulfo. Un champ électrique va être appliqué sur la plaque permettant une émission de lumière par l'anticorps secondaire et ainsi la quantification.

La méthode Luminex fondée sur les billes et la cytométrie, développée entre autres par la société Luminex, permet d'accroître le débit et la sensibilité de détection. Celle-ci est basée sur la détection simultanée de deux signaux de fluorescence et l'utilisation de billes. Chacune de ces billes est couplée à un anticorps spécifique de l'analyte à quantifier et possède une fluorescence spécifique qui va être détectée par l'appareil de mesure. Un anticorps de détection va être ajouté pour détecter l'analyte capturé par les billes. L'appareil détecte alors la fluorescence de la bille, permettant d'identifier l'analyte analysé et l'anticorps secondaire pour quantifier la quantité d'analyte capturé par les billes. De fait, cette technique est adaptée au multiplexage, où plusieurs analytes peuvent être quantifiés dans le même volume ou puits. L'appareil permettant la détection des signaux émis par les billes est basée sur la technologie de cytométrie en flux. Ainsi, il est possible de doser des dizaines d'analytes dans un laps de temps très réduit (**Figure 5**).

Le développement de trousse dédiées est en constante évolution et touche différents domaines de la biologie (immunologie, oncologie, neurosciences ou encore métabolisme) et de nombreuses espèces (Humain, Souris ou Bovin). Au sein du laboratoire, un kit a été développé en collaboration avec Merck-Millipore pour l'espèce bovine. Cette trousse permet la détection de 15 cytokines permettant l'étude de la réponse immunitaire innée (5 cytokines), adaptative (5 cytokines) ou du recrutement cellulaire (5 chimiokines). La complexité du développement de ce type de trousse concerne principalement la dynamique des standards externes car tous les couples anticorps/analytes n'interagissent pas de la même façon (affinité, avidité), et des interactions entre analytes sont possibles et doivent être maîtrisées. Ainsi, il est possible de mesurer la concentration d'une centaine d'analytes en parallèle, même si la plupart des trousse se limitent à une vingtaine, voire une quarantaine de cibles tout au plus. Le format de dosage varie avec l'utilisation de microplaques contenant 96 ou 384 puits.

Cette méthode nécessite de faibles volumes (25 µl maximum en l'absence de dilution) et elle est très répétable et sensible. Il est possible de sélectionner les analytes pour créer des trousse à façon. Les données collectées par l'appareil sont de 2 types soit une valeur de fluorescence moyenne sur les billes ou MFI (pour Mean Fluorescence Intensity), soit une valeur de concentration après conversion des valeurs de fluorescence à l'aide de la référence externe. Les 2 mesures peuvent être utilisées pour l'analyse avec pour chacune, des avantages et des

inconvenients qui seront examinés dans la partie concernant les méthodes d'analyse de ce type de données.

Ces deux dernières méthodes affichent des résultats similaires avec une sensibilité accrue pour la méthode Luminex [319], [320].

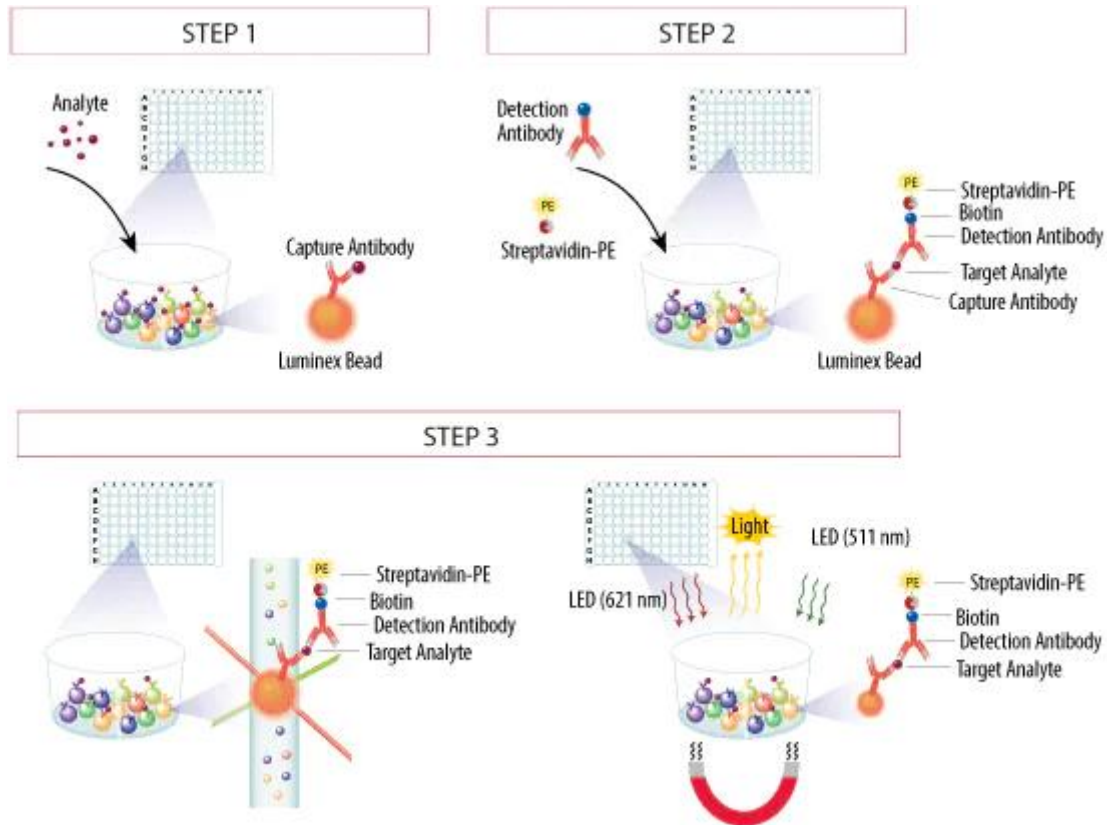


Figure 5 : Protocole détaillé de la technologie Luminex (R&D Systems)

1.3.2.3. Bilan

Comme nous l'avons vu, le système biologique fonctionne grâce à l'utilisation de plusieurs supports de l'information. Le premier support est l'ADN et correspond au sujet d'étude de la génomique. L'objectif est d'étudier les séquences d'ADN. La transcriptomique correspond à l'étude des ARN messagers issus du génome et qui donneront les protéines, sujet d'étude de la protéomique. La dernière grande discipline correspond à la métabolomique qui a pour objectif d'étudier les liens entre ces protéines et les divers sucres et lipides présents dans une cellule. Toutes ces informations sont liées entre elles et permettent d'étudier des mécanismes via plusieurs approches (**Figure 6**).

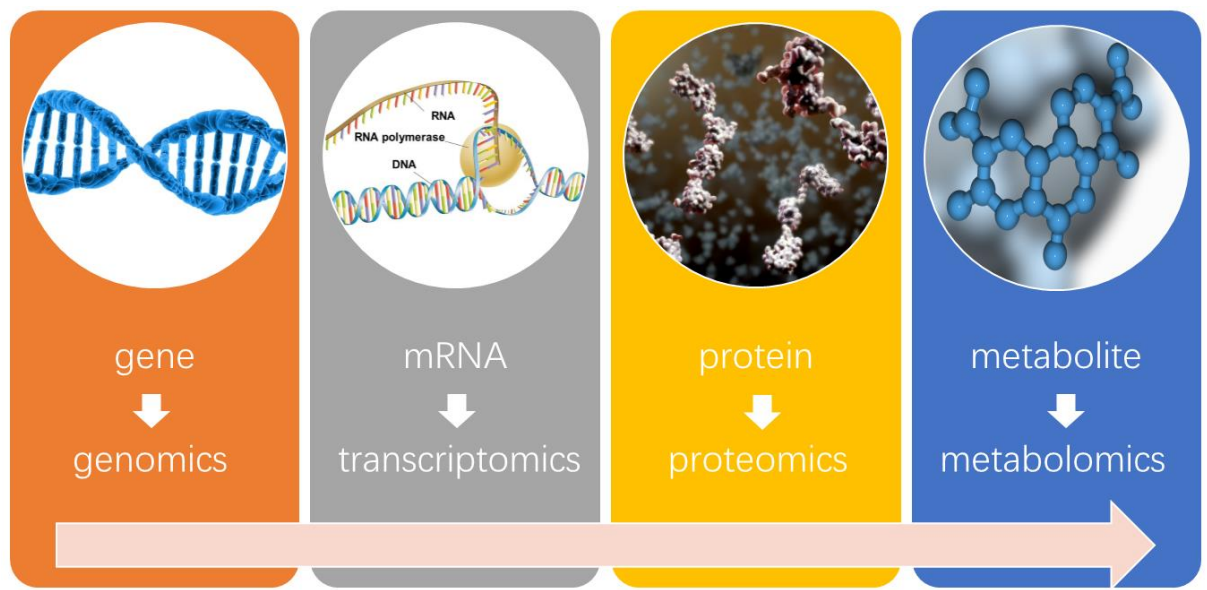


Figure 6 : Schéma des relations entre les différentes couches d'informations au sein d'un échantillon biologique (creative-proteomics.com)

1.3.3. Méthodes d'analyse des données transcriptomiques par séquençage de l'ARN (RNA-seq)

1.3.3.1. Méthodes bio-informatiques appliquées au traitement des données transcriptomiques

1.3.3.1.1. Le séquençage

1.3.3.1.1.1. Préparation des échantillons

Le séquençage haut-débit a révolutionné l'analyse du transcriptome. Les technologies disponibles aujourd'hui permettent d'étudier l'expression de tous les gènes exprimés du génome sans *a priori*. Le séquençage est l'étape permettant de déterminer la succession des bases nucléotidiques présentes dans un acide nucléique, après conversion en ADN. Les séquenceurs utilisés actuellement sont dits à haut-débit car ils permettent la lecture d'un très grand nombre de séquences. De plus, le temps nécessaire pour les obtenir a constamment diminué, alors que la quantité des données produites n'a cessé d'augmenter. Par exemple, le séquenceur Illumina HiSeq 3000 permet de générer 700 Gb de données par flow-cell en 2,5 jours, permettant une profondeur d'analyse conséquente. Le dernier séquenceur de chez Illumina, le NovaSeq permet de générer plus de 1000 à 3000 Gb en fonction de la flow-cell utilisée, et le tout en 44 heures. Néanmoins, pour pouvoir séquencer les acides nucléiques d'un échantillon biologique, il est nécessaire de respecter des critères de pureté, de quantité et de

qualité des acides nucléiques, ADN ou pour le transcriptome ARN, minimales. La quantité d'ARN minimale est d'environ 100 ng/ μ l et pour la qualité, les ratio 260/280 pour déterminer la contamination par des protéines et le ratio 260/230 pour déterminer la présence de sels doivent être tous les deux supérieurs à 1,8-2,0. L'étude du RNA integrity number (RIN) est utile pour connaître l'état de l'ARN extrait et l'absence de dégradation par les nucléases. Idéalement, il doit être supérieur à 8,5 sur une échelle de 0 à 10 (**Figure 7**).

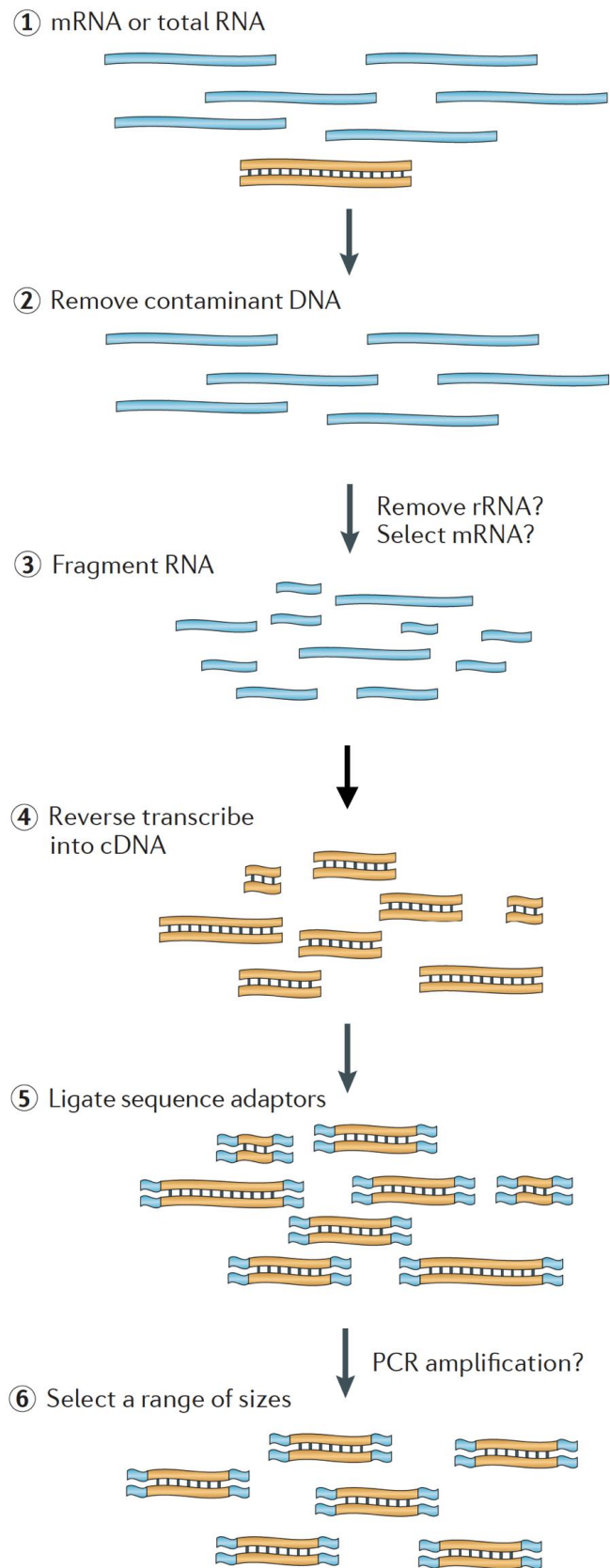


Figure 7 : Etapes pour la préparation d'échantillons pour le séquençage

Les ARN présents dans l'échantillon sont un mélange de différents types tel que l'ARN ribosomique, l'ARN de transfert, ou encore l'ARN messager. La majorité de ces molécules correspond à des ARN ribosomiques. Pour les études de transcriptomique par la technologie RNA-seq, seul l'ARN messager est étudié ; il est donc nécessaire d'éliminer les autres types d'ARN. Il existe 2 méthodes très utilisées pour dépléter les échantillons d'ARN ribosomique, la première consiste à utiliser des billes magnétiques poly-T permettant la sélection des ARN messagers. Elle a pour avantage d'être très spécifique mais elle a pour inconvénient de ne fonctionner que sur les échantillons issus d'eucaryotes car il n'existe pas de queue poly-A chez les procaryotes. La seconde méthode consiste à utiliser des billes ciblant les ARN ribosomiques. Cette méthode fonctionne chez toutes les espèces mais nécessite une optimisation. Elle est moins efficace et permet la récupération des ARN messagers mais aussi des ARN non codants (Figure 8).

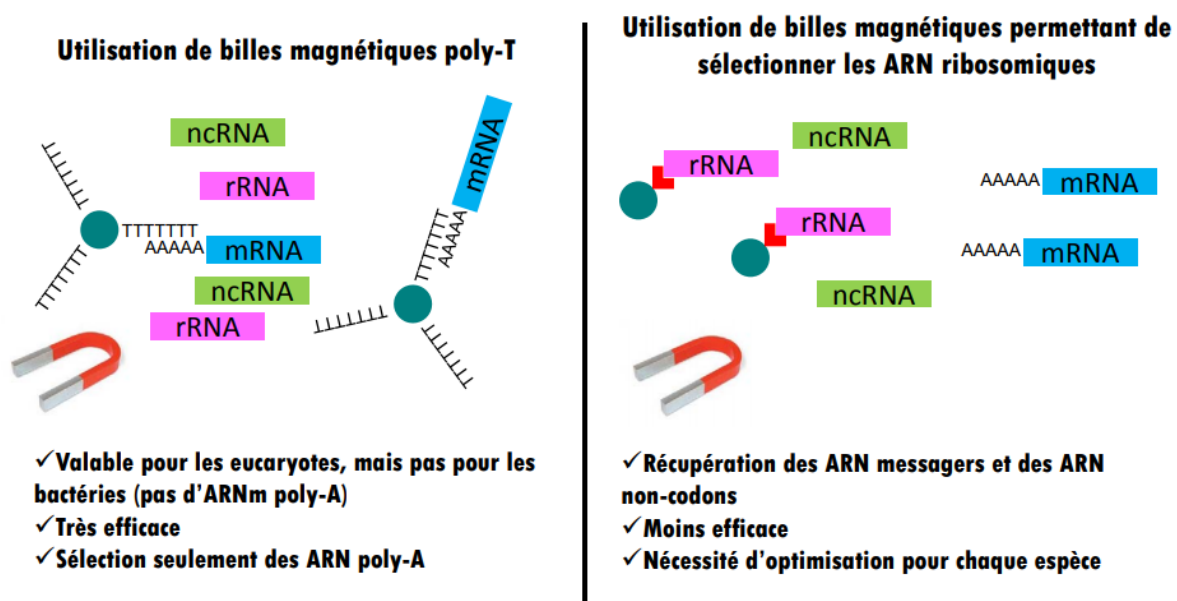


Figure 8 : Bilan des méthodes de purification des ARNm (Bouchez, 2016)

Cette première vérification est faite à partir de l'étape de préparation des librairies. Une fois les ARNm purifiés, une synthèse du brin ADN complémentaire est réalisée, suivie d'une séparation des extrémités pour permettre l'adénylation des extrémités 3'. Une liaison à des adaptateurs est ensuite nécessaire pour identifier les séquences et permettre l'amplification par PCR. La dernière étape consiste à vérifier la qualité des librairies obtenues (Figure 9).

Après cette préparation, les échantillons sont transférés sur une flowcell et l'amplification via la génération des clusters va permettre l'étape proprement dite de détermination de la séquence.

Chaque flowcell contient 8 lignes pouvant chacune contenir un ou plusieurs échantillons. Le séquençage par synthèse est ensuite effectué pour déterminer la succession des bases nucléotidiques.

Après la préparation des bibliothèques, le séquençage des fragments obtenus va être réalisé et produit des « reads ». Il est possible de séquencer un seul des brins (Single end) ou les deux brins (Paired end). L'avantage de cette dernière est de permettre d'éviter les multiples alignements sur le génome à cause des zones répétées. Néanmoins, il est nécessaire d'utiliser un kit de préparation des bibliothèques permettant l'étude double brin. Le prix du kit est identique mais cette technologie entraîne une augmentation du coût du séquençage.

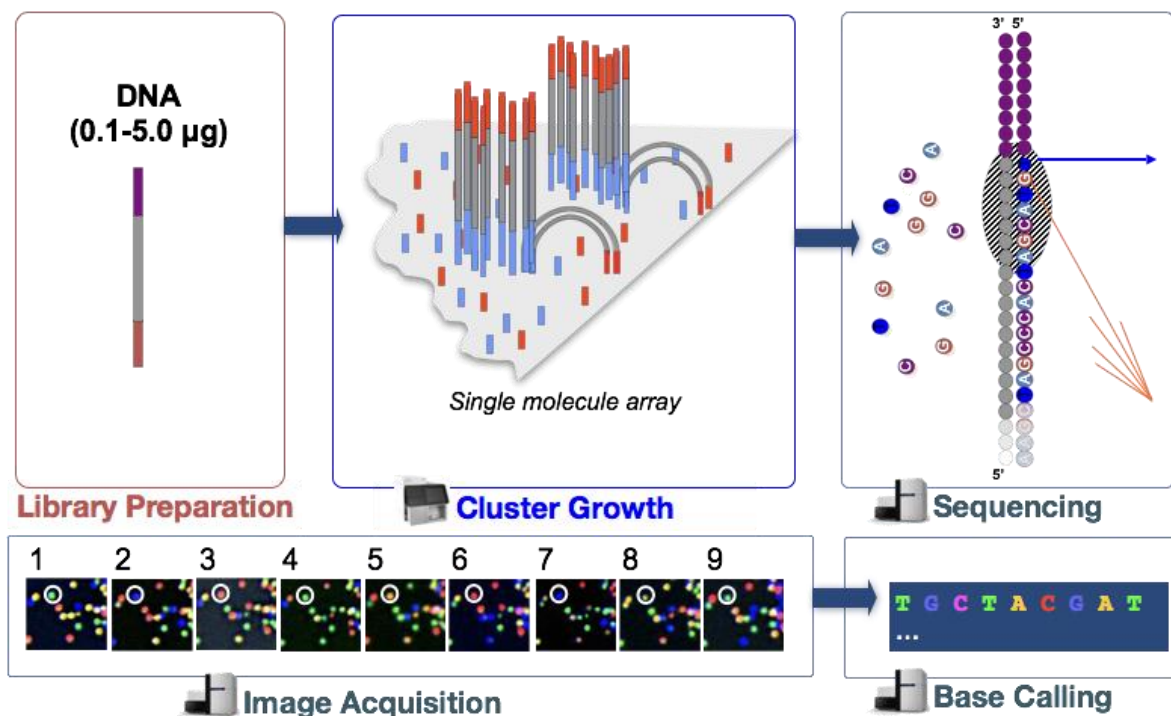


Figure 9 : Fonctionnement du séquençage (Harvard Chan Bioinformatics Core (HBC))

1.3.3.1.1.2. Analyse de la qualité et nettoyage des séquences

1.3.3.1.1.2.1. Le démultiplexage

A la fin du séquençage, les données obtenues correspondent à des fichiers BCL contenant l'information de tous les individus présents sur une ligne de la flowcell. Ce format BCL ou *Binary base call* est le fichier de sortie brute du séquenceur. Pour l'analyse, il est nécessaire de le convertir au format FASTQ et cela passe par l'étape de démultiplexage. En effet, dans la plupart des analyses de séquences, il y a plusieurs individus par ligne et il est donc nécessaire

d'attribuer les séquences à l'individu correspondant. Pour cela chaque échantillon contient un adaptateur spécifique permettant de trier les séquences. Le logiciel majoritairement utilisé pour cette étape est *bcl2fastq*. Cette étape est présente dans l'ensemble des technologies d'étude transcriptomique à haut débit (RNA-seq, single cells RNA-seq, spatial transcriptomique) (Figure 10).

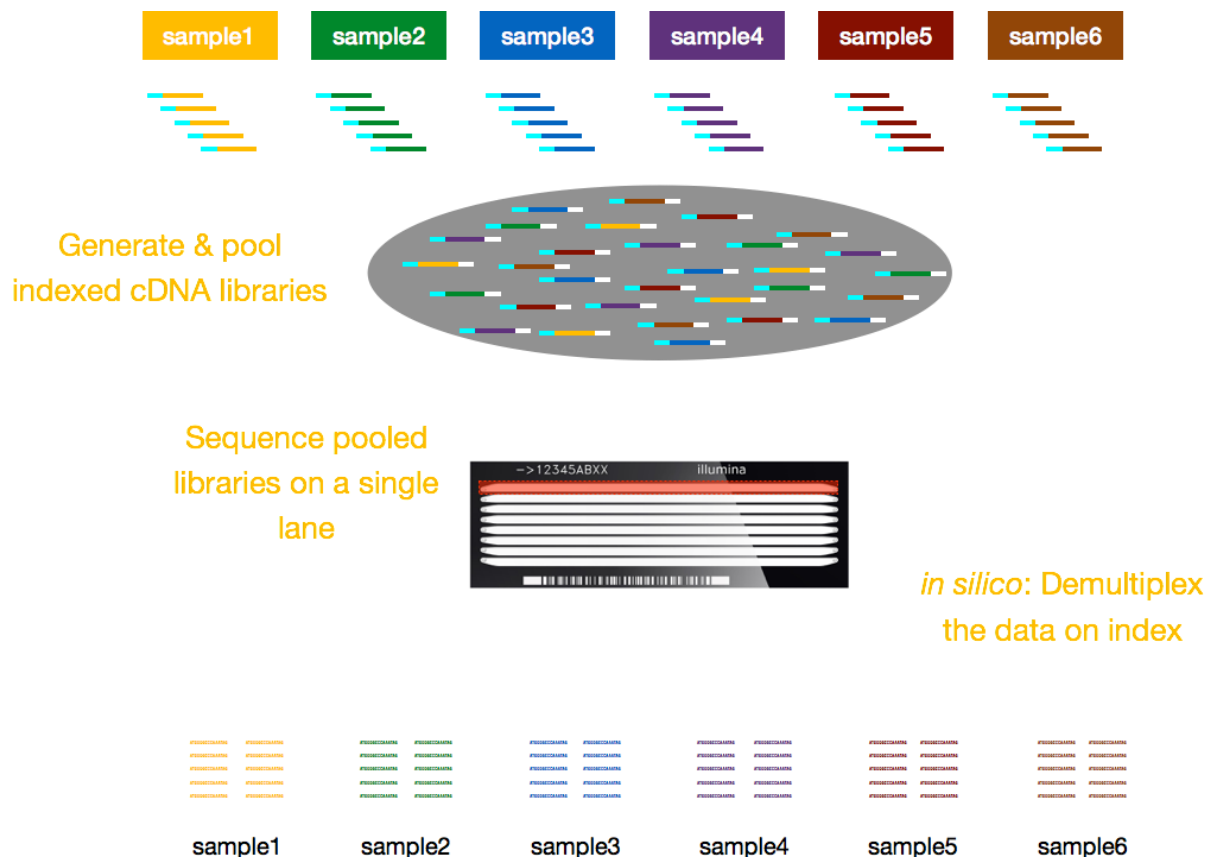


Figure 10 : Fonctionnement du demultiplexage de séquences (Harvard Chan Bioinformatics Core)

1.3.3.1.1.2.2. Le fichier FASTQ

Le fichier FASTQ est un format type texte qui a été développé par Jim Mullikin lors de l'année 2000. Ce fichier permet de stocker la séquence obtenue à l'étape du séquençage, ainsi que la qualité de cette séquence et des informations liées à l'expérience. Ce fichier correspond à la fusion des fichiers FASTA (séquence) et des fichiers QUAL (qualité) sous la forme d'un seul fichier pour faciliter sa manipulation. Néanmoins, ce type de fichier ne donne aucune information relative au génome étudié. Il contient 3 lignes par segment de séquence, avec chacune une information précise :

La première ligne correspond à l'identifiant de la séquence. Chaque information est séparée par deux points. Voici un exemple d'identifiant :

HWI-ST1136 :117:HS055:3:1101:1134:22441:N:0:GCCAAT

HWI-ST1136	Nom du séquenceur
117	Identifiant du run
HS055	Identifiant de la flowcell
3	Le numéro de la ligne
1101	Le numéro du tile
1134	La coordonnée X sur la flowcell
2244	La coordonnée Y sur la flowcell
1	Le numéro de la paire si analyse paired end (1-2)
N	N si la séquence passe le filtre de qualité sinon Y
0	Aucun des bit contrôles n'est activé
GCCAAT	Index de la librairie (si multiplexage)

La seconde ligne est la séquence nucléotidique brute correspond à un enchainement de A (Adénine), C (cytosine), G (guanine) et T (Thymine).

La dernière ligne correspond à la qualité de la séquence désignée phred. Durant les dernières années, 2 méthodes existaient pour le calcul de ce score (variant Sanger, variant Solexa/Illumina). Depuis 2011, un seul variant est utilisé, le score Sanger, calculé via la formule suivante :

$$Q_{sanger} = -10\log(p)$$

Où p correspond à la probabilité d'erreur calculée lors de l'identification de la base par le séquenceur.

Celle-ci varie entre 0 et 50. Un score de 10 indique un pourcentage d'erreur sur une base de 10% alors qu'un score de 50 indique une probabilité erreur de 0,00001%. Ce score est codé d'un point de vue informatique via le langage ASCII (American Standard Code For Information Interchange). Chaque chiffre est codé par un caractère unique.

1.3.3.1.1.2.3. L'analyse qualité

La première étape après l'obtention des séquences FASTQ est l'étude de leur qualité. Le logiciel le plus utilisé pour cette étude est le logiciel *FASTQC*. Il a été créé par l'équipe de Simon Andrews (Babraham Institute, Cambridge, Royaume-Uni). Au départ, ce logiciel a été développé pour les données génomiques. De ce fait, certaines analyses indiquent des erreurs liées à la nature des données transcriptomiques. Ce logiciel nous fournit diverses informations sur la séquence dont voici les plus importantes

- « *Basics statistics* » permet d'avoir des informations générales tels que le nom du séquenceur, le nombre total de fragments, le nombre de séquences de faible qualité, la taille de celles-ci, ainsi que le pourcentage de GC.
- « *Per base sequence quality* » résulte en un graphique exprimant le score Phred en fonction de la position de la base et pour chaque read. Les résultats sont sous forme de boxplot avec la médiane en rouge. C'est grâce à ce graphique qu'il est possible de savoir si un nettoyage préalable des reads est nécessaire pour éliminer les séquences ou les portions de séquences dont la qualité est insuffisante.
- « *Per tile sequence quality* » permet de dénombrer le nombre de reads pour chaque score Phred. Ce graphique permet de vérifier que le séquençage s'est déroulé conformément aux attentes et que les séquences sont de bonne qualité à toutes les positions.
- « *Per sequence quality score* » nous indique la proportion d'une base donnée à une position donnée sur un read.

Le logiciel *MultiQC* a les mêmes fonctions que *FASTQC* mais il possède aussi des fonctions bio-informatiques liées à l'alignement et à la quantification. De plus, il a l'avantage de combiner l'ensemble des rapports de plusieurs séquences en un seul permettant un résumé synthétique.

1.3.3.1.1.2.4. Le nettoyage des séquences

Le nettoyage des données après l'analyse qualité n'est pas systématique. Cependant celui-ci est fortement recommandé car il permet d'enlever les séquences de mauvaise qualité, de supprimer les séquences contaminantes si elles sont présentes, et d'enlever les adaptateurs (**Figure 11**). En effet après le nettoyage, le Phred correspondant à la qualité des séquences est bien meilleur. Le logiciel *cutadapt* fonctionnant sous Linux permet de supprimer les adaptateurs mais aussi d'écartier les séquences de mauvaise qualité [321]. Le logiciel actuellement le plus utilisé est

Trimgalore qui correspond à la compilation du logiciel *FASTQ* pour l'analyse qualité et du logiciel *Cutadapt* pour le nettoyage des séquences.

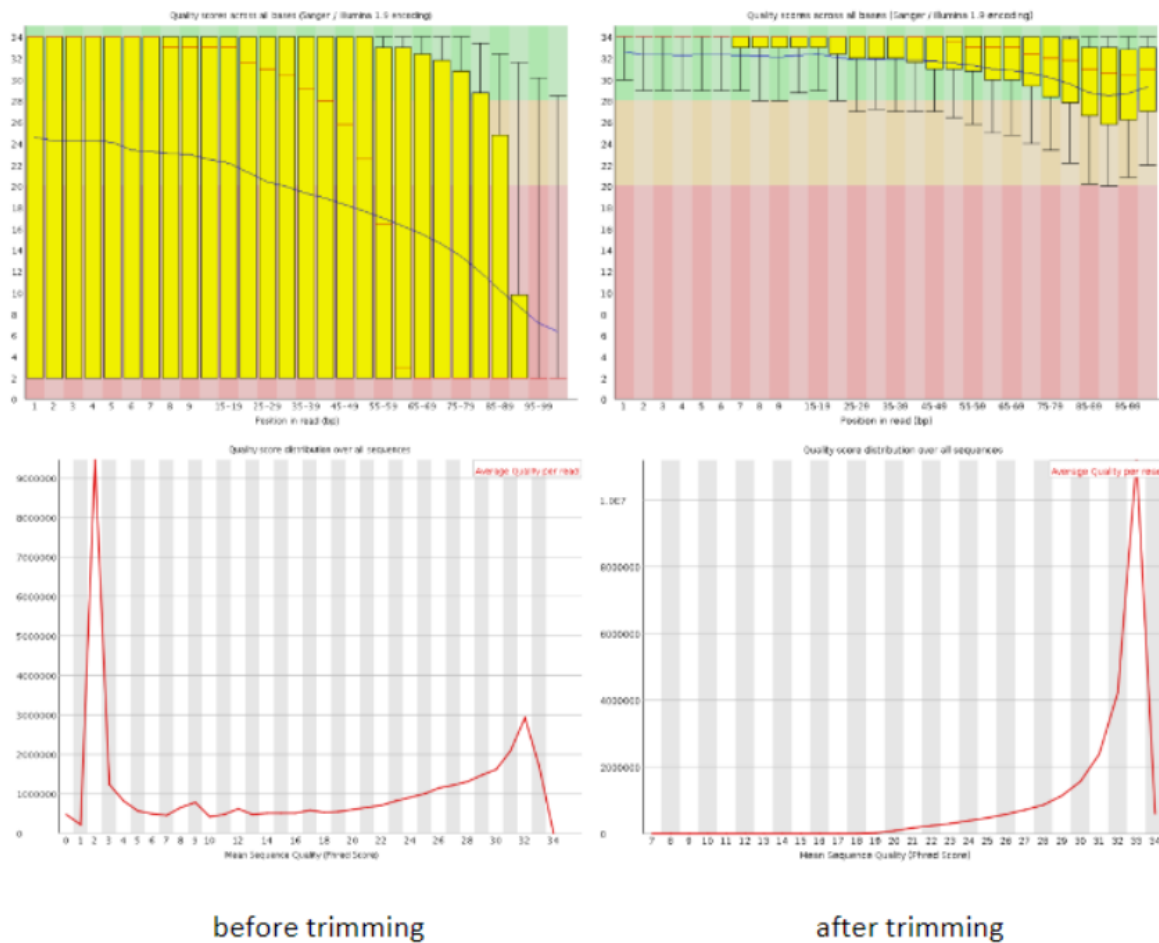


Figure 11 : Effet du nettoyage des séquences via le logiciel Trimgalore (Trimgalore user guide)

1.3.3.1.1.3. Alignement et pseudo-alignement de l'ARN

L'alignement a pour but de comparer les séquences nucléotiques ou « reads » issus du séquençage avec le génome de référence de l'espèce d'où provient l'échantillon biologique. Il existe de nombreux logiciels permettant cet alignement avec pour chacun des avantages et des inconvénients. Le phénomène d'épissage constitue la plus grande difficulté de ces algorithmes. En effet, lors de la transcription, les ARN transcrits vont subir des coupures et des ligatures pour éliminer les introns pour ne garder que les exons. La plupart des algorithmes utilisés aujourd'hui prennent en compte ce phénomène. Les logiciels se divisent en 2 grandes catégories, en fonction du principe utilisé [322].

Les algorithmes les plus anciens (à partir de 2010) comme par exemple *STAR*, *HISAT2*, utilisent l'alignement complet. Ils ont donc une bonne précision car ils comparent les séquences

obtenues avec celle du génome de référence de l'espèce étudiée. La quantification des gènes et des transcrits est réalisée via le dénombrement des « reads » chevauchant chaque gène ou isoforme. Néanmoins pour cela, ces utilitaires ont besoin de beaucoup d'espace mémoire pour fonctionner (32 gigaoctets en moyenne pour *STAR*), mais ils ont aussi un temps de calcul plus long (plusieurs heures). Tous ces algorithmes n'utilisent pas exactement la même approche.

La première méthode est dite « *d'exon first* ». Le principe est d'aligner les séquences sur le génome de référence afin de définir les exons, puis de faire des lectures pour identifier les jonctions entre ces exons. De nombreux algorithmes l'utilisent, dont les plus connus sont *Tophat* [323] ou *Bowtie2* [324]. Néanmoins ces algorithmes ont été remplacés au détriment de ceux utilisant la seconde méthode dite du « *seed and extend* ». Le principe est d'aligner une partie de la séquence et d'étendre l'alignement petit à petit. Celui-ci est utilisé par *STAR* [325] ou encore *HISAT2* [326].

A l'inverse, d'autres algorithmes utilisent la méthode du pseudo-alignement via un jeu de séquences issues du génome au lieu de l'ensemble de l'information génétique de référence. L'avantage de ces méthodes est le gain de temps ; puisqu'une vingtaine de minutes suffit par échantillon, tout en gardant une précision comparable aux méthodes classiques décrites précédemment. C'est le cas de l'algorithme *Kallisto* qui est basé sur la méthode des k-mers (courte séquence de k bases). La première étape de cet algorithme est la création d'un index tout comme les aligneurs classiques. Néanmoins, cet index porte le nom de graphe de Bruijn (T-DBG) et est réalisé à partir des k-mers que l'algorithme a identifiés dans le génome. Ce graphique contient la relation entre le k-mer et le ou les transcrit(s) d'origine permettant *in fine* de déterminer le gène en question. Puis, pour chaque read, *Kallisto* cherche à déterminer quels transcrits présents dans l'index sont compatibles avec le read qui a été découpé préalablement en k-mers afin d'être identifiable par l'algorithme [327]. *Salmon* est un logiciel utilisant un concept similaire à celui de *Kallisto*, à savoir le quasi-mapping [328] (**Table 5**).

Les fichiers de sortie après alignement correspondent à des fichiers de type SAM (Sequence Alignment Map), ou BAM (Binary Alignment Map) qui est la version binaire du type SAM. Ces types de fichiers ont été créés par Heng Li et Bob Handsaker en 2009 lors du projet *1000 génomes*. Pour étudier ces fichiers, le logiciel *SAMTOOLS* a été développé [329]. Il permet d'étudier ou d'obtenir les données d'une région délimitée du génome.

Table 5 : Etude entre les différents types d'aligneurs utilisées dans l'analyse RNA-seq

Nom	Type d'alignement	Caractéristiques	Date de création	Référence
Hisat1 / 2	Alignement total	Temps de calcul relativement faible Moins performant que les autres	2015 / 2019	[326], [330]
TopHat1 / 2	Alignement total	Fort pourcentage d'alignement	2009 / 2013	[323], [331]
Bowtie1 / 2	Alignement total	Fort pourcentage d'alignement	2009 / 2012	[324], [332]
Bwa	Alignement total	Meilleur ratio	2009	[333]
STAR	Alignement total	Pourcentage d'alignement / Temps de calcul	2013	[325]
Kallisto	Pseudo-alignement	Faible temps de calcul	2016	[327]
Salmon	Pseudo-alignement	Faible temps de calcul	2017	[328]

1.3.3.1.1.4. La quantification

L'un des intérêts fondamentaux du RNA-seq est la quantification de l'expression de gènes ou de transcrits comme identifiés via l'alignement décrit précédemment. La quantification des séquences alignées pour un gène ou un transcrit est corrélée positivement à son niveau d'expression dans l'échantillon examiné.

Cufflinks est une suite permettant le traitement des données obtenues à la suite de l'alignement. Il a été développé en 2010 par Core Trapnell [334]. Il permet l'assemblage de transcrits ainsi que l'estimation de leur abondance. Il est aussi possible de comparer les noms des transcrits et de faire une analyse différentielle de ceux-ci, même si ces deux fonctions sont aujourd'hui

moins utilisées par la communauté scientifique. Cette suite nécessite le fichier d'annotation du génome, ainsi que des fichiers SAM pour chacun des échantillons à traiter. Le point le plus intéressant de ce logiciel est la reconstruction des transcrits.

Stringtie est une suite logicielle développée en 2015 par Mihaela Pertea. Elle permet l'assemblage de transcrits ainsi que leur dénombrement, avec une forte ressemblance avec *Cufflinks*. Il utilise cependant un nouvel algorithme avec un assemblage *de novo* pour assembler et quantifier les multi-transcrits correspondant à chaque locus. Il est utilisable aussi bien sur des courtes que des longues séquences nucléotidiques.

Les outils *Cufflinks* et *Stringtie* ont des fonctions similaires. Néanmoins, le temps moyen de traitement pour un échantillon est d'une heure pour *Stringtie* alors qu'il est de huit heures pour *Cufflinks*. De plus, *Stringtie* assemble correctement de 32 à 53% de transcrits en plus que son homologue. Il est aussi meilleur pour la reconstruction des gènes dont l'abondance est faible, ceux contenant beaucoup d'exons ainsi que de multiples isoformes [335].

Cette étape de reconstruction est importante à faire avant de procéder à la quantification afin qu'un maximum de transcrits soit identifié.

HTSeq est une librairie développée en utilisant le langage Python par Simon Anders en 2015; il permet l'étude des données de séquençage à haut débit [336]. Dans cette librairie, un outil HT-seq count permet la quantification des séquences. Tout comme ses homologues, il nécessite un fichier d'annotation du génome. Cet outil ne tient compte que des séquences alignées non ambiguës (1 seul alignement).

RSEM (pour *RNA-seq by Expectation Maximization*) est une méthode de quantification développée par Bo Li et Colin N Dewey en 2011 [337]. Il s'agit du logiciel le plus utilisé aujourd'hui pour la quantification des transcrits en RNA-seq. Il peut être utilisé sur des transcriptomes obtenus *de novo* car il ne nécessite pas obligatoirement un génome de référence contrairement aux autres méthodes pré-citées. Ce logiciel a une bonne habileté à résoudre un problème non résolu par les méthodes d'alignement classiques, comme par exemple les gènes multi-alignés sur le génome.

Un second programme permet la quantification via des fichiers d'alignement (BAM/SAM). Il s'agit du logiciel *FeatureCounts* issu de la suite *Subreads*. Celui-ci permet une quantification rapide utilisant une faible capacité mémoire. Il ne nécessite que les fichiers BAM ainsi que le fichier d'annotation du génome au format GTF, qui est celui déjà utilisé pour l'alignement. Ce programme possède des options permettant la quantification quel que soit l'acide nucléique de départ [338]. L'ensemble des méthodes présentées précédemment hormis *HTseq* car non inclus

ont des performances similaires [339]. Dans une seconde étude, RSEM semble la méthode ayant les meilleures performances [340].

1.3.3.1.2. Méthodes biostatistiques appliquées aux données génomiques et/ou transcriptomiques

Une fois que les gènes et/ou les transcrits ont été quantifiés, il est nécessaire d'utiliser l'analyse biostatistique pour permettre leur visualisation ou connaître l'expression différentielle entre conditions. Une grande part de ces outils a été développé via le langage R très utilisé dans le domaine des biostatistiques, notamment ceux appliqués aux données génomiques tels que *DESeq2* ou encore *EdgeR* qui sont les plus utilisés et seront présentés dans les parties suivantes. Néanmoins, de nombreux logiciels sont développés dans un autre langage informatique, le langage Python le préférant pour sa rapidité de calcul et sa simplicité d'utilisation. Les données obtenues après la quantification se présentent sous la forme d'une matrice contenant les gènes en lignes et les différents échantillons en colonnes.

La première étape est de visualiser la distribution des données pour identifier des données ou des échantillons aberrants. La première étude à réaliser est d'observer la profondeur de séquençage (nombre de « reads ») pour chacun des échantillons. En effet, un nombre de « reads » très variable peut avoir un impact sur les résultats de l'analyse statistique. Pour aller plus loin dans la visualisation des données, il est possible d'utiliser une heatmap ou une ACP.

La heatmap est une représentation en deux dimensions de données statistiques utilisant une échelle de couleur. Elle a été développée pour l'étude météorologique ou géographique. Néanmoins, avec l'explosion des sciences des données, des outils ont été développés dans R pour l'analyse des données à haut débit telles que des données transcriptomiques avec *Pheatmap* [341].

- L'heatmap d'une matrice de corrélation permet la détection d'échantillons aberrants ou « outliers ». Cette matrice utilise l'expression de tous les gènes pour évaluer les liens entre les différents échantillons. La corrélation peut être déterminée via plusieurs méthodes statistiques telles que celle de Pearson qui calcule un lien linéaire entre les variables. Les méthodes de Kendall et de Spearman ont l'avantage de pouvoir calculer une corrélation non linéaire. Les deux premières sont les plus utilisées. Via la visualisation à l'aide d'une heatmap, il est possible d'étudier les relations entre les différents échantillons et d'identifier un ou plusieurs échantillon(s) qui se différencient

des autres, comme par exemple à cause d'un nombre de reads anormalement plus faible ou d'un défaut de la qualité qui n'aurait pas été identifié lors des étapes précédentes.

- Cette représentation peut aussi être utilisée pour étudier l'expression des gènes après un passage dans les outils dédiés. Cela sera détaillé dans la partie relative aux logiciels utilisés pour l'étude des gènes différentiels.

La seconde représentation possible est l'Analyse en Composante Principale ou ACP. Cette méthode appartient à la catégorie des analyses multivariées. Elle permet de résumer l'ensemble des informations des 20 à 40 000 gènes sur 2 à 3 dimensions en limitant la perte d'informations. Cette méthode sera détaillée dans la partie représentation des données. Par cette méthode, il est possible de confirmer la présence d'échantillon(s) aberran(t)s, une position dans la distribution très excentrée par rapport à celle des autres, indiquant un profil d'expression génique très différent. En fonction de la question posée, cette représentation peut donner une première indication sur l'effet de tel ou tel facteur en lien avec la question expérimentale (effet d'un traitement par exemple). La distribution des gènes est également importante à observer car une grande partie de ceux-ci ne sont pas exprimés et le nombre de reads correspondant est nul. La plupart des logiciels les exclut car ils peuvent parasiter l'analyse. Il est donc utile de les supprimer si cela n'est pas fait automatiquement pour améliorer la qualité de l'analyse.

Pour aller plus loin dans l'analyse, il est nécessaire de travailler avec des packages spécialisés dans la détermination des gènes différentiels. Les trois logiciels les plus utilisés sont *DESeq* [342] et surtout son évolution *DESeq2* [343], *EdgeR* [344] ainsi que *limma* [345]. Dans cette partie, uniquement les packages *DESeq2* et *EdgeR* seront détaillés.

DESeq2 est un algorithme développé par Michael Love en 2014. Il fait suite à une première version qui est *DESeq*. *EdgeR* lui est un ensemble logiciel développé par Mark Robinson en 2011. Ils sont tous les deux basés sur l'hypothèse d'une distribution négative binomiale (ou loi de Pólya) correspondant à une distribution de probabilité discrète de type Poisson. Néanmoins, il existe des différences majeures entre ces deux outils dont la première correspond à l'étape de normalisation de la distribution des données.

Le nombre de reads varie entre les échantillons, ce qui peut être interprété de façon erronée comme une variation d'origine biologique. Néanmoins, il ne s'agit pas du seul biais possible. En effet, en plus des variations biologiques, il existe des variations techniques introduites au moment de la préparation des échantillons. La préparation de la librairie peut donc causer un biais dans l'analyse du transcriptome. L'effet de la flowcell ainsi que de la lane sont aussi à prendre en compte. Néanmoins, ces biais n'ont pas tous le même impact sur l'analyse. Les biais

d'origine biologique ont des conséquences majorées par rapport à ceux causés par la préparation des librairies ou encore ceux liés au matériel utilisé lors du séquençage. Une normalisation des données est donc nécessaire pour réduire l'ensemble de ces perturbations.

Les méthodes de normalisation se distribuent en quatre grandes catégories : ajustement des distributions, nombre de reads effectifs, longueur du gène et teneur en GC. Certaines de ces méthodes utilisent les comptages de reads alors que d'autres utilisent la taille de la librairie. C'est cette dernière qui est la plus utilisée via 2 méthodes qui varient selon les logiciels. Il s'agit de la méthode « Trimmed Mean of M-values (TMM) » pour *EdgeR* et de la méthode « relative log-expression (RLE) » pour *DESeq2*. De toutes les méthodes de normalisation disponibles, ce sont les méthodes implémentées dans *DESeq2* (RLE) et *EdgeR* (TMM) qui sont les plus robustes. D'un point de vue analytique, ces 2 méthodes sont cependant équivalentes [346].

Table 6 : Principales méthodes de normalisation de données RNA-seq

Nom	Catégorie	Package utilisable	Date de création
RLE	Taille de librairie	EdgeR, DESeq2	2010
TMM	Taille de librairie	EdgeR	2010
RPKM	Longueur du gène	EdgeR	2009
Quartile normalization	Ajustement de la distribution	EdgeR	2010
Total read count adjustment	Ajustement de la distribution	EdgeR	2008

De nombreuses méthodes existent pour normaliser les données ; il est parfois difficile de s'y retrouver. En 2013, une étude comparant l'efficacité de l'ensemble de ces méthodes a été réalisée. Les méthodes de l'ajustement sur le nombre total de counts et la RPKM ont des résultats similaires à l'exploitation de données non traitées. D'autres critères ont été évalués tels que la distribution, le nombre de faux positifs ou encore l'intra-variance. Il en ressort que les meilleures méthodes sont celles basées sur la taille de la librairie (RLE et TMM) (**Table 6**) [347].

Une fois la normalisation réalisée, l'étape suivante est la détermination des gènes différentiels via un modèle complet ou réduit. Le modèle linéaire généralisé ou *glm* permet de comparer différentes conditions d'expérience pour aller plus loin que la simple comparaison 2 à 2. Ce modèle est une généralisation de la régression linéaire. Le modèle réduit, lui, ne contient que

certaines de ces conditions, pour isoler l'effet d'une seule d'entre elles par exemple. Une fois le modèle identifié, une liste de gènes est obtenue avec une valeur de p (*p-value*) et une variation de l'expression exprimée en log2 (*log2foldchange*). Le *log2foldchange* correspond au rapport de l'expression entre une condition A et une condition B prise comme référence. Plus la valeur absolue est élevée, et plus la différence d'expression du gène entre les 2 conditions est grande. La *p-value* ajustée correspond à la significativité du test corrigée pour les tests multiples. Dans le cas de la détermination de gènes différentiels, un ajustement de la valeur de p via un test post-hoc est nécessaire car plus le nombre de tests réalisés est grand, et plus le risque de détecter des gènes faussement significatifs augmente.

Dans de nombreux cas, il s'agit soit du test de Bonferroni, qui est le test le plus sélectif, soit du test de Benjamin-Hochberg (BH), qui est moins sévère et le plus couramment utilisé pour ce type d'analyse. La correction de BH est une méthode faisant partie des corrections de type FDR (pour *False Discovery Rate*). Elle corrige la *p-value* en prenant en compte 2 critères, le nombre de *p-values* étudiées ainsi que le taux de *p-values* non significatives. Il s'agit d'une des méthodes de corrections les moins sévères et elle est donc utilisée dans des cas où l'on veut garder le maximum de gènes à étudier. Le test de Bonferroni est basé uniquement sur le nombre de *p-values*, ce qui, en conséquence, la rend plus sévère. Elle est utilisée lorsque l'on veut garder une liste réduite de gènes très significatifs.

Les règles en vigueur indiquent de sélectionner les gènes ayant une valeur de p ajustée inférieure à 0,05 ou 0,01, si l'on veut faire une sélection plus forte. Ces tests utilisant le nombre de gènes dont la *p-value* est inférieure à un seuil, il est donc important de restreindre l'analyse aux seuls gènes dont l'expression n'est pas nulle, au risque d'allonger inutilement la liste des gènes ou transcrits analysés, et de réduire *in fine* celle de ceux qui seront considérés comme significatifs. La sélection peut aussi être faite sur la différence d'expression en choisissant des gènes dont le *log2FC* est supérieur à une certaine valeur, pour ne conserver que les gènes qui varient beaucoup entre 2 conditions. Cependant, comme nous le verrons plus tard, cette sélection n'est pas souhaitable chaque fois qu'une analyse fonctionnelle des gènes/transcrits identifiés comme différents est envisagée à la suite de cette première étape.

D'autres méthodes pour la détermination de la liste des gènes différentiels existent. Chaque outil a son propre but : *Sleuth* est utile pour la recherche d'isoformes différentielles, mais aussi pour la recherche de gènes différentiels dans le cas de plans expérimentaux complexes [348]. Les trois autres logiciels vus précédemment sont utilisés dans tous les cas de figure et permettent l'identification des gènes différentiels quel que soit le dessin expérimental ou le

nombre de réplicats biologiques. Plusieurs comparaisons des différentes méthodes d'alignement et de recherche des gènes différentiels utilisant les outils les plus utilisés actuellement ont été publiées [349]. Quelle que soit la méthode d'alignement, l'article conclut au fait que ce sont *Limma* et *DESeq2* qui donnent les résultats les plus robustes, c'est-à-dire qu'ils sont similaires quelle que soit la méthode d'analyse utilisée. Une autre étude publiée en 2016 a montré que les résultats obtenus avec EdgeR et DESeq2 étaient très similaires [346]. Une fois la liste des gènes différentiels déterminée, différentes méthodes de visualisation sont utilisées pour résumer l'information : il peut s'agir d'une heatmap ou de volcano plots. Comme vu précédemment, la heatmap permet de visualiser l'expression des gènes différentiels dans plusieurs conditions. Les volcano plots sont des graphiques en nuage de points très utilisés pour la comparaison deux à deux. En effet, ce type de graphique permet de représenter l'ensemble des gènes, dont ceux différemment exprimés en fonction de la différence d'expression et de la valeur de p ajustée, et des seuils retenus à partir desquels les gènes sont considérés comme différents entre les deux conditions.

DESeq2 propose aussi une méthode dans le cas où le facteur Temps est pris en compte ; cette méthode est très utilisée pour l'étude de la réponse à la vaccination par exemple. Il permet d'étudier le nombre de reads pour les gènes les plus différentiels, à différents moments de la réponse analysée. De plus il est possible d'identifier ceux qui varient le plus fortement et de présenter leur variabilité dans plusieurs conditions soit en fonction du temps, soit en fonction du groupe [350].

L'étude des gènes différentiels est très limitée et au final relativement préliminaire dans les étapes de description de la variation d'expression du génome. De nombreuses bases de données et outils ont été développés tels que *GSEA* pour identifier et décrire les fonctions associées à ces gènes et dont l'activité varie en fonction des paramètres biologiques étudiés.

1.3.4. Méthodes d'analyse de données non transcriptomiques

1.3.4.1. Méthodes d'analyses des données protéiques complexes : exemple des cytokines

Les méthodes de dosage protéique à haut débit produisent un ensemble de données de grande taille, dont l'analyse nécessite souvent des outils spécifiques pour leur traitement et leur visualisation.

Tout comme l'analyse de la qualité des données transcriptomiques, il est nécessaire de vérifier la validité des valeurs obtenues avec ces méthodes de débit moyen. Le package R *DrLumi*

permet de vérifier la qualité des données collectées et de construire, via des algorithmes implantés, les références externes en choisissant le modèle le plus adapté pour leur analyse [351]. Il permet de traiter le bruit de fond, de déterminer les limites de quantification, et d'identifier les échantillons aberrants. L'analyse de la qualité des données passe par l'étude de la courbe des standards externes, et notamment via l'étude de la répétabilité. Le pourcentage du coefficient de variation (ou CV) est un critère pour mesurer la répétabilité de l'analyse. Ce pourcentage est toujours plus élevé dans les points extrêmes ; c'est pour cela qu'il est nécessaire de restreindre l'intervalle des données analysées à la partie médiane de la référence externe, de façon à conserver un CV le plus faible possible et ainsi obtenir une détermination plus précise et répétable. Dans le cas des méthodes utilisant des billes pour la quantification d'analytes, un nombre minimal de billes examinées est également requis parmi les critères de qualité et de validation de la mesure. Toutes ces analyses permettent d'effectuer le contrôle qualité de l'expérience.

L'interprétation des résultats peut se faire à partir des données brutes ou corrigées de la fluorescence mesurée (MFI), ou bien après conversion de ces valeurs avec les références externes, en concentration (généralement en pg/mL). Plusieurs études publiées ont montré que l'analyse à l'aide des MFI était plus robuste que celle utilisant les concentrations, et que des différences majeures dans le résultat final peuvent être observées en fonction du choix de la méthode d'analyse [352]. L'avantage de l'analyse fondée sur les mesures de fluorescence est qu'elle s'affranchit de l'utilisation d'une référence externe. En théorie, il n'y a pas de limites à l'analyse et d'effet plateau en cas de valeurs extrêmes. Néanmoins, cette dernière ne peut être réalisée que si les échantillons ont été traités de la même manière et avec le même appareil, alors que la méthode d'étude basée sur les concentrations permet de s'affranchir de ces contraintes, et de comparer des résultats obtenus dans différents contextes d'analyse (site, appareil, essai). Chacune des méthodes a ses propres avantages et inconvénients ; néanmoins, les différences entre ces deux méthodes ne sont pas toujours identifiables et elles varient en fonction de la valeur des échantillons par rapport à la distribution des valeurs de la référence.

L'analyse des cibles (chimiokines, cytokines) contenues dans ces tests immuno-enzymatiques se fait la plupart du temps, une cible après l'autre alors que leur expression et/ou leurs fonctions sont quelquefois proches. Une méthode d'analyse modulaire connue sous le nom de CYTOMOD [353] a été développée dans le but d'étudier non plus les analytes un par un, mais de manière plus globale à travers des modules d'expression voisine. Les valeurs mesurées de chaque cytokine sont d'abord normalisées via une étape de régression. Pour un analyte donné,

les valeurs de tous les individus constituant l'expérience sont comparées à la concentration moyenne des cytokines via une régression. Les concentrations de cytokines ajustées sont définies comme les résidus issus de la régression. Une étape de classification hiérarchique est ensuite réalisée pour *in fine* créer des modules de cytokines ayant le même profil d'expression, et la même évolution dans l'expérience. L'expression des modules peut être étudiée et mise en corrélation via une étape d'association avec des informations indépendantes sur les individus, comme des scores cliniques par exemple. Cette méthode permet de résumer l'information en utilisant quelques variables. Néanmoins, le développement de ces méthodes d'analyse est encore récent, et mériterait des développements complémentaires.

1.3.4.2. Utilité des données caractéristiques d'un animal

Les caractéristiques propres à chaque animal lors des essais dans des élevages commerciaux sont très utiles pour les études cliniques, comme dans nos travaux, mais elles peuvent aussi servir dans d'autres contextes (santé de l'animal, étude nutritionnelle). Elles sont souvent mises en relation avec des informations omiques pour étudier, comprendre, et interpréter certaines conditions particulières comme une maladie, la réponse à un traitement par exemple. Différents types de données peuvent être disponibles ou collectés spécifiquement pour ces études ;

Les données d'élevages, qu'elles soient cliniques (éventuellement sous la forme d'un score), métaboliques (hormones, métabolites) ou sanguine (numération cellulaire) sont analysées de manière classique et ne nécessite pas de méthodes statistiques aussi complexes, que celles vues précédemment. Néanmoins elles diffèrent des données obtenues lors d'essais expérimentaux. La nature des prélèvements est plus restreinte dans le contexte des élevages. De plus, la présence de valeurs manquantes est souvent plus grande et peut donc requérir l'utilisation de méthodes permettant de substituer ces données.

Avec l'automatisation de l'acquisition des données, la taille des ensembles à analyser est de plus en plus grande et nécessite l'utilisation d'outils informatiques plus performants, notamment ceux qui sont disponibles dans l'ensemble R.

En effet, les données obtenues peuvent contenir des informations manquantes, comme c'est souvent le cas pour des données collectées à grande échelle. Le package R *visdat* [354] permet de détecter et de visualiser les informations manquantes. Alors que l'analyse statistique requiert des tables de données exhaustives, il est possible de faire de l'imputation de données pour résoudre l'écueil de l'absence de données. Cette méthode est utilisée lorsque l'analyse nécessite la totalité des valeurs de l'ensemble des variables analysées ou si le nombre de données

manquantes est trop grand et au-delà d'une limite qui varie en fonction de l'effectif (<70%). Cette imputation est aussi réalisée avec les données omiques et c'est notamment le cas dans le cadre des analyses de single cells via les « dropout ». Ce sont des cellules où un gène exprimé modérément dans une cellule qui ne l'est pas dans une autre cellule du même type. Ce type de profil peut altérer l'analyse et donc nécessiter une imputation. Néanmoins de nouvelles méthodes se développent pour empêcher ces situations [355]. Il existe des méthodes peu sophistiquées comme le remplacement par la valeur la plus fréquente, ou par la moyenne globale. Ces méthodes sont peu robustes et d'autres méthodes plus satisfaisantes ont été développées. Le package *missMDA* [356] permet de gérer les informations manquantes dans le cadre d'analyses factorielles. Cet algorithme impute les données de manière à ce qu'elles n'aient pas d'impact sur l'analyse. Le package *MICE* [357] est l'un des plus utilisés et suppose que les données manquantes suivent une distribution aléatoire. Il contient de nombreuses méthodes statistiques assez sophistiquées en fonction du type de variable.

Les informations apportées par ces données de terrain sont très importantes et très utilisées dans les intégrations de données qui seront vues dans l'axe suivant. En effet, ces informations déterminent l'état d'un animal dans un domaine précis (immunitaire, métabolique, alimentaire) qui peuvent être mises en relation avec des informations transcriptomiques ou protéomiques dans le cas d'étude individu-centrées et donc de détecter un mécanisme cellulaire lié. Dans le cadre d'étude épidémiologique, ces informations sont cruciales car elles permettent de connaître l'état de l'animal et de déterminer l'impact d'une infection ou d'un problème alimentaire sur celui-ci.

1.3.5. La représentation et l'intégration de données

1.3.5.1. Objectifs et intérêt de l'intégration de données

L'intégration des données biologiques est permise par le développement de la biologie intégrative computationnelle (ou pan-omique), comme déjà indiqué précédemment. Le développement des techniques omiques a entraîné la production de grands jeux de données à propos de divers compartiments biologiques à partir de l'ADN/ARN (génomique et transcriptomique) ou encore des protéines et des métabolites (protéomique et métabolomique). L'utilisation d'outils informatiques et mathématiques est nécessaire pour extraire l'information biologique de ces ensembles complexes et peu exploitables de façon immédiate et directe. L'intégration de données a pour objectif de les rendre accessibles et compréhensibles en extrayant l'information biologique essentielle.

La co-expression des gènes peut avoir comme défaut de ne pas toujours correspondre à la réalité biologique, avec pour cause, la relation entre deux gènes qui n'est pas toujours systématique. L'intégration de données omiques avec des informations biologiques de types différents (phénotype, caractères particuliers ou signes cliniques) permet d'étudier les voies biologiques et les fonctions fortement corrélées avec le génotype ou le phénotype, et ainsi de comprendre à l'échelle individuelle, l'effet d'une infection ou d'un fond génétique particulier par exemple. L'intégration des données a été utilisée pour l'étude de systèmes biologiques, d'une maladie ou encore de l'effet du microbiote dans un contexte donné. Cette méthode a aussi été très utilisée et l'ai toujours pour découvrir des marqueurs génétiques ou protéiques d'une maladie. Depuis les dernières années, de nombreux algorithmes ou outils sont régulièrement développés dans le but d'intégrer les données omiques, principalement les données transcriptomiques et protéomiques. Cette analyse de données à haut débit nécessite l'utilisation de méthodes d'étude et de visualisation dédiées qui seront détaillés dans les parties suivantes.

1.3.5.2. Les méthodes utilisées

Lors de l'étude de données générées à haut débit, les représentations simples tels que les courbes ou les barplots ne suffisent plus à les rendre compréhensibles. Il est donc nécessaire d'utiliser des méthodes de visualisation spécialisées avec des logiciels adaptés. Les méthodes de représentation adaptée aux données à haut débit sont une première étape vers l'intégration de données, dont les représentations en composantes principales. Ces méthodes peuvent être complétées par la classification, dite aussi clusterisation, afin de rechercher des profils proches ou similaires. Enfin, la corrélation entre les variables et/ou les individus est une étape essentielle à l'intégration de données.

1.3.5.2.1. Les méthodes d'analyse en composantes principales (ACP, MFA ...)

Les méthodes d'analyse en composantes principales permettent de résumer des données de grandes dimensions tout en maximisant la quantité d'informations retenue dans la présentation finale. Il existe des représentations pour chaque type de données : elles peuvent être quantitatives, qualitatives ou un mélange des deux. Dans le cas des analyses de transcriptome, l'Analyse en Composante Principale ou ACP est la méthode la plus utilisée du fait du caractère quantitatif des données. L'ACP a été développé par Harold Hotelling en 1930, avec l'aide des travaux de Karl Pearson.

Cette méthode permet l'analyse des données tels que des gènes ou des transcrits caractérisant un groupe d'échantillons ou d'individus. Elle permet de résumer l'expression de milliers de gènes en deux ou trois dimensions via la création de nouvelles variables que sont les composantes principales. *In fine*, chaque point présenté dans le nuage de l'ACP correspond à un échantillon. L'ACP va faire en sorte de déterminer les axes qui permettent la meilleure dispersion des échantillons. Pour réaliser l'ACP, les données vont être centrées, c'est à dire soustraite de la valeur moyenne de la variable. Cette étape ne modifie pas la dispersion des échantillons. Les données peuvent aussi subir une réduction, en les divisant par leur écart type. Cette réduction a un effet sur la dispersion, et elle est nécessaire lorsque les variables sont exprimées avec des unités différentes. Cette étape n'est pas obligatoire en fonction de la question posée. Une non-réduction des données peut entraîner une ACP peu informative si certaines variables ont une forte variance, et donc une contribution forte qui peut masquer les autres variables. En revanche une réduction peut transformer des variables peu significatives en données informatives [358].

L'ensemble de l'information correspond à 100% de l'inertie de l'ACP. Le choix du nombre d'axes à considérer peut se faire de diverses manières, soit en considérant un nombre d'axes permettant d'atteindre une inertie cumulée de 80% au moins, soit en utilisant une méthode graphique dite de l'éboulis des valeurs propres. Ce diagramme décrit le pourcentage d'inertie par axe. La méthode du coude consiste à déterminer à partir de quel nombre d'axes, l'inertie diminue plus lentement et de considérer tous les axes en amont de cette inflexion. Un autre critère correspond au critère de Kaiser. Via ce dernier, les axes dont l'inertie est inférieure à la variabilité d'une variable ne sont pas considérés. Cette variabilité est calculée par la formule suivante : 100 divisé par le nombre de variables. Si les deux premiers axes de l'ACP correspondent à 90% de l'inertie cumulée, cela indique que les deux premières composantes principales résument 90% de l'information. Cette représentation est très utilisée dans les analyses de RNA-seq afin d'avoir un résumé global de l'expression en fonction des conditions de l'analyse.

L'AFM ou Analyse Factorielle Multiple est une méthode adaptée pour l'étude des individus ou échantillons caractérisés par plusieurs catégories de variables structurées. Chaque jeu de données introduit pour l'analyse est pondéré afin de donner la même contribution aux variables dans ce jeu. Cette méthode permet d'étudier une question en utilisant différents jeux de données (cliniques, cytokiniques, transcriptomiques ou d'élevage par exemple) et de déterminer

graphiquement les variables qui sont corrélées entre elles. Cette méthode est une extension de l'ACP lorsqu'on utilise uniquement des données quantitatives.

L'ACP et l'AFM sont les méthodes les plus communes parmi les méthodes factorielles. Néanmoins, il existe d'autres méthodes tels que l'Analyse Factorielle des Correspondances (AFC), l'Analyse des Correspondances Multiples (ACM) ou encore l'Analyse Factorielle des Données Mixtes (AFDM). Toutes ces méthodes factorielles reposent sur le même principe général mais elles varient en fonction des données entrantes. L'AFC utilise des tableaux de contingence en entrée, l'ACM un tableau de données qualitatives alors que l'AFDM utilise un tableau de variables qualitatives et/ou quantitatives (**Figure 12**). Ces méthodes peuvent être couplées avec des méthodes de catégorisation pour donner notamment la méthode HCPC qui sera décrite dans la partie suivante.

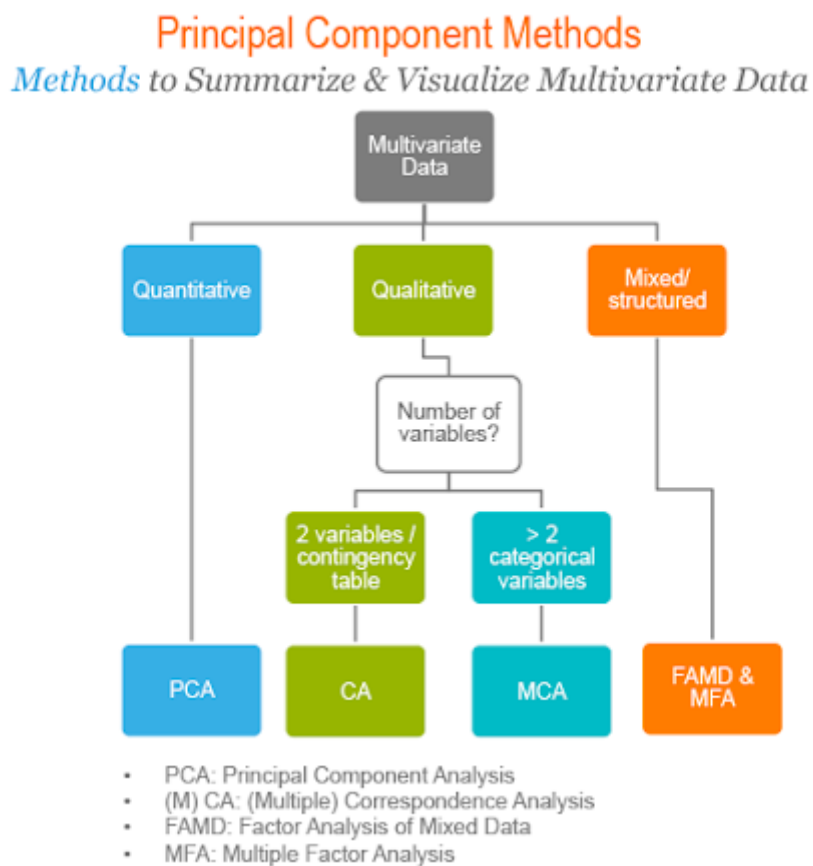


Figure 12 : Schéma résumant les différentes méthodes de réduction et visualisation de données multivariées et leur condition d'utilisation (Kassambara, 2017)

1.3.5.2.2. La classification ou clusterisation (CAH, K-means)

L'objectif principal de la classification est de rechercher les similarités existantes entre échantillons ou individus en utilisant des variables associées. De nombreuses méthodes existent notamment via le développement de l'apprentissage automatique (*machine learning*). Les méthodes peuvent appartenir à deux catégories que sont la classification supervisée et celle non supervisée. La principale différence entre les deux catégories est que pour la première, la connaissance de la classification est souvent disponible avant l'analyse, contrairement à la seconde. Les algorithmes suivants font partie de la classification non supervisée. Les 2 familles principalement utilisées sont les méthodes hiérarchiques tels que la classification ascendante hiérarchique (CAH) et les méthodes centroïdes, comme les K-means. Néanmoins d'autres algorithmes existent mais leur utilisation est moins appropriée pour les données biologiques. Cependant, le calcul de la distance entre les individus est un socle commun entre toutes ces méthodes.

Les classifications hiérarchiques utilisent une matrice de distance calculée au préalable via différents algorithmes pour catégoriser les échantillons ou les individus. La méthode la plus utilisée est la Classification Hiérarchique Ascendante (CAH) qui consiste à partir des individus pour remonter vers les clusters. La matrice peut être calculée en choisissant la distance minimale entre les individus (*single linkage*), la distance maximale (*complete linkage*), ou la distance moyenne (*average linkage*), ou encore la méthode de Ward. Cette dernière est la plus utilisée. Elle permet de regrouper 2 classes en maximisant l'inertie interclasse par l'utilisation de la distance de Ward qui correspond à la distance entre les deux barycentres élevés au carré. Cette distance est pondérée par rapport aux effectifs de chacune des classes. Les méthodes hiérarchiques ont l'avantage de permettre de déterminer le nombre de classes optimal. Néanmoins, elles nécessitent un temps de calcul plus long. Cette méthode est donc adaptée pour des études avec un faible nombre d'échantillons. Le résultat final correspond à un dendrogramme et il peut être découpé pour présenter les différentes classes. Ce découpage peut être réalisé visuellement ou en observant les sauts d'inertie. La littérature suggère de choisir le partitionnement ayant la plus grande perte d'inertie relative.

Les classifications centroïdes, dont la plus connue est la méthode des k-moyennes ou *k-means* nécessitent de choisir au préalable le nombre de classes attendu. Il s'agit de l'inconvénient majeur de cette méthode. Pour débiter cette analyse, il est nécessaire de déterminer au hasard un groupe d'échantillons qui serviront de point de départ. Ces points représentent les classes,

puis l'algorithme va associer chaque individu restant avec la classe qui lui est la plus proche, et ainsi de suite jusqu'à obtenir le nombre de classes choisi. L'avantage de cette méthode est qu'elle nécessite moins de temps de calcul et peut donc être utilisée sur des jeux de données déjà grands. Néanmoins, le résultat de cette catégorisation dépend énormément des individus choisis pour l'initialisation. Pour remédier à ce problème, de nombreux algorithmes réalisent plusieurs itérations afin de garder le partitionnement qui minimise la variance au sein d'une classe.

Les deux méthodes vues précédemment sont les plus courantes dans ce type d'analyse, mais chacune possède des avantages et des inconvénients. Une nouvelle approche a été développée, l'HCPC ou *Hierarchical Clustering on Principal Component* qui combine la méthode de la classification ascendante (CAH), celle des k-means ainsi que les représentations des composantes principales [302]. Celle-ci fonctionne de la manière suivante : la première étape consiste à réaliser une analyse via les méthodes des composantes principales (ACP, AFM, ACM, AFDM), en fonction du type de données à analyser, puis de réaliser une classification hiérarchique. Il est possible de déterminer le nombre de composantes à choisir lors de la première étape. Puis l'algorithme va définir le nombre optimal de catégories et utiliser la méthode des k-means pour améliorer la classification des individus. Cette méthode permet de diminuer le bruit de fond du jeu de données et donc d'obtenir une classification plus robuste.

Une autre catégorie de classifications très utilisée regroupe les méthodes de densité, dont la plus connue est *DBSCAN (Density Based Spatial Clustering of Applications with Noise)*. Cette méthode permet à la fois de regrouper les individus mais aussi de détecter ceux qui s'écartent ou sont aberrants. Elle nécessite au minimum deux paramètres pour être utilisée : la distance maximale entre échantillons pour appartenir à une même classe, ainsi que le nombre minimal d'individus pour former une classe. Puis l'algorithme va créer les classes au fur et à mesure en fonction de ces indications.

Pour conclure, il existe de nombreuses méthodes permettant la classification avec pour chacune, des avantages et des inconvénients en fonction de la nature des données disponibles. Avec le développement du *machine learning*, de nouvelles méthodes se développent telles que les réseaux de neurones qui sont très utilisés pour la classification d'images par exemple. Cette méthode est inspirée des réseaux de neurones biologiques et est utilisée dans le cadre de question d'apprentissage ou de traitement des signaux. Ils sont utilisés dans de nombreux domaines aujourd'hui comme notamment les chatbots. Leur utilisation nécessite des

compétences statistiques et informatiques car une majorité de ces réseaux sont créés via le langage Python.

1.3.5.2.3. L'analyse de corrélation (*WGCNA*, *Mixomics*)

En analyse statistique, la corrélation est une méthode permettant d'étudier la relation entre deux variables. Cette relation peut être corrélée positivement si les deux valeurs évoluent dans le même sens ou négativement si elles évoluent dans des sens opposés. Les trois méthodes d'analyse de la corrélation sont celles de Pearson, de Spearman et de Kendall ayant chacune leurs spécificités. Le coefficient de corrélation de Pearson mesure une dépendance linéaire entre deux variables. Ce test nécessite que la distribution des données soit normale. Pour les méthodes de Kendall et de Spearman, l'analyse est basée sur le rang des données. Ce sont donc des méthodes plus robustes et ne nécessitant pas l'hypothèse de normalité de la distribution des données. Elles sont utilisées par des algorithmes plus poussés pour l'analyse de la corrélation entre des données biologiques tels que *WGCNA* ou encore *MixOmics*. Les logiciels utilisant ces méthodes statistiques seront développés dans la partie à propos des réseaux biologiques.

1.3.5.2.4. La déconvolution cellulaire

La déconvolution cellulaire est une méthode visant à estimer la proportion des différentes cellules au sein d'un échantillon à partir de l'expression génique. En effet, chaque type cellulaire possède un profil d'expression génétique qui lui est propre. De nombreux algorithmes ont été développés en parallèle de l'avènement des techniques omiques. Dans le cadre des études transcriptomiques et notamment du RNA-seq, cette technologie s'avère très intéressante car elle permet d'améliorer l'interprétation des résultats. La majorité des algorithmes de déconvolution utilise une table de référence contenant les informations géniques de différentes populations cellulaires et ont pour objectif d'estimer la contribution de chaque profil de signature au niveau global du signal. Néanmoins, il existe des méthodes dites sans référence dont l'objectif est d'estimer à la fois les signatures de type cellulaire et la contribution de chaque signature (**Figure 13**).

Le but est donc de confronter ces tables de référence avec les tables normalisées issues des analyses avec *DESeq2* ou *EdgeR*. Cette méthode est aussi beaucoup utilisée dans le cadre des études épigénétiques avec des algorithmes dédiés qui ne seront pas détaillés dans cette partie.

La déconvolution cellulaire dans les études transcriptomiques se font majoritairement avec des méthodes basées sur les références. De nombreuses méthodes existent dont les plus utilisées

ont été décrites dans la **table 7**. Chacune de ces méthodes ne peuvent pas être utilisés dans tous les contextes. De plus, certaines ne permettent pas de déterminer de façon optimale, une population cellulaire donnée. Des études de comparaison de méthodes existent permettant de déterminer la méthode la plus adéquate à un contexte, un tissu et une population donnée [359]–[361]. Il est possible d’importer ses propres tables de référence dans de nombreux algorithmes. Les limites de ces analyses concernent la fiabilité de la référence ainsi que le manque de références pour certaines populations cellulaires. En effet, la référence est sensible à de nombreux critères tels que le tissu ou la maladie étudiée. De plus, il existe de nombreux types cellulaires dont leur profil génique n’est pas disponible ou ne l’est que partiellement. L’arrivée des références basées sur les méthodes de single cell RNA-seq a permis partiellement encore d’améliorer ces limites. Une autre limite existe et concerne le détail des types cellulaires. Il est encore impossible de déterminer les sous-types de populations immunitaires, tels que les lymphocytes T CD4+ Th17 par exemple au sein des échantillons.

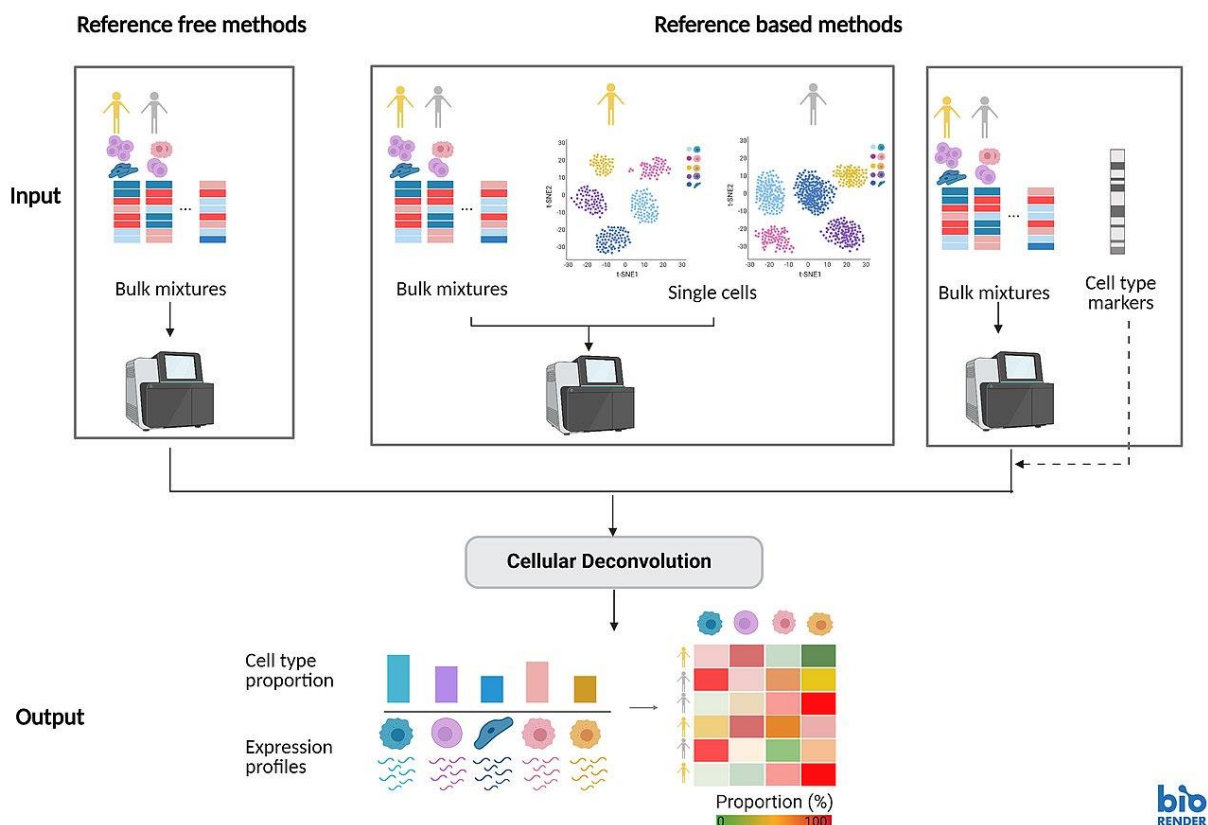


Figure 13 : Fonctionnement de la déconvolution cellulaire

Table 7 : Recapitulatif des méthodes de déconvolution cellulaire les plus utilisées dans l'étude du système et de la réponse immunitaires

Méthode	Types cellulaires	Categorie	Année	Référence
CIBERSORT	22 types cellulaires	Basée sur la référence	2018	[362]
EPIC	5 types cellulaires	Basée sur la référence	2017	[363]
xCells	64 types cellulaires	Basée sur la référence	2017	[364]
Mcp-counter	9 types cellulaires	Basée sur la référence	2016	[365]
Quantiseq	11 types cellulaires	Basée sur la référence	2019	[366]
DeconvSeq	4 types cellulaires	Basée sur la référence	2019	[367]
CDseq		Sans référence	2019	[368]

1.3.5.3. L'analyse à l'échelle des fonctions biologiques

1.3.5.3.1. Identifiant génique

Dans le domaine de la biologie, les gènes peuvent posséder plusieurs appellations redondantes qui leur ont été données au fur et à mesure des progrès réalisés sur la connaissance de leurs fonctions. Par ailleurs, cette dénomination est quelquefois propre à chaque espèce et il est alors difficile de faire un parallèle entre elles ; cette difficulté est encore plus marquée dans le cas des protéines. De plus, certaines confusions existent en confondant le nom du gène avec celui de la protéine.

Le comité HUGO a été chargé d'assigner un nom unique à chaque gène, à dire d'experts. Ce comité a permis la création de la base de données HGNC (<https://www.genenames.org/>) contenant les informations relatives à chaque gène avec le nom approuvé et les noms alternatifs, pour l'espèce humaine. Les gènes sont assez souvent caractérisés par des symboles décrivant

leurs fonctions comme par exemple CXCL10 (C-X-C motif chimiokine ligand 10) qui correspond à un gène codant pour une chimiokine. Dans la base HGNC, il est aussi possible d'obtenir les noms des gènes homologues dans les autres espèces (souris, rat, vache ...). Avec le développement des bases de données, d'autres identifiants de gènes sont apparus tels que l'identifiant *Ensembl* ou encore *Gene* (NCBI). Ensembl ID [369] a été développée par l'institut de bioinformatique européen et peut s'appliquer à des gènes, des transcrits voire des protéines. Il permet d'avoir un identifiant sûr et vérifié. C'est celui qui est utilisé pour les alignements en utilisant un génome de référence. Il est composé de 3 parties : la première correspond au préfixe de l'espèce (ENSBTA pour *Bos taurus*), suivi d'un suffixe correspondant au type de fonction (E : Exon, G : gene ou P : Protein), puis par l'identifiant numérique spécifique du gène ou de la protéine (avec 11 chiffres). Cette nomenclature permet d'avoir un identifiant unique quel que soit l'espèce. L'autre identifiant très utilisé est le Gene ID développé par NCBI [370]. Il correspond à un numéro spécifique du gène dans une espèce donnée (CXCL10 humain : 3627, CXCL10 bovin : 615107).

Grace à des logiciels comme le logiciel *Biomart* [371], la conversion des symboles de gènes en identifiants *Ensembl* et inversement, est possible permettant de réaliser les analyses fonctionnelles entre espèces, en utilisant l'information biologique bien plus abondante pour des espèces comme l'homme ou la souris par rapport aux espèces domestiques.

1.3.5.3.2. Analyse de fonction biologique

Aux prémices de l'étude du transcriptome, l'absence de base de données permettant l'étude des gènes identifiés comme d'intérêt était un réel problème. Pour résoudre cet écueil, le projet Gene Ontology a débuté en 1998 dans le but de développer des bases de données pour décrire les fonctions associées aux gènes et fournir des moyens techniques-pour cette analyse. Ce projet a entraîné le développement du langage GO afin d'étudier des phénomènes biologiques en s'affranchissant de la barrière d'espèce.

L'étude des fonctions des gènes via ce langage est divisée en trois axes que sont leur présence dans les composants cellulaires, leur rôle comme fonction molécule notamment dans le cas des enzymes, et enfin leurs fonctions biologiques. Chaque phénomène biologique est défini par un « terme GO » dont chacun contient un identifiant unique, le nom du phénomène biologique, les bases de données de référence de ce terme, et enfin les gènes ou fonctions associés à celui-ci. L'ensemble de ces informations ont permis le développement des analyses d'enrichissement de fonctions. A partir d'un ensemble de gènes considérés différentiels, il est possible d'étudier les

fonctions biologiques les plus altérées par la modification d'expression de ces gènes et de comprendre ainsi dans quel processus biologique interviennent ces gènes différentiels. Cette méthode d'enrichissement est divisée en plusieurs étapes. Une fois les gènes différentiels obtenus via les packages *EdgeR* ou *DESeq2*, leurs identifiants sont utilisés dans des algorithmes et mis en relation avec des bases de données pour sélectionner les processus biologiques (Processus biologiques, composant cellulaire ou fonction moléculaire) auxquels ils appartiennent et qui sont susceptibles d'être modifiés. Ces logiciels utilisent des méthodes statistiques similaires à ceux utilisés pour la détermination des gènes différentiels.

Plusieurs outils ont été développés en utilisant ces bases de données tels que *TopGO* [372], *clusterProfiler* [373] ou *Funrich* [374]. Ce système de termes GO possède des limites comme celle d'avoir des fonctions générales et peu spécifiques pour chaque type cellulaire ou organe. Néanmoins pour pallier ce problème, de nombreuses autres bases de données GO ont été développées avec chacune leur spécificité.

Aujourd'hui, le nombre de processus biologiques couverts par les termes GO est de 28748, avec 11153 concernant les fonctions moléculaires, et 4184 pour les composants cellulaires. Cet ensemble de termes est issu de l'annotation de 7 931 218 études dans 4 743 espèces, dont 204 espèces comprenant plus de 1000 annotations. Dans le domaine particulier de l'étude de la réponse immunitaire, les termes GO liés aux processus biologiques et notamment « *immune system process* » sont très intéressants. Les termes GO immunitaires correspondent à 892 processus biologiques immunitaires et associés à 2 875 gènes et issus de 9 606 annotations permettant des études globales mais aussi plus spécifiques de la réponse immunitaire [375]. Toutes les méthodes d'enrichissement de fonctions sont plus ou moins liées aux termes GO. Une autre des limites de la méthode GO est liée à la sélection des gènes différentiels qui est faite de manière arbitraire, notamment dans le choix du seuil de signification statistique (même si la valeur de 5% est la plus souvent admise), mais aussi de l'amplitude de la variation de l'expression ou de la différence entre conditions et pour laquelle aucune valeur seuil n'est admise et qui varie beaucoup selon les études. Les limites de cette méthode sont la pertinence de l'information qui peut ne pas toujours être vraie, ainsi que les sources des informations qui peuvent être erronées. Heureusement, un travail de vérification est réalisé en permanence, dans des bases de données vérifiées et permet de réduire ces erreurs.

1.3.5.3.3. GSEA

GSEA pour *Gene Set Enrichment Analysis* est une méthode computationnelle développée en 2003 dans le but de permettre une étude fonctionnelle des données transcriptomiques [376]. Cette méthode utilise une base de données de signatures moléculaires (MSigDB) [377] qui est divisée en neuf sets majeurs permettant une étude spécifique de ce type de données. Un des sets de cette base de données contient notamment les termes GO. Très utilisée au début des analyses de transcriptome avec les puces, son utilisation s'est maintenue pour l'analyse des données de RNA-seq. Au contraire de la méthode GO, les données utilisées par cette méthode d'analyse correspondent à l'ensemble des gènes différentiels obtenus via *DESeq2* ou *EdgeR* qu'il soient significatifs ou non. Ce choix de données est possible car les gènes peu exprimés ne contribuent pas au calcul du score d'enrichissement (ES). Cette note est calculée en classant tout d'abord les gènes en fonction de leur expression via un test statistique similaire au test de Kolmogorov-Smirnov. La significativité de ce score est estimée puis normalisée, et un test post-statistique comme ceux détaillés précédemment, est utilisé. La plupart du temps, il s'agira d'une correction de Benjamini-Hochberg afin d'obtenir un nombre maximal de fonctions. Les gènes sont ensuite comparés à la base de données MSigDB pour déterminer les fonctions auxquelles ils appartiennent. Il est possible de choisir un jeu de données au sein de cette base afin d'obtenir des résultats spécifiques en fonction du contexte de l'étude et de la nature des phénomènes biologiques qui sont étudiés. Récemment, un package supplémentaire a même été développé pour l'analyse du scRNA seq.

Plusieurs outils complémentaires ont été développés pour effectuer ces analyses. Certains sont disponibles via des interfaces web tels que *WebGestalt* [378], *David* [379] et *ToppGene* suite [380] ; d'autres sont des bibliothèques de R comme *gprofileR2* [381] ou *TopGO* [372]. De nombreux logiciels ont des utilisations spécifiques tels que *PlantRegMap* développés pour l'étude des plantes [382].

Une alternative *SEA* pour *Simpler Enrichment Analysis* a été proposée avec une simplification de l'algorithme. Une autre limite concernant *GSEA* est l'obtention de résultats qui lui sont propres et qui dépendent fortement de l'algorithme utilisé pour grouper les gènes. La méthode *Spectral Gene Set Enrichment* (SGSE) a été proposée avec un test non supervisé. *ssGSEA* est une extension de *GSEA* permettant de calculer le score d'enrichissement pour un échantillon donné et un gène donné. Cette méthode est notamment utilisée par le logiciel *Reactome* afin d'identifier les fonctions affectées [383].

1.3.5.3.4. Modules et analyse modulaire

En parallèle du développement de GO et de GSEA, d'autres méthodes suivant une approche légèrement différente ont été développées ; il s'agit de l'analyse modulaire. Cette méthode est principalement appliquée aux études transcriptomiques réalisées à partir d'échantillons sanguins. Cette approche fut initialement développée par Damien Chaussabel en 2009 [384]. Elle a conduit à l'élaboration de modules spécialisés pour l'étude de la réponse immunitaire. Chacun de ces modules est caractérisé par un identifiant de module (exemple : M1.1), par le nombre de gènes qu'il contient, ainsi que par la fonction générale ou spécifique du module. Ces derniers ont été développés avec l'étude transcriptionnelle de cellules immunitaires présentes dans le sang (comme les PBMC) dans divers contextes infectieux ou cancéreux. Les gènes ont ensuite été regroupés en catégories via leur profil d'expression par une classification basée sur la méthode des k-means. Les modules ont ensuite été consolidés par un processus itératif en vérifiant leurs caractéristiques lors de différents contextes infectieux (*S. aureus*, *Influenza A*) ou immunitaires (mélanome, diabète). Une fois déterminée, leur fonction générale a été décrite. Au total, 28 catégories ont été décrites et elles correspondaient encore à des fonctions très générales. En 2014, Li [385] a développé de nouveaux modules plus spécifiques. Le champ de ces modules n'est plus seulement immunitaire, il est aussi lié à la localisation, à la fonction moléculaire, à la signalisation et aux processus biologiques qui sont associés au module. Finalement, 607 modules ont été développés, avec plus de 300 modules ayant des fonctions précisément déterminées. Trente-neuf correspondent aux modules de Chaussabel avec des fonctions plus larges que ceux décrits par Li. Chacun de ces modules contient entre 9 et 347 gènes. Quarante-neuf d'entre eux correspondent à des processus biologiques (cycle cellulaire, adhésion, ...), 103 à des fonctions immunitaires, 25 à la localisation cellulaire, 48 à des fonctions moléculaires et 15 à leur rôle dans la signalisation.

L'étude de fonctions à l'aide de ces modules est possible via un package R baptisé *Tmod* [386]. Chaque module est représenté par un disque dont l'échelle colorimétrique indique le niveau de la valeur statistique (p-value) de la fonction, et sa taille, la proportion de gènes activés ou inhibés parmi le nombre de gènes total du module en question. Cette méthode permet d'étudier les processus immunitaires très précisément dans un contexte infectieux, et permet de résumer l'ensemble des fonctions activées sous la forme d'une seule représentation graphique [387]. Au cours de l'année 2020, une nouvelle version de ces analyses modulaires a été publiée [388]. Elle comprend 382 modules annotés au total, et fonctionne avec le langage R. Elle permet de réaliser une comparaison 2 à 2 entre groupes, mais aussi entre les individus au sein des groupes avec une représentation sous la forme d'une heatmap. L'analyse de l'expression différentielle

des modules, ainsi que celle des gènes dans chacun des modules, est plus aboutie permettant une analyse approfondie par rapport à la version précédente de ce logiciel. Cette méthode a été utilisée dans le cadre d'une étude de l'effet des neutrophiles sur le psoriasis à partir de différents jeux de données disponibles dans la littérature et elle offre des perspectives très intéressantes [389].

1.3.5.3.5. Etude des voies de signalisation et d'activation cellulaire

L'étude des processus biologiques via GSEA, GO ou l'analyse modulaire permettent d'apporter de nombreuses informations à partir de la liste des gènes différentiels entre conditions. En parallèle des méthodes précédentes, il est possible de déterminer l'implication des voies de signalisation et d'activation cellulaires. Celles-ci sont définies comme une série d'interactions entre des molécules afin de permettre la synthèse d'un produit ou d'entraîner un changement d'état de la cellule. Ces voies ou *pathways* sont impliquées dans le métabolisme, la régulation de l'expression génique ainsi que dans les transmissions des signaux au sein de la cellule et entre cellules. Dans le domaine d'étude de l'immunité, ces voies conduisent notamment à la production de molécules telles que les chimiokines et les cytokines qui jouent un rôle dans le recrutement cellulaire ou encore l'activation des cellules immunitaires.

Plus de 200 bases de données sur ces processus sont disponibles avec pour chacune, des spécificités. Voici les bases les plus génériques et les plus couramment utilisées [390].

- *KEGG pathway* [391] est un ensemble de bases de données liés aux voies métaboliques et aux composés biochimiques.
- *Wikipathways* [392] est une ressource communautaire dédié aux pathways biologiques. Ce site contient différents types de pathways (métabolique, signalisation) ainsi que plusieurs espèces (Homme, souris, levure ...)
- *PANTHER* [393] est un système de classification basé sur des bases de données biologiques. Ce système fait partie du projet *Gene ontology*. Il permet d'analyser des listes de gènes ou de protéines, d'obtenir des informations sur un gène ou une protéine donnée et d'identifier des fonctions de gènes non caractérisés dans certaines espèces via les relations évolutives avec des gènes déjà connus.
- *Reactome* [394] est à la fois une base de données multi-espèces et un logiciel permettant l'analyse de pathways. Il contient les différents pathways de signalisation, métabolomiques, etc.

- *InnateDataBase* [395] est une base de données publiques sur les gènes, protéines et pathways liés à la réponse immunitaire, en particulier son versant inné. Tout comme *Reactome*, il permet aussi de réaliser des analyses de pathways à partir d'une liste de gènes.

1.3.5.3.6. Réseaux biologiques

L'étude des réseaux biologiques constitue une autre approche pour l'analyse des données de l'analyse transcriptomique. Un réseau biologique est un ensemble de constituants qui sont liés entre eux par des relations plus ou moins directes. De nombreux réseaux existent en biologie tels que les réseaux d'interactions protéine-protéine, de signalisation ou encore les réseaux des réactions métaboliques. Les réseaux de co-expression (GCN) sont des représentations où chaque nœud est un gène, et ces nœuds sont potentiellement reliés entre eux s'il existe une co-expression. Ces GCN sont très utilisés avec les données transcriptomiques à haut débit. Les relations entre gènes peuvent être des corrélations positives ou négatives et sont calculées en utilisant les méthodes de corrélation décrites précédemment telles que la méthode de corrélation de Pearson, et celle de Spearman. Des méthodes de calcul de distances peuvent aussi être utilisées pour étudier les relations entre deux gènes comme la distance euclidienne. Les 2 logiciels majoritaires pour l'analyse de la co-expression et des fonctions sont *CEMITool* et *WGCNA*.

L'ensemble des logiciels qui vont être présentés utilisent des approches statistiques pour la détermination de fonctions. Les approches les plus utilisées sont le test hypergéométrique et le score d'enrichissement comme le GSEA.

Le test hypergéométrique est une méthode basée sur la loi hypergéométrique pouvant être utilisé dans le cas où l'on veut déterminer la probabilité d'obtenir une ou plusieurs fonctions en simultanée. Cette loi est très similaire à la loi binomiale mais applicable à plusieurs fonctions en simultanée. Les gènes différentiels obtenus sont comparés à la base de données et une probabilité sous forme de p-value est calculée. L'objectif de GSEA est le même que le test hypergéométrique, c'est-à-dire déterminé si une fonction est enrichie via l'expression des gènes différentiels. Pour cela, il calcule une statistique dite locale pour chacun des gènes et une liste de gènes est ordonnée selon la valeur obtenue quelle soit positive ou négative. Ces listes sont ensuite comparées à une base de données contenant les fonctions et les gènes associés, et un score est calculé par la somme des statistiques locales des gènes présents dans la fonction.

La méthode *WGCNA* (pour *Weighted Gene Co-expression Network*) fut la première méthode développée pour l'analyse de la co-expression des gènes [396]. Elle permet de créer un réseau

de gènes co-exprimés en réalisant une étude de corrélation (Spearman) entre ceux-ci. Une identification des modules par classification hiérarchique est utilisée. Les modules sont ensuite mis en relation avec des informations externes, comme par exemple des données cliniques ou protéomiques ainsi qu'avec des données sur les fonctions liées aux gènes. Des algorithmes sont ensuite utilisés pour mettre en relation les modules entre eux et déterminer s'il existe des liens entre ceux-ci. *In fine*, il peut permettre d'identifier des biomarqueurs via des tests de causalité. Alors que cette méthode a été développée il y a plus de 15 ans, elle est encore très utilisée aujourd'hui car elle est la plus aboutie et permet l'analyse du plus grand nombre de fonctions. Elle présente cependant quelques limites qui seront énumérées lors de la discussion sur le principal concurrent qui est le logiciel *CEMITool*.

CEMITool (pour *Co-expression modules identification Tool*) est une méthode permettant d'identifier les co-expressions entre gènes [397]. Ce logiciel contient une méthode innovante de filtration non supervisée des gènes basée sur la distribution gamma inversée. Il permet aussi de faire des analyses d'enrichissement de fonctions via la méthode *GSEA* et de réaliser une analyse *ORA* (pour *Over Representation Analysis*) permettant de déterminer les modules de fonctions les plus significatifs via un test hypergéométrique par l'utilisation du package *clusterProfiler* [373]. Ce logiciel a beaucoup de similarités avec *WGCNA* telles que l'analyse *ORA*, et l'association de modules de fonctions avec des phénotypes. En outre, il permet aussi d'identifier les interactions entre les gènes au sein de chaque module de co-expression en utilisant diverses bases de données externes (*Reactome*, *ToppFun*, *Ingenuity*). Le logiciel exporte l'ensemble de ces analyses sous la forme d'un résumé au format HTML. Comparé à ce logiciel, les possibilités d'analyse des fonctions sont plus réduites avec *WGCNA* et les données doivent être préalablement préparées notamment avec *DESeq2* (Normalisation, Suppression des gènes non détectés).

String (*Search tool for the retrieval of interacting genes/proteins*) est une base de données biologique et un logiciel permettant d'étudier les interactions protéine-protéine. Celle-ci contient des informations issues de multiples sources comme GO, KEGG ou encore Réactome. Elle contient des informations sur plus de 5000 organismes et plus de 3 milliards d'interactions. L'étude des réseaux d'interactions « prédits » peut permettre de comprendre des processus biologiques dans un contexte particulier. Ce logiciel peut être utilisé pour étudier les interactions prédites dans une nouvelle espèce via l'orthologie de ces réseaux. Il permet aussi d'étudier les gènes et *in fine* les fonctions dans lesquels ils interviennent au sein de ces réseaux. L'atout majeur de ce programme est la représentation des réseaux qui est personnalisable via le

logiciel *Cytoscape* [398]. Au contraire des précédents, celui-ci ne permet pas de faire des études de corrélation avec des données biologiques externes.

Ingenuity ou *IPA* (pour *Ingenuity Pathway Analysis*) est un logiciel développé et commercialisé actuellement par la société QIAGEN. Il permet l'analyse, l'intégration et l'interprétation de données omiques au sein d'une seule et même plateforme [399]. Son objectif est d'identifier les voies de signalisation ou métaboliques, d'étudier des réseaux de protéines et les fonctions biologiques associées. Il permet aussi de prédire l'activation ou l'inhibition des facteurs de transcription à partir de l'expression des gènes tout comme *Reactome* ou *ClusterProfiler*. Enfin, il est possible de comparer les voies et les processus biologiques modifiés entre différentes conditions. Ce logiciel permet d'étudier les gènes différentiels obtenus après une analyse avec *DESeq2* ou *EdgeR*, et il est possible de comparer différentes analyses 2 à 2 afin d'étudier différentes conditions biologiques en parallèle. Le problème majeur de ce logiciel est le coût de la licence d'utilisation qui peut être prohibitif à l'échelle d'une équipe, au contraire de ceux cités précédemment.

Parmi les logiciels libres de droit, *Reactome* est, comme vu auparavant, une base de données mais il permet aussi de faire des analyses de fonctions. La première de celles-ci est l'étude des voies biologiques dans de nombreuses espèces comme l'homme mais aussi l'espèce bovine, dans notre cas. Cette base de données contient des informations sur des voies liées à la réponse immunitaire, au métabolisme, au cycle cellulaire et aux maladies. La seconde fonction de ce logiciel est l'analyse de données sous la forme d'une liste de gènes, et dans ce cas, il est possible d'utiliser tous les identifiants alternatifs d'un gène. L'analyse peut aussi passer par une analyse d'expression et le logiciel utilise alors les tables de comptage produites via trois méthodes : la méthode *ssGSEA* décrite précédemment, la méthode *Camera* qui correspond à un algorithme similaire à *GSEA* implanté dans le package *limma* ou encore l'algorithme *PADOG* (pour *Pathway Analysis with Down Weighting of Overlapping Genes*) qui est une méthode basée sur le calcul d'un score (t-score), et qui est pondérée selon la fréquence d'apparition des gènes dans les pathways.

Les méthodes *PADOG* et *Camera* permettent des études fonctionnelles en comparant 2 groupes d'échantillons entre eux, alors que la méthode *ssGSEA* établit la liste des fonctions enrichies au sein de chaque échantillon.

Mixomics est un package développé pour R qui permet l'exploration et l'intégration de données biologiques avec une sélection possible des variables à prendre en compte [400]. Ce package est très complet car il ne contient pas moins de 19 méthodes d'analyse développées

spécifiquement pour les données omiques. L'avantage de ce logiciel est la diversité et la qualité de la représentation graphique pour les différentes variables. Les méthodes utilisées varient selon le but et la nature du jeu de données. Les outils employés pour l'exploration d'un seul jeu de données sont l'ACP et la sparse ACP. La sparse ACP est une méthode permettant d'éliminer les données superflues et donc de permettre d'avoir une analyse plus simple à interpréter. En effet, elle utilise la méthode lasso qui permet de sélectionner des ensembles de variables. Cette méthode peut être utilisée sur de nombreux modèles statistiques. Cet outil permet aussi l'intégration de données pour deux jeux correspondant à des mesures sur les mêmes individus via l'utilisation des méthodes de « *Canonical Correlation Analysis (CCA)* ou de « *Partial Least Squares (PLS)*. La PLS est une méthode de régression basée sur la covariance permettant de maximiser la variance. L'avantage de cette dernière par rapport à l'ACP notamment, est l'impact moins marqué des données manquantes. *Mixomics* permet aussi de classifier des groupes d'individus via des méthodes discriminantes tels que la *PLS-DA* et la *sparse-PLS-DA*. Enfin, ce package contient des fonctions permettant l'analyse de données répétées avec des sources multi-niveaux. La fonction *DIABLO* implantée dans ce package est une méthode de réduction de dimensions multivariée et ne nécessite pas d'hypothèse. L'algorithme construit des combinaisons de variables et maximise la corrélation entre les types de données afin d'identifier un petit ensemble de marqueurs.

2. Etude expérimentale

2.1. Objectif de la thèse

Le choix de réduire l'utilisation des antibiotiques pour les réserver à la santé humaine fait partie des évolutions récentes de notre société avec des répercussions fortes sur l'élevage. Il a pour but de réduire le développement de bactéries résistantes aux antibiotiques, qui rendrait des infections banales, difficiles à soigner. Dans le contexte actuel de la prévention de l'antibiorésistance, le développement de méthodes alternatives pour lutter contre les maladies est nécessaire. L'amélioration des conditions d'hygiène générale et l'amélioration des contextes dont on sait qu'ils sont propices au développement des maladies ont déjà porté leur fruit. Néanmoins, ils ne sont pas suffisants et doivent être complétés par des méthodes prophylactiques, comme la vaccination. C'est dans ce cadre que s'inscrit mon projet de thèse. L'objectif est d'améliorer les connaissances vis-à-vis de la vaccination pour la prévention de maladies très présentes chez les veaux pour les infections pulmonaires et les vaches laitières pour les mammites. La méthodologie et la collecte de nouvelles informations s'appuient sur l'utilisation de méthodes à haut débit et les outils de biostatistique appliqués à ces ensembles nouveaux et particuliers de données.

2.2. Etude 1 : Etude transcriptomique des mécanismes de protection permis par la vaccination vis-à-vis d'une mammites à *E. coli* chez la vache laitière

Article publié dans *npj Vaccines*

2.2.1. Objectif et protocole

Les mammites sont la cause numéro un d'utilisation des antibiotiques chez la vache laitière. Ces infections sont principalement dues à *S. aureus* à l'origine de mammites sub-cliniques, *S. uberis* et *E. coli* causant des mammites cliniques, parfois sévères. Les méthodes prophylactiques tels que la vaccination vis-à-vis de ces pathogènes, et notamment d'*E. coli* sont encore peu efficaces. Cette faible efficacité a pour origine les zones d'ombres concernant les connaissances de la réponse immunitaire permettant une protection vis-à-vis de cette bactérie.

Les objectifs de ce projet sont

- d'étudier deux protocoles de vaccination utilisant des voies d'immunisation différentes, le premier utilisant la voie d'immunisation intramusculaire seule, et le second associant une immunisation par voie intramusculaire à la voie mammaire, en vue de protéger la glande mammaire vis-à-vis d'une infection par *E. coli*.
- d'analyser les mécanismes immunitaires induits par la vaccination qui permettent la protection des animaux vis-à-vis de cette infection.

Cette étude fait suite à un premier travail décrivant l'étude expérimentale, l'évolution clinique après une épreuve expérimentale de mammites avec la même bactérie que celle utilisée pour l'immunisation, qui a été précédemment publiée (Herry, et al 2017).

Dans le but de connaître les effets de ces deux protocoles d'immunisation, un essai expérimental sur 18 vaches au total a été réalisé. Trois groupes de 6 animaux ont été constitués :

- Un groupe *témoin* : injection de l'adjuvant seul (Adj) à deux reprises et au même moment que les injections vaccinales dans les autres groupes
- Groupe *Intramusculaire* : Immunisation avec des bactéries *E. coli* P4 en présence d'un adjuvant par voie intramusculaire avec une deuxième injection dans les mêmes conditions réalisée 60 jours plus tard.
- Groupe *Intramammaire* : La première immunisation est la même que celle réalisée pour le groupe intramusculaire (*E. coli* P4 + adjuvant) alors que la deuxième intervention est

effectuée par voie intramammaire avec un surnageant de culture d'*E. coli* concentré par ultracentrifugation.

Dans le premier article publié par Herry et al., il avait été montré que les animaux du groupe *intramammaire* avait une quantité de bactéries dans le lait à la suite de l'inoculation qui était plus faible que celle des autres groupes 40 heures après infection (hpi). De plus, les animaux immunisés quelles que soient les modalités utilisées ont une diminution de la production de lait moins forte, un score clinique systémique plus faible et une leucopénie moins forte au pic de l'infection (à 12 hpi).

Dans l'étude présentée ici, une analyse cinétique du transcriptome à trois dates (0h correspondant au temps de référence, 12h soit à l'acmé de la phase aiguë, et 40h correspondant au début de la phase de résolution de l'infection et de l'inflammation associée), et une mesure de la production de cytokines ont été réalisés sur des échantillons de sang et de plasma, respectivement. En outre, un examen du transcriptome sanguin a été réalisé à 3 dates de la phase d'immunisation et 7 jours après la deuxième immunisation. Enfin, pour analyser la réponse locale induite par l'immunisation mammaire, la purification et l'analyse de l'expression des gènes de lymphocytes T CD4 du tissu mammaire a été réalisée 16 hpi.

2.2.2. Résultats

2.2.2.1. Transcriptome et protéome sanguin, des méthodes efficaces pour l'étude d'une mammite ?

Notre premier objectif était de déterminer si l'analyse du transcriptome sanguin permettait d'étudier la réponse immunitaire associée à l'infection mammaire par *E. coli*. L'étude du profil transcriptomique à l'aide d'une analyse en composantes principales et d'une représentation heatmap a montré une forte différence entre les trois dates au cours de l'évolution de la mammite. Via une étude des fonctions mobilisées, utilisant les bases de données GSEA, nous avons pu montrer une activation des fonctions liés au *recrutement des neutrophiles* et à la *reconnaissance de pathogènes bactériens* à la phase précoce de la mammite, soit 12 hpi. Concernant la phase tardive soit 40 hpi, ce sont plutôt les fonctions liées à *l'activation des macrophages* et à la *réponse humorale* qui sont présentes. Au pic de l'infection, une augmentation de la production des chimiokines CCL-2, CCL-4, CXCL10 et des cytokines inflammatoires (TNF- α , IL-6) est présente en lien avec la réponse inflammatoire mammaire.

Ces résultats montrent de façon originale qu'il est possible de suivre l'évolution d'une mammite provoquée par *E. coli* grâce à l'analyse du profil de l'expression des gènes dans le sang. En

outre, il existe une forte augmentation de la production de cytokines et de chimiokines, dont il est difficile de savoir si elles proviennent du tissu mammaire, ou si elles sont produites dans d'autres sites (comme le foie) en réponse à l'inflammation mammaire, ou encore à la diffusion de molécules phlogogènes.

2.2.2.2. La vaccination, un effet au niveau systémique ?

Le nombre de gènes différentiels entre les temps 0 et 40 heures post-infection n'est pas le même entre le groupe Intramusculaire (IM) et le groupe Intramammaire (IMM), avec un beaucoup plus petit nombre de gènes différentiels dans ce dernier groupe, indiquant un retour du profil d'expression à l'état initial, ce qui peut être interprété comme une guérison plus rapide. Concernant les fonctions activées, elles sont similaires à celles observées lors de l'étude du groupe témoin hormis une activation de la réponse *interféron* dans les groupes vaccinés à 40 hpi. L'effet de l'infection est beaucoup plus marqué que celui lié à la vaccination si l'on considère le nombre de gènes différentiels identifiés.

Concernant la réponse cytokinique, une forte activation de la réponse interféron (présence de quantités augmentées d'IFN γ et de CXCL10) est observée dans le groupe Intramusculaire, et dans une moindre mesure dans le groupe *Intramammaire*. Néanmoins, la réponse inflammatoire (avec IL-1 β , IL-6 et TNF- α) est moins marquée dans les groupes vaccinés. A l'aide d'une étude de corrélation, nous avons montré que cette production d'interféron était corrélée au score clinique mammaire indiquant un possible effet négatif. A ce stade de l'analyse, il n'y a pas de différence de la réponse liée à l'IL-17.

2.2.2.3. Quel effet de l'immunisation sur la réponse systémique ?

Pour identifier les mécanismes liés à la vaccination, nous avons étudié la réponse transcriptomique au moment de la phase d'immunisation, et en particulier au 7^e jour après la première (I7), et après la deuxième injection quelle que soit la voie utilisée (B7). Nous avons constaté une augmentation des gènes différentiels dans les deux groupes vaccinés, mais le profil de réponse est différent puisque seuls 30% des gènes différentiels dans chaque groupe sont partagés, indiquant que 2/3 de la réponse repose sur des gènes et des fonctions différentes. Une étude de la réponse immunitaire via l'utilisation des modules définis par Chaussabel et Li ont montré de nettes différences entre les 2 groupes vaccinés. En effet, une activation du module lié à la *réplication des lymphocytes T CD4* est identifiée à I7 dans le groupe Intramammaire. Concernant la réponse à B7, les modules liés aux neutrophiles et aux monocytes sont communs, alors que les modules liés à la *signalisation cellulaire* et notamment aux TLR sont exclusivement présents dans le groupe Intramammaire.

2.2.2.4. La réponse des neutrophiles, une fonction indispensable lors d'une mammite ?

Lors de l'analyse approfondie des données cliniques et transcriptomiques, la moitié des animaux du groupe Intramammaire ont montré une maladie plus bénigne indiquant que la vaccination avait un effet protecteur plus marquée dans ce groupe. Pour en comprendre les mécanismes, une étude comparant les animaux dont la guérison était la plus rapide aux congénères du même groupe, nous avons pu identifier deux fonctions spécifiquement activées dans les premiers. Ces réponses sont liées à la *signalisation cellulaire liée aux TLR* ainsi qu'à la *réponse des neutrophiles*.

2.2.2.5. L'étude de l'immunité locale, un incontournable pour comprendre les mécanismes de protection ?

En lien avec la réponse adaptative, la réponse et le recrutement des neutrophiles peuvent être induits par les lymphocytes T CD4+ via la production de certaines cytokines comme l'IL-17A. L'étude des voies de signalisation de ces cellules purifiées à partir du parenchyme mammaire des vaches immunisées et infectées a montré une forte activation des voies liées aux cytokines (IL-10, TNF- α et IL-17A) dans les lymphocytes T CD4+ des animaux du groupe Intramammaire par comparaison à ceux du groupe témoin. Cette activation cytokinique est plus marquée au sein des lymphocytes T CD4+ après l'infection mammaire indiquant qu'ils sont activés dans ce contexte. En parallèle, nous avons montré une augmentation légère de la concentration d'IL-17A dans le tissu et le lait.

2.2.3. Conclusion générale

L'étude du transcriptome sanguin à travers une cinétique constituée de plusieurs temps identifiés comme critiques au cours du déroulement de la mammite, a permis d'étudier la réponse systémique associée à une infection locale comme la mammite provoquée par *E. coli*. La vaccination lorsqu'elle est effectuée uniquement par voie intramusculaire n'a pas eu d'effet bénéfique sur l'évolution et l'issue de l'infection, par comparaison aux groupes témoin et intramammaire, possiblement à cause de l'induction d'une forte réponse interféron qui apparaît délétère. La réponse associée à l'IL-17A produite par les lymphocytes T CD4+ permettant le recrutement des neutrophiles dans le parenchyme mammaire semble être la réponse à privilégier par la vaccination pour protéger efficacement les animaux de la mammite provoquée par *E. coli*.

ARTICLE OPEN



Th17-related mammary immunity, but not a high systemic Th1 immune response is associated with protection against *E. coli* mastitis

Nathan Cebron¹, Sarah Maman², Sarah Walachowski¹, Blandine Gausserès¹, Patricia Cunha³, Pascal Rainard³ and Gilles Foucras¹✉

Vaccination against bovine mastitis lags behind despite high demand from the dairy industry and margin for efficacy improvement. We previously compared two immunization protocols against *E. coli* using either only the intramuscular route or a combination of intramuscular and mammary ductal routes, also known as 'prime and pull' strategy. A homologous mammary challenge during the memory phase showed that immunization favorably modified the mastitis course, notably in locally immunized cows in comparison to intramuscular and control adjuvant-only groups. Here, we performed whole-blood profiling through RNA-seq transcriptome and plasma cytokine 15-plex analyses at time points of the *E. coli* mastitis that showed significant clinical and laboratory differences among the groups. Diminished production of inflammatory cytokines and increased IFN γ were detected in the blood of immunized cows, where a T lymphocyte activation profile was evidenced at 12-h post infection. Acute phase neutropenia was less severe in these cows, and pathways related to neutrophil diapedesis and monocyte activation were also present. Furthermore, three intramammary-immunized cows showing faster healing and shorter mastitis duration had gene profiles that differed from their counterparts, but without any clue for the mastitis susceptibility difference. Inasmuch, when gene expression of CD4 T cells was assessed in mammary tissue, enrichment of IL-17-associated pathways was identified in the quarters of intramammary-immunized cows not only after challenge but also in the control quarters that were not infected. These findings indicate that local immunization mobilizes protective mechanisms that rely on the settlement of type 3 immunity-related CD4 T cells prior to infection.

npj Vaccines (2020)5:108; https://doi.org/10.1038/s41541-020-00258-4

INTRODUCTION

Mastitis is a major disease of dairy cows¹, which develops upon infection of the mammary gland by various bacterial species. Clinical cases are mainly due to *Escherichia coli* and *Streptococcus uberis*, while *Staphylococcus aureus* is more frequently isolated from chronic subclinical mastitis. Infection can easily be detected due to leukocyte recruitment in mammary tissue, and ultimately in milk, leading to an increased milk somatic cell count (SCC), an indicator widely implemented in dairy control programmes².

Mastitis prevention and cure heavily rely on antibiotics, so that mastitis is the main reason for their common usage in dairy cattle³. Alternatives to antibiotic treatment are needed to reduce their use and the spread of antibiotic resistance⁴. Development of strategies to increase mammary defences is a major possibility, and among immune-based approaches, vaccination would be a good path to follow⁵. Unfortunately, marketed vaccines based on bacterins, such as the *E. coli* J5, suffer a certain number of limitations that lessen their generalized utilization in dairy herds. These vaccines have shown a low reduction of clinical mastitis incidence upon natural exposure, and they mainly reduce the severity of mastitis and milk losses^{6,7}. Since new developments are needed to overcome the low efficacy of available commercial vaccines, the slow progress is partly due to the lack of knowledge about the way vaccines operate in the mammary gland and about the exact protective mechanisms against infection.

In effect, the local immune response relies essentially on early and efficient recruitment of neutrophil granulocytes to control bacterial infections. Antibodies in the IgG1 and G2 isotypes, and to a lesser extent IgA since their concentration is lower in ruminants compared to other mammals, do not seem to play a major role although they are generally induced by current vaccines.

Vaccine delivery via mucosal routes is an interesting alternative to parenteral administration⁸. Local immunization through respiratory, digestive, and genital mucosal surfaces has been shown to induce local immunity associated with tissue-resident CD4 and CD8 memory T cells^{9–11}. The mammary gland does not have a mucosal surface strictly speaking, although it is sometimes affiliated to the mucosal common immune system. Surprisingly, the induction of a robust immune response against mammary pathogens through intramammary (IMM) immunization has rarely been reported in dairy ruminants. Pioneering work by Lascelles¹² used living staphylococci in dairy ewes, but the main limitation of these studies is that no information is available on the mammary status and ensuing inflammation before injection of the challenge inoculum. The presence of high SCC at the time of experimental infection may have prevented the proliferation of the bacterial inoculum and thus the development of mastitis. More recently, immunization locally with killed *Streptococcus uberis* had favorable effects on the reduction of mastitis frequency on the next lactation in dairy cows¹³. Local injection with a low number of dead *E. coli* bacteria also had some protective effects, although it

¹IHAP, Université de Toulouse, ENVT, INRAE, 31076 Toulouse, France. ²SIGENAE, GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, 31326 Castanet Tolosan, France. ³ISP, INRAE, Université de Tours, 37380 Tours, France. ✉email: gilles.foucras@envt.fr

is doubtful if immunity was really present at the time of challenge due to the low severity of the induced mastitis in the control group¹⁴.

In a vaccination trial against bovine *E. coli* mastitis previously reported by us¹⁵, we used a local administration of *E. coli* crude antigens after a primary intramuscular (IM) immunization, also known as 'prime and pull' strategy. Compared to IM route alone, we showed that the booster injection via the ductal route improved the immunization protocol efficacy. The antibody response did not seem to play a significant role in the immune defences in this case, and the development of a cell-associated immunity with interferon- γ (IFN γ) production in the milk of immunized cows was evident. Indeed, the cell-mediated response may be essential to improve the protective response against *E. coli* mastitis.

To improve our knowledge about vaccine-based cellular immunity against *E. coli* infection, we performed a whole-blood transcriptome analysis on the samples collected during the mastitis trial, and also the immunization phase of the study previously described¹⁵. By combining high-throughput assessment of whole-blood transcriptome and cytokine protein expression, we describe here the systemic response during *E. coli* mastitis. Furthermore, we provide evidence that the settlement of interleukin-17 (IL-17)-producing T helper cells in the mammary tissue improves immunity to *E. coli* mammary infections.

RESULTS

Clinical and biological differences among experimental groups upon *E. coli* mammary challenge

The mastitis trial contained two different immunization treatment groups (IM or IMM, $n = 6$ per group), and a control (CONT, $n = 6$) group that only received the same adjuvant intramuscularly at the indicated dates (Fig. 1a). The ensuing results have been previously described in more details over a course of 10 days¹⁵. Here, we summed up the main clinical and biological differences observed during the challenge. Indeed, in order to look for mechanisms and the specific effects of each immunization scheme, three time points (0, 12, and 40 hpi) were selected as the most discriminant among the groups, where significant differences were evidenced (Kruskal–Wallis, $p < 0.05$), and corresponding to the acute (12 hpi) and the late (40 hpi) phases of the *E. coli* mastitis. Several criteria are considered: the systemic clinical score (Fig. 1b), white blood cell (Fig. 1c), and neutrophil (Fig. 1d) counts; milk production (Fig. 1e); the mammary clinical score (Fig. 1f), and milk bacteriological load (Fig. 1g).

In addition, blood samples collected at day 0 (I0), and 7 days after the IM immunization (I7) and the booster injections (B7) were included in the analysis. In total, 107 samples from six dates and six animals in each of the three groups were prepared for transcriptome analysis; one sample (CONT, 12 hpi) was missing due to the low amount of RNA recovered after extraction. The whole-blood transcriptome was established using RNA sequencing with an average value of 10.3 million (R1 + R2) reads for each sample (Supplementary Table 1). The dataset obtained with FeaturesCount after STAR alignment comprised 24,617 bovine genes. At the acute phase of *E. coli* mastitis (12 hpi), the number of reads was significantly lower than at the other time points (Kruskal–Wallis, $p < 0.05$), except in the IMM group (Supplementary Table 1). One sample (IM, B7) was removed from the dataset due to an outlier profile in the principal component analysis (PCA). In total, 106 samples were available for the statistical analysis.

Whole-blood transcriptome through RNA-seq identifies the main biological functions altered by *E. coli* mastitis

As a blood transcriptome related to bovine *E. coli* mastitis has never been reported before, we first investigated changes in gene

expression in the CONT group. A PCA showed that the expression profile varied according to the time, with a clear separation between 0 and 12 hpi (PCA axis 1: 53.8%), and between 0 and 40 hpi (PCA axis 2: 24.3%) (Fig. 2a). Differentially expressed genes (DEGs) were identified with DESeq2 at these two time points in a comparison with the pre-inoculation time point. The numbers of DEGs were 5214 (with 2653 over- and 2561 under-expressed genes), and 3385 genes (2202 over and 1183 under), at 12 and 40 hpi, respectively. DEG lists were mostly different at the two times, as a minority of genes was common between the two sets as shown in a Venn diagram (with 832 over and 447 down-regulated DEGs) (Fig. 2b). Volcano plots showed determining DEGs for the 12 hpi (Fig. 2c) and 40 hpi (Fig. 2d) comparisons, and the lists of the top 10 over- or under-expressed DEGs are available in Supplementary Table 2 for 12 and 40 hpi time points. To further depict the response and identify blocks of commonly regulated genes and functions, a heatmap was prepared (Fig. 2e), with samples clustering according to the time. The top 500 most DEGs were divided into four modules that typified individual time points. The function modules were determined with ToppFun software and are presented with a dendrogram to evaluate correlations among functions. At 12 hpi, *Cellular response to lipopolysaccharide and Regulation of cytokine mediated signaling pathway* were the most activated functions, while *Humoral response* and *Macrophage activation* characterized the late phase at 40 hpi. *Neutrophil migration* and *Cellular response to reactive nitrogen species* were significantly detected at both times. Plasma concentrations of IL-6, tumor necrosis factor- α (TNF α), IL-10 cytokines, and CCL2, CCL4, CXCL10 chemokines were significantly increased at 12 hpi (Kruskal–Wallis, $p < 0.01$), and returned to initial values at 40 hpi; only IL-1RA plasma concentration was increased at both 12 and 40 hpi (Supplementary Fig. 1).

Comparison with the transcriptome profiles of IM- and IMM-immunized cows

In order to compare the blood transcriptomes among the groups, we first looked for DEGs in the two immunized groups, independently. After DESeq2 analysis, DEGs were identified at 12 and 40 hpi in both groups. The lists showing the top 10 genes with up and down expression in IM and IMM groups are provided in Supplementary Tables 3 and 4. The DEG numbers were not different between the groups at 12 hpi (Fig. 3a, c), but it was much lower in the IMM group at 40 hpi (Fig. 3b, d). Venn diagrams are presented in Supplementary Fig. 2 to determine the number of commonly deregulated genes. Interestingly, the number of DEGs was lower in the immunized cows compared to the controls at 12 hpi, and the difference was even larger for the IMM group at 40 hpi, where DEG number represented only 10% of the numbers obtained in the IM and CONT groups, supporting the idea that systemic healing was earlier in this group. Altogether, samples clustered essentially according to the Time in a global heatmap (Fig. 3e), except three samples from IMM cows at 40 hpi that grouped together with the samples collected before inoculation. Bacterial excretion in these cows was the lowest after this time point (<100 bacteria/mL) and their mammary infection was actually cleared after 48 h, and twice faster than other cows. Indeed, expression profiles of these three IMM cows cannot be distinguished between 0 and 40 hpi, indicating that gene expression almost returned to the pre-infection state in these cows. They are further defined as cured cows on the basis of both their lower clinical scores and their gene expression profiles. The top 500 DEGs divided into four modules determined with ToppFun software and were essentially related to immune functions (Fig. 3e). At 12 hpi, activated functions were *Defense response to bacterium*, and *Neutrophil chemotaxis*, as shown before for the CONT group. At 40 hpi, *Type I interferon signaling* and *Humoral immune response* were significantly activated in the IM group,

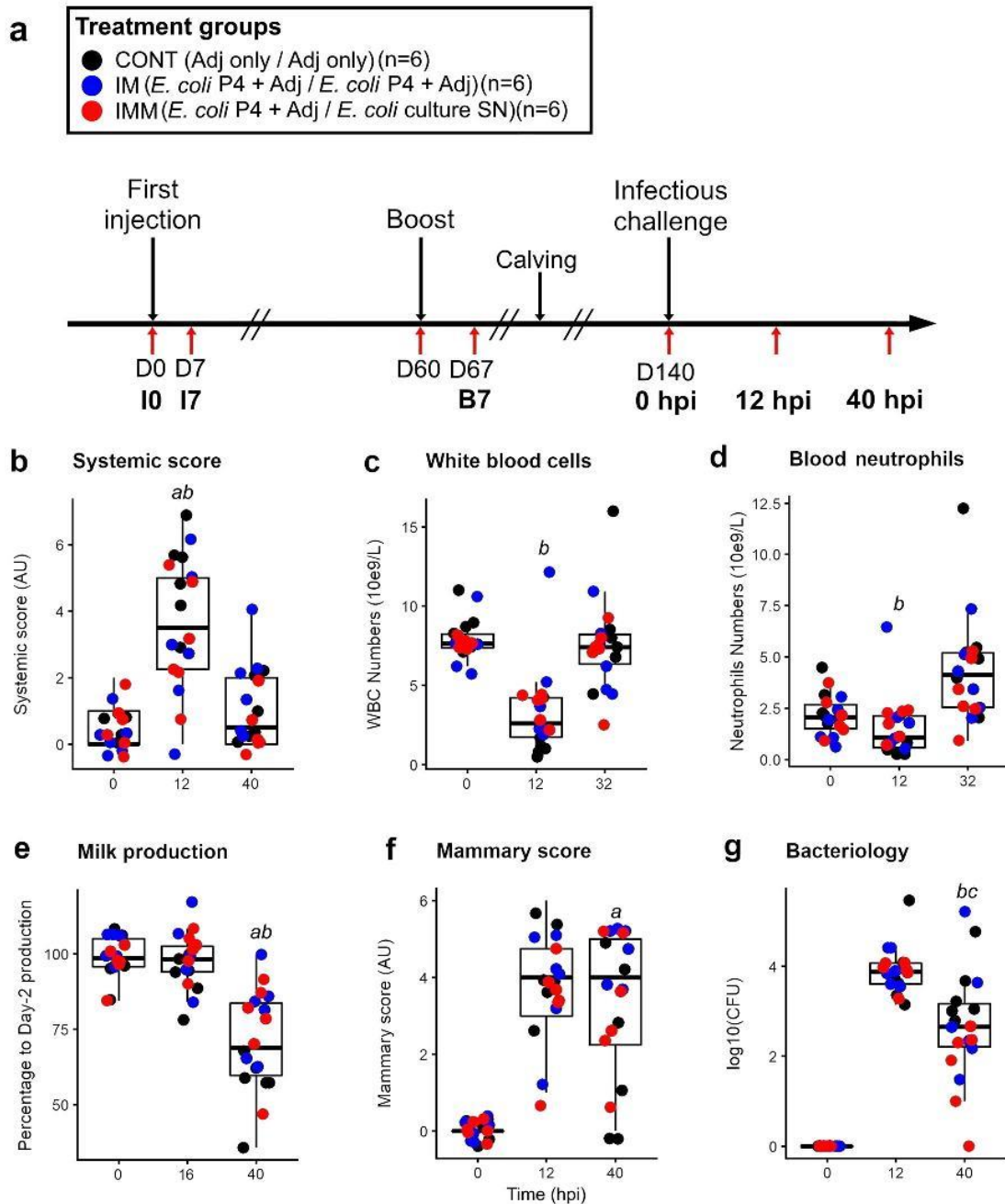
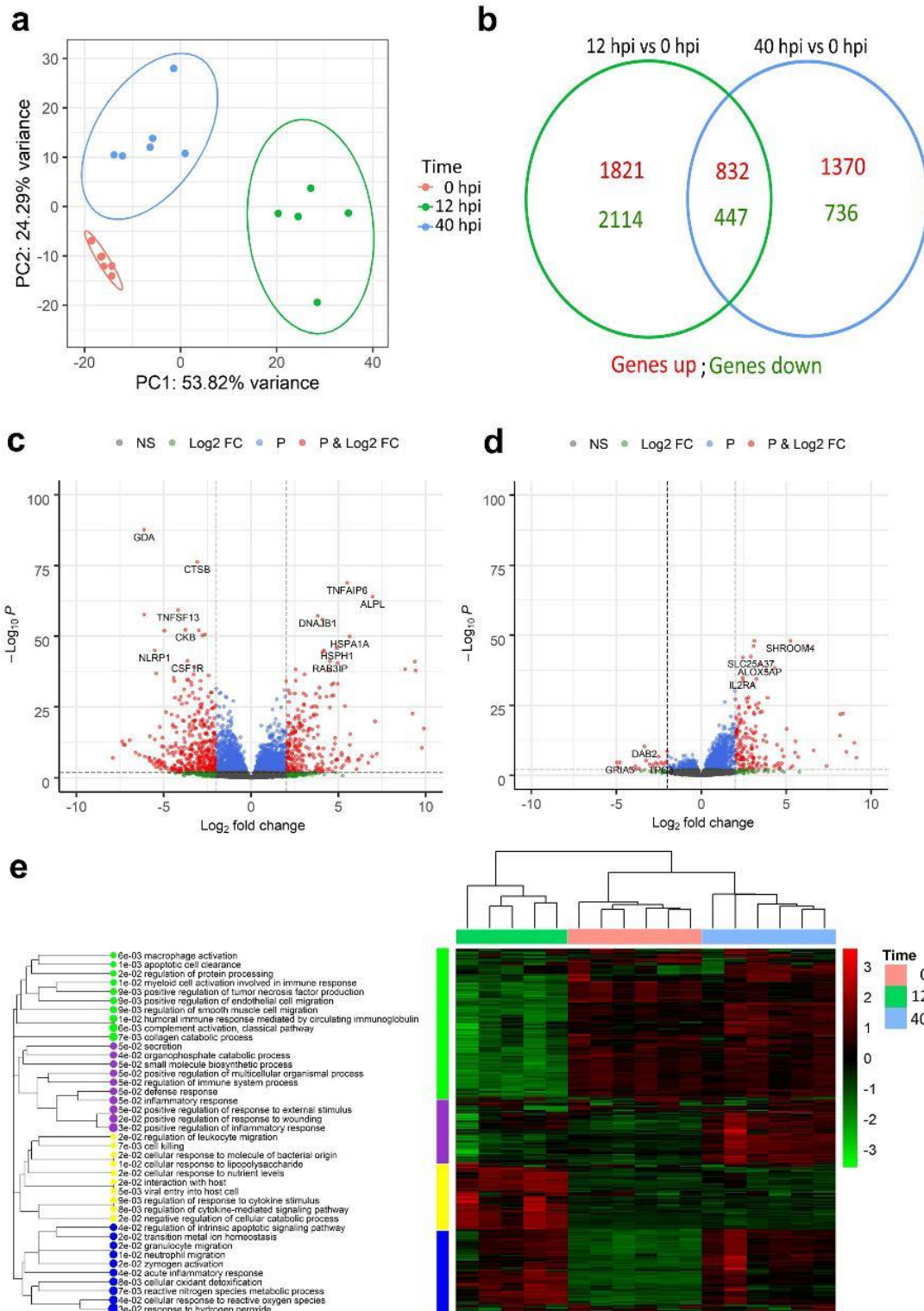


Fig. 1 Clinical, hematological, and mammary response of the cows according to their immunization group at three informative time points of the *E. coli* mastitis course. **a** Experimental scheme with immunization and experimental challenge phases. Systemic (**b**) and mammary (**f**) clinical scores were established as composite values of several signs recorded by observers blind to the treatment. Milk production expressed as the relative percentage of milk produced at each time compared to the average production before inoculation (**e**). Milk bacteriology (**g**) expressed as the log 10 value of colony-forming units per mL of milk, white blood cells (**c**) and neutrophils (**d**) counts are presented at the same times. Each dot corresponds to an individual cow with colors indicating the treatment group and the box plot showing the mean and standard deviation at each time point. Letter indicate a difference (Kruskal-wallis, $p < 0.05$) between groups (**a**: CONT vs IM; **b**: CONT vs IMM; **c**: IM vs IMM).

which showed the slowest recovery rate. The profiles were identical at time 0 hpi among the groups, indicating the absence of any ongoing immune activation or physiological difference among the groups. Plasma cytokine dosage showed that the production of pro-inflammatory cytokines IL-1 β , IL-6, and TNF α , but also CCL2 and CCL4 chemokines was significantly lower in the immunized compared to the CONT cows at the acute phase (12

hpi), indicating a lower inflammatory response upon infection in the IM and IMM groups (Supplementary Fig. 3). Furthermore, IFN γ and CXCL10 concentrations were higher in the IM group at 40 hpi, in agreement with the activation of interferon-related pathways in this group during the late phase of the response (Fig. 3e).

To identify the biological functions that correlated with the systemic and mammary scores as presented above in Fig. 1b, f, we



used weighted correlation network analysis (WGCNA) on this set of gene expression data. Twelve modules significantly correlated with the mammary and/or systemic scores. Notably, a T helper type 1 cell-related module was activated at 12 hpi in the immunized groups, indicating that immunization protocols

through either IM or IMM routes changed the course of the immune response during mammary infection (Supplementary Fig. 4). These results also suggest that immunization promoted a T cell response that was mobilized in blood early during the mastitis course.

Fig. 2 Whole-blood transcriptomic response of control cows during the course of an *E. coli* mastitis. Differentially expressed genes (DEGs) between times were assessed using Deseq2. **a** A principal component analysis on the top 500 most differentially expressed genes of control cows at 0, 12, and 40 h post infection (hpi) shows that the expression profile of blood cells changed according to the time during the *E. coli* mammary infection. A Venn diagram presents the number of DEGs between time points, and the number of genes in common between the two comparisons (between 12 and 0 hpi and between 40 and 0 hpi) (**b**). Volcano plots were used for determining DEGs in the CONT group between 12 and 0 hpi (**c**), and 40 and 0 hpi (**d**). Thresholds were arbitrarily established for a q value < 0.01 and a $\log_2\text{FoldChange} > 2$ and genes above these thresholds are depicted in red dots. **e** A Heatmap of the top 500 DEGs during infection in control group is presented using Euclidian distribution analysis. Results of the functional analysis with ToppFun is presented as a dendrogram showing the relationship among biological functions and blocks of commonly regulated genes.

Systemic and local routes of *E. coli* immunization elicit different gene expression profiles

Since immunization changed the susceptibility to *E. coli* mastitis, we investigated DEGs after the first intramuscular and booster injections to decipher the immune-related functions mobilized by each kind of immunization scheme. The comparison of transcriptome data did not reveal any difference among the groups before immunization (I0). Analysis of the primary response at 7 days post-IM injection (I7) indicated that 600 genes were DE in the *E. coli*/adjuvant ($n = 12$ cows) compared to the adjuvant alone condition ($n = 6$ cows) (Fig. 4a). DEGs were related to immune functions with *Secretion*, *Myeloid leukocyte activation*, and *Immune effector process* (Supplementary Table 5). After the booster injection (B7), through either the IM or the mammary ductal (IMM) routes, gene expression profiles were significantly different among the groups. Interestingly, the number of DEG was greater in the immunized groups compared to the CONT group, with 389 and 416 genes upregulated in the IM and IMM groups, respectively (Fig. 4b). These results suggest that the presence of *E. coli* antigens and microbe-associated molecular patterns like the lipopolysaccharide (LPS), in addition to the adjuvant which was the same between CONT and IM groups had a strong effect on the secondary response. DEGs in the IM group were related to *Interferon alpha/beta signaling* and *Cytokine signaling in immune system* functions, whereas DEGs in the IMM group were related to *Neutrophil degranulation* and *Chemokine signaling pathway* functions. Venn diagram indicated that 399 upregulated genes were common between IM and IMM groups. These genes were related to *Neutrophil degranulation* and *Innate immune system* functions (Supplementary Table 6). Furthermore, while no module was activated in the CONT group, four modules were commonly upregulated in the immunized groups with *Enrichment in monocytes* and *Immune activation—generic cluster* as the main modules. The module *TLR and inflammatory signaling* was activated only in the IMM group at B7 time point (Fig. 4c). So, the injection of antigens from *E. coli* culture supernate in the mammary gland lumen stimulated a systemic response which is essentially different from that induced by the IM immunization with killed bacteria in adjuvant in previously sensitized cows.

Blood transcriptome profile in IMM cured cows correlated with early cure

We then paid attention to the comparison between the two IMM subgroups, whose clinical signs and blood transcriptome differed significantly at the late mastitis phase (40 hpi). To define the peculiarities of the cured cows, their blood profile was further scrutinized at all the available dates. A PCA was performed on the top 500 most variable genes from the I0 and B7 data (Fig. 5a). At the B7 time point, PCA showed a significant difference between cured and non-cured cows indicating that the response to immunization of these particular cows was different and correlated with early cure. To further this hypothesis, a modular analysis was performed with the tmod R package. Two modules are common among all animals with *Enrichment in monocytes* and *Immune activation—generic cluster* as the main modules. Three

modules were further activated in the cured animals and functions of these modules were *Enriched in neutrophils*, *Cell cycle and transcription*, and *TLR and inflammatory signaling* (Fig. 5d). These functions upregulated in the cured subgroup support the idea that the differences of the systemic response are related to neutrophil mobilization, and are probably an indirect consequence and not the true cause of the difference of cure quickness due to immunization.

To look for the genes and immune functions differentially mobilized between cured/not cured IMM cows during the mastitis, a PCA at 12 and 40 hpi was performed with the top 500 most DEGs, showing similar profiles at 12 hpi, but again a marked difference at 40 hpi (Fig. 5b, c). Analysis with DESeq2 identified only 26 genes (21 up and 5 down in the cured group), whereas at 40 hpi, the number of DEGs was 307 genes with 22 genes upregulated in the cured and 285 in the non-cured subgroup, respectively. Biological functions are presented in Supplementary Table 7. Again, the functional differences do not point to a probable cause of early healing in the three cured cows compared to the rest of the group.

In the comparison between IM and IMM groups, the fact that no DEG, nor function, except neutrophil mobilization, lead to a mechanism explaining the better outcome in locally immunized cows, leaves at least two possibilities: (i) these were very transient during the mastitis course and were not identifiable at the chosen time points for analysis or (ii) the functional differences are not present in the blood, because they reside in the mammary tissue in link with the local immunization.

Th17-polarized CD4 T cells predominate in the mammary tissue of IMM quarters and are activated after the challenge with *E. coli*

Local immunization has been previously associated with recruitment and establishment of CD4 and CD8 T cells in the targeted tissues. Furthermore, since no difference could be identified in the blood, we made the hypothesis that the mechanisms responsible for the difference of mastitis outcome may develop in the mammary tissue. To examine the phenotype of the T helper cell compartment in link with the mammary immunization, six cows (3 CONT, 3 IMM) were challenged once again using the same conditions (Fig. 6a). CD4 T cells (CD45+ CD4+ cells) were isolated by fluorescence-acquired cell sorting at 16 hpi from mammary tissue upon collagenase treatment. Four different conditions were examined: one challenged and one healthy control quarters from three CONT and three IMM cows. RNA sequencing of pools of purified (purity $> 95\%$) CD4 T cells showed a transcriptional profile where IL-17-related pathways were significantly overrepresented in cells from the IMM cows. Profiles of DEGs were more related to the immunization than to the challenge status of the mammary quarters, suggesting that resident cells of the Th17 phenotype pre-existed the *E. coli* mammary challenge. Analysis of biological functions with ToppFun software is shown on a dendrogram presented in Fig. 6b. The top 500 genes were divided into four modules. Module 1 was related to the IMM group and correlated with *Neutrophil degranulation* and *Regulation of BAD phosphorylation*. Module 3 correlated with *Th17 cell differentiation* and *CD40L signaling pathway*. These modules were expressed in IMM infected

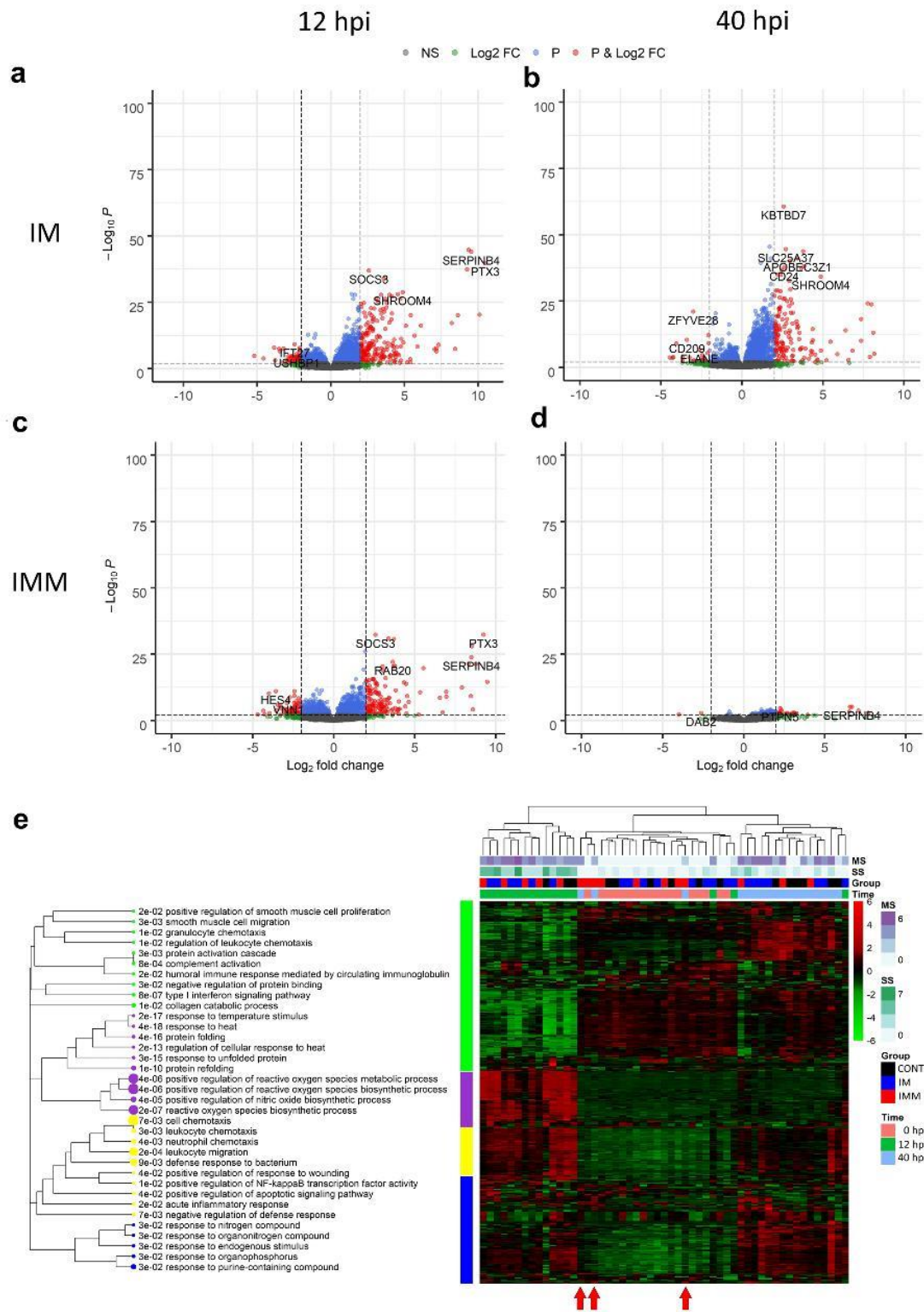


Fig. 3 Whole-blood transcriptomic analysis of the response of immunized cows during the *E. coli* mastitis course. DEGs at two time points (12 and 40 hpi) in comparison with 0 hpi were identified with Deseq2. Volcano plots comparing gene expression differences between 0 and 12 hpi (**a, c**), and at 0 and 40 hpi (**b, d**) are presented for the intramuscular (IM) and the intramammary (IMM) group, respectively. Red dots correspond to genes with a q value < 0.01 and a $\text{Abslog}_2\text{FoldChange} > 2$. **e** Heatmap of the top 500 DEG in control and immunized groups during the course of infection is presented. Functional analysis was performed with ToppFun and a dendrogram shows the relationship among biological functions of commonly deregulated genes according to the mammary and systemic scores, the time, and the treatment group. MS mammary score, SS systemic clinical score. Red arrows indicate IMM cows with 0 and 40 hpi profiles that were undistinguishable (further designed as cured cows).

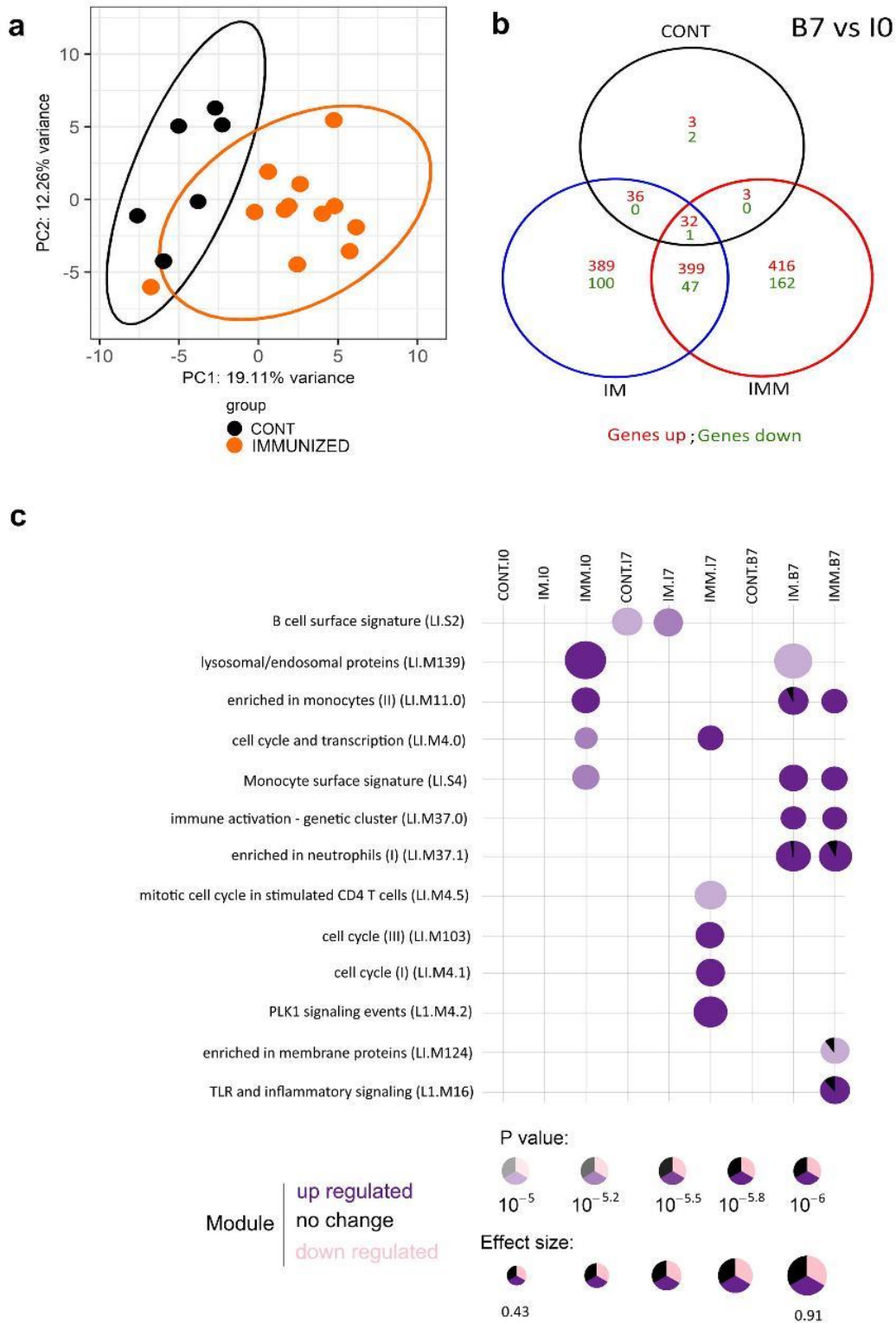


Fig. 4 Whole-blood transcriptomic response to immunization in control and immunized groups. **a** A principal component analysis (PCA) on the top 500 most DEG of CONT (black) and immunized IM + IMM (orange) cows after the first injection (I7) shows that the expression profile of blood cells changes upon systemic immunization. **b** Venn diagram of identified genes in control, intramuscular and intramammary groups at 7 days after booster injection (B7) compared to I0 (before immunization). **c** A modular analysis on B7 gene expression shows the significant difference (p value $< 10^{-5}$) of upregulated biological functions between CONT, IM, and IMM groups.

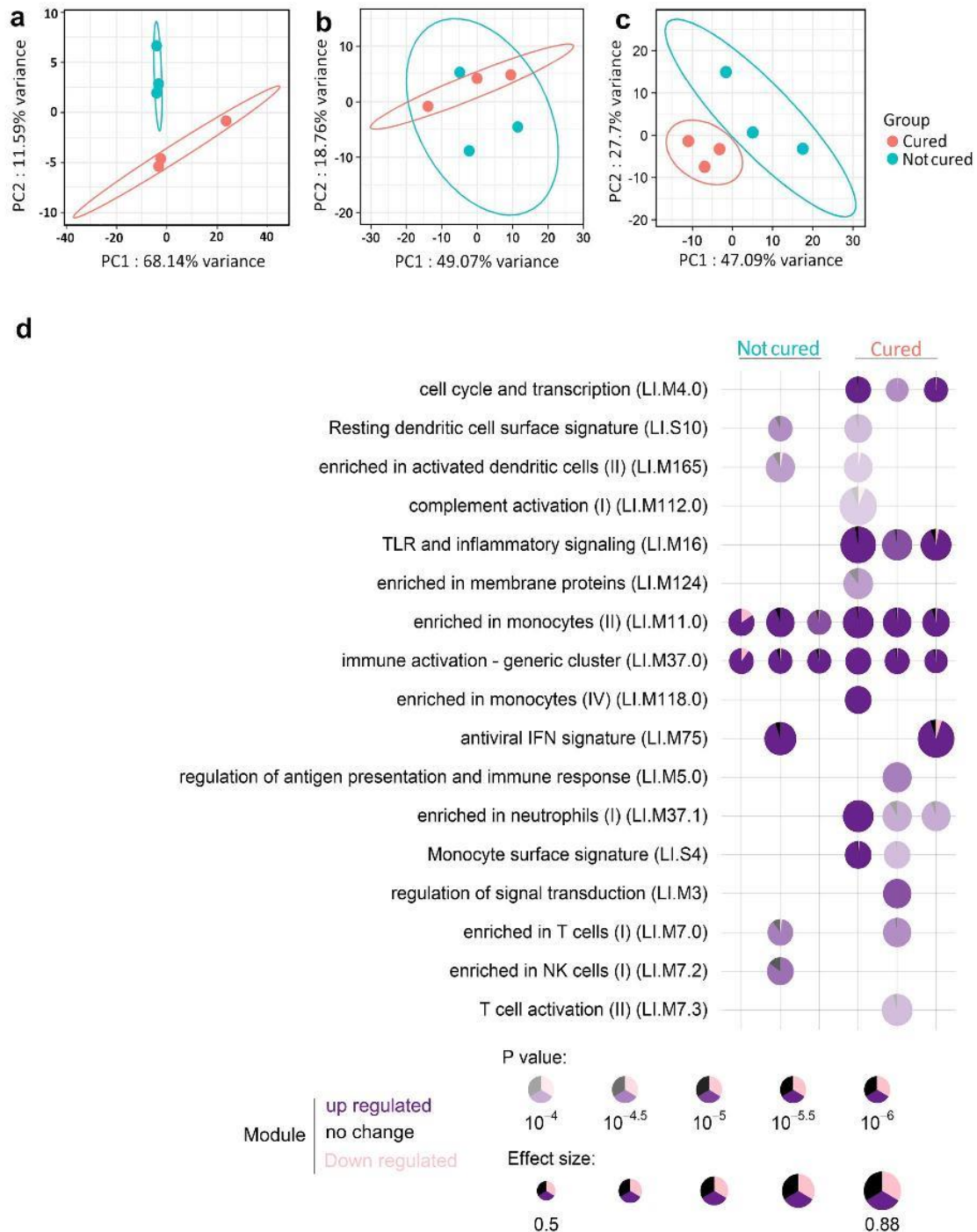


Fig. 5 Response upon the mammary booster injection and correlation with the outcome of the *E. coli* mammary challenge. **a** A principal component analysis of gene expression profiles (top 500 most variable genes) for the intramammary-immunized IMM cows at B7. **b** At 12 h post infection and **c** and at 40 h post infection; **d** a modular analysis at B7 shows the difference of upregulated biological functions between cured and not cured cows (p value above 10^{-4}).

and non-infected glands. Module 2 was related to *IL-17 signaling pathway* and *Cytokine-cytokine receptor interaction* and expressed in IMM infected quarters only. Modules related to the infection were mainly correlated with *IL-17 signaling* and *Neutrophil migration*, suggesting that the local Th17 response promoted mobilization of neutrophils, which was protective against *E. coli* infection.

The list of DEG identified with DESeq2 analysis was used to evaluate a difference of pathway activation between the CONT and IMM groups upon infection with Ingenuity Pathway Analysis software. The gene network showed that an IL-17-related network was strongly activated in the IMM group during infection (Fig. 6c).

To confirm the overproduction of IL-17A in immunized quarters, cytokine secretion was measured in milk before *E. coli* inoculation

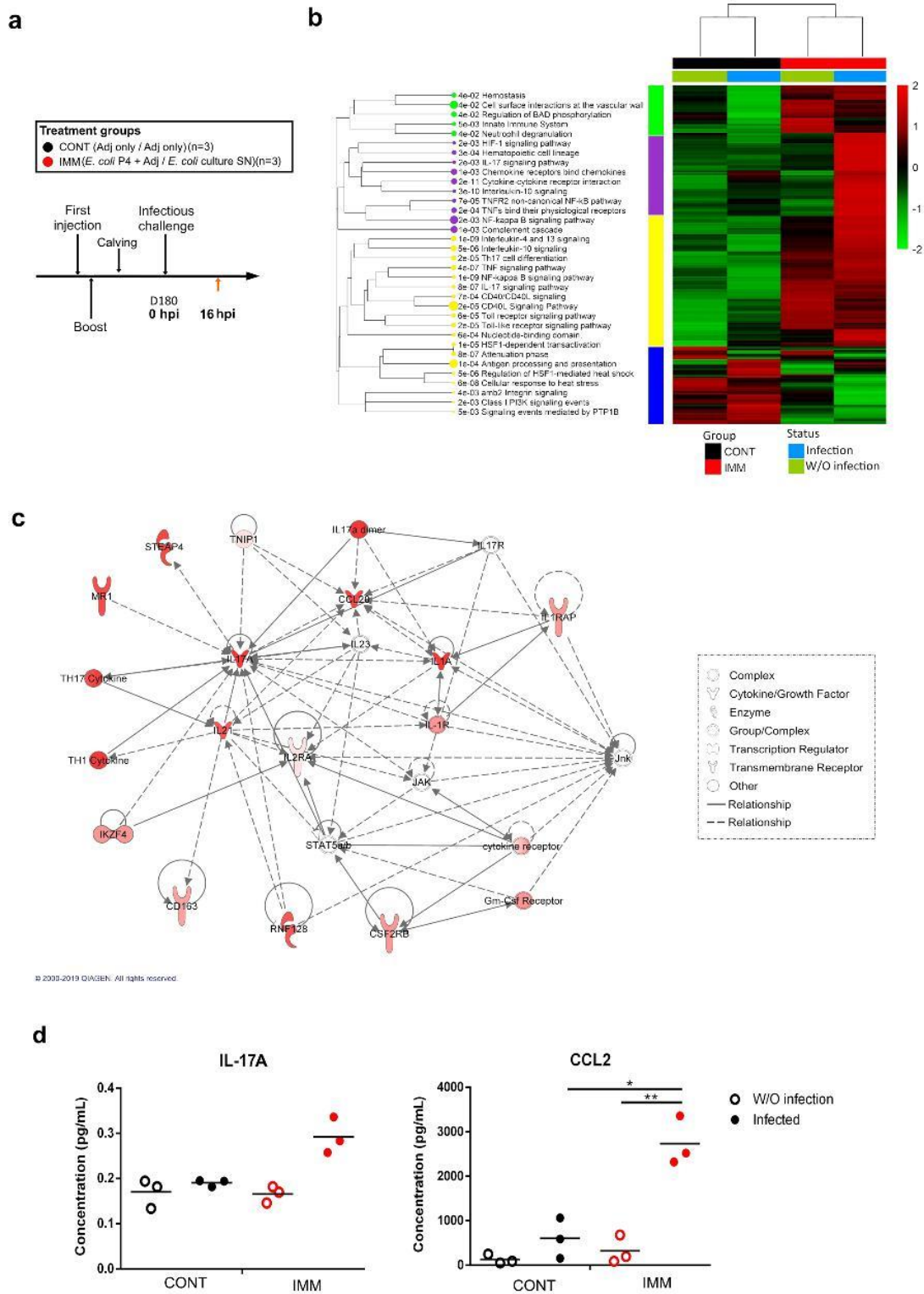


Fig. 6 Gene expression profile of CD4 T cells isolated from the mammary tissue of control and challenged quarters from locally immunized and control cows. a Experimental scheme with immunization and challenge phases. DEGs were assessed between four conditions corresponding to challenged and control quarters in CONT and IMM cows. **b** A heatmap showing the top 500 DEGs according to the conditions is presented. Functional analysis performed with ToppGene is presented as a dendrogram of functions in each block of correlated genes (p value < 0.05). **c** Gene network of the DEGs in CD4 T cells between the infected IMM and infected CONT conditions is presented and shows the relationship with IL-17 cytokine. **d** Tissue IL-17A and CCL2 production in non-infected and challenged quarters of CONT and IMM cows was determined on a pre-defined amount of tissue homogenate with a Milliplex assay at 16 hpi. Individual values are shown for $n = 3$ cows per group.

and just before tissue collection at 16 hpi. Despite the low number of animals, IFN γ and IL-17A secretion tend to be higher in the milk of locally immunized cows, while IL-10 secretion was lower (Supplementary Fig. 5), in accordance with the previous report¹⁵. When cytokines were measured in the tissues of infected quarters, IL-17A, and CCL2, a IL-17-induced cytokine were more abundant in IMM cows (Fig. 6d); a difference of IL-1 α was also detected (Supplementary Fig. 6). It is worth noting that milk and tissue concentrations were poorly correlated, except for IFN γ cytokine.

Altogether, these data support the idea that the local booster injection promoted the settlement of basal Th17 immunity in mammary tissue, further amplified by the *E. coli* infection.

DISCUSSION

Our aim was to decipher the immune mechanisms associated with the local route of immunization which revealed more effective at protecting the mammary gland against infection by *E. coli*. We analyzed blood and milk samples collected on the occasion of our previous challenge study¹⁵. We provide evidence that the two modes of immunization that were evaluated in parallel mobilized a different immune response. The data strongly suggest that the injection of antigens from *E. coli* culture supernate to the local route enabled the establishment in the mammary tissue of Th17-polarized CD4 T cells that were undetectable in non-immunized cows.

We reported previously that the clinical signs of mastitis were less severe in immunized compared to control cows, with faster bacterial clearance and a moderate leukopenia. Nevertheless, there was also a notable difference in response between systemic (IM) and local (IMM) immunization routes, as the inflammatory response persisted longer in the IM group. We suggest now that this was associated with a heightened signaling related to Type I interferon pathway that may be the result of TLR4-Myd88 signaling by LPS released from *E. coli* bacteria in the infected quarters. It could also be the result of a late influx of T cells that was detected in the milk of the IM group after 40 hpi. These cells were circulating in blood at the time of the mammary challenge and were recruited into the inflamed mammary tissue due to the expression of chemokines and endothelial adhesion molecules. Some may have proliferated locally due to antigen recognition or became activated non-specifically by the local production of cytokines. They may thus have contributed to the increase of the inflammatory response during the late phase of the mastitis, while healing was already ongoing in the CONT and IMM groups.

We chose to evaluate blood cell transcriptome by RNA-seq rather than mammary tissue or milk cell gene expression as previously described by us^{16,17} or other groups¹⁸. Indeed, this approach allowed us to do a kinetics study and to look for mechanisms over time on a tissue whose cell composition is more stable than that of milk during the mastitis course. Indeed, upon infection, recruited cells in milk are essentially neutrophils (>90–95%) as previously shown¹⁵, whereas these represent no more than a third of the milk cells in healthy glands. Furthermore, RNA quality as defined by RNA integrity number (RIN) assessment is better on blood cells than on milk cells, where many leukocytes are dead or apoptotic/dying cells. If neutrophils represent essential terminal effectors against mammary infection, they do not represent the central type of cells induced by immunization. Conversely, CD4 and CD8 T cells are only a minority in the milk of infected glands, as their number in milk at the time of inoculation is very low (often lower than 10⁴ cells/mL). Later, they are really difficult to isolate from abundant and highly activated neutrophils upon infection with strongly inflammatory Gram-negative bacteria.

We chose to retain six crucial time points, three during the infection course to decipher the protective mechanisms, with 0 hpi corresponding to our time point of reference, and three during

immunization with time I0 corresponding to the time point before the first injection of immunization. In the CONT group, the blood transcriptome during the infection phase was informative about the response to an *E. coli* infection, with good clustering of the samples according to the Time. Blood leukocyte gene expression changed quickly and massively upon mammary infection with a large number of deregulated genes. This is partly due to changes in the blood leukocyte composition¹⁵ but also to changes in cell activation and functions. Indeed, functions related to neutrophils and bacterial recognition are highly activated at the peak of infection, whereas at 40 hpi, these functions are replaced by monocyte/macrophage activation and B cell response. The blood transcriptome actually revealed essential functions that are known to be mobilized or activated during the mastitis course. Blood represents a good compartment to sample to study the immune response to a mammary infection as also shown elsewhere¹⁹. This may be due to the high blood flow through the lactating mammary gland which is a well perfused, metabolically active organ.

Immunized cows had less significant changes in their blood transcriptome than the controls, in accordance with the systemic changes that were smaller in these cows, as shown in previous reports indicating that immunization with the *E. coli* J5 bacterin diminished clinical signs and associated symptoms like milk production drop, a parameter that is well correlated with the severity of *E. coli* mastitis²⁰. Indeed, the number of DEGs at 12 hpi compared to the 0 hpi time point is smaller in the immunized groups compared to the control group with 44% and 57% fewer DEGs for the IM and IMM groups, respectively. On the other hand, at 40 hpi, this decrease is only seen in the IMM group with a 90% decrease of DEGs, confirming the shorter duration of the mastitis in the IMM group in accordance with recorded systemic and mammary clinical signs. This is confirmed by the lower production of inflammatory cytokines (IL-1 β , IL-6, and TNF α) and chemokines (CCL2 and CCL4). In addition, high production of IFN γ was detected at 40 hpi in the IM group that corresponded to the probable activation of Th1-related CD4 or CD8 T cells that were primed by systemic immunization. This production was lower in IMM cows, maybe because they received only one injection of vaccine in adjuvant or because antigens from *E. coli* culture supernate in the mammary lumen were not able to expand or differentiate the cells to the same magnitude and/or phenotype as did the systemic immunization. As to the IL-17 cytokine, no difference of secretion could be detected in the plasma at any time point¹⁵.

To look for the pathways correlating with the systemic and/or mammary scores, WGCNA was implemented. It enabled the identification of two different modules that were associated with lymphocytic functions in the IM group and granulocyte degranulation in the IMM group. The indication that T cell activation was weaker in the IMM group compared to the IM group was interesting and led us to investigate CD4 T cells in the mammary tissue, and not only the blood cell compartment.

We provide evidence that IL-17-producing CD4 T cells are enriched in the tissue of bacterial antigen-exposed mammary glands. The antigen repertoire of these cells is not known and whether they are specific to *E. coli* antigens remains to be determined as previously shown for CD8 T cells in another tissue²¹. The role of IL-17 cytokine in immune responses against extracellular bacteria is now well established²². This cytokine is known to participate in the local defence by inducing the production of CXCL8 by epithelial and/or stromal cells during infection by fungi and extracellular bacteria, which in turn induces neutrophil influx and antibacterial peptide secretion at the infection site²³. Its role in bovine mastitis has been highlighted in several reports. In a mastitis model with *E. coli* infection in mice, the IL-17 response in the mammary gland is instrumental in *E. coli* clearance and reduction of mammary inflammation²⁴. In dairy

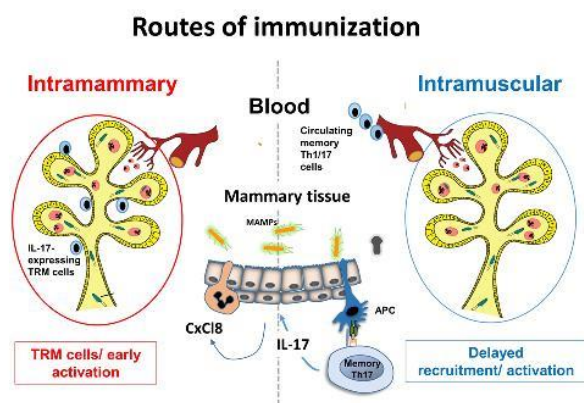


Fig. 7 Local mammary immunization establishes resident memory CD4 T cells in the mammary tissue on the contrary to the intramuscular route of immunization. The schematic describes our current view of the beneficial effect of intramammary immunization and compares the two routes of immunization. Systemic injection of killed bacteria in adjuvant produces a pool of Th1/Th17 circulating CD4 T cells. Injection of *E. coli* culture supernate antigens via the teat ductal route enables settlement of CD4 TRM cells in the mammary tissue that are poised to express IL-17 and mobilize a protective response through neutrophil recruitment early during the course of an *E. coli* mastitis.

cows, despite reports of IL-17 detection in mastitis cases^{25,26}, clear evidence was still lacking about its role in reducing mastitis severity.

IL-17-expressing CD4 T cells may be activated by antigen-presenting cells locally or IL-17 expression follows the secretion of IL-23 and activation of T mammary-resident memory (TRM) cells. IL-17 cytokine by acting on epithelial and stromal cells induces CXCL8 secretion and recruitment of neutrophils as shown by the upregulation of this pathway in the modular analysis. Unfortunately, by measuring cell recruitment in milk every 4 h, we were unable to show an influx difference among the groups. The number of cells in milk may not be completely correlated with precocity or intensity of the recruitment in the tissue since neutrophils need to cross the epithelial barrier to reach the mammary lumen and be identifiable in milk. Furthermore, neutrophil bactericidal activity may differ among conditions depending on the cytokine/chemokine pattern related to the type 3 activity they have encountered. Again we were not able to show any difference of opsonophagocytosis and killing activity among the groups using an *in vitro* assay¹⁵, suggesting that the antibody response is not central to the mechanisms of bacterial clearance and mastitis healing. Unfortunately, the production of antibacterial peptides was not measured in the secretion of infected quarters to look for other mechanisms related to type 3 immunity.

What determines the phenotype of TRM cells in the mammary gland is another important question to resolve. For priming immunization, we used an adjuvant that is known to promote the development of cellular response in cattle, with a Th1/Th17 profile as previously shown using a model antigen like ovalbumin²⁷. When we evaluated IFN γ and IL-17 production using a whole-blood assay with heat-killed bacteria in the same cows, cytokine production in IM and IMM cows was well above that of CONT cows. Although a high degree of variability was seen from cow to cow, IFN γ and IL-17 production increased after the booster injection in the IM group, suggesting that immunization induced circulating CD4 T lymphocytes producing IL-17A and/or IFN γ , contrary to what was seen in the IMM group. Intra-ductal injection of concentrated *E. coli* culture supernate at this step helped to recruit T cells within the tissue. Pro-inflammatory cytokine

response induced locally by *E. coli*-derived LPS may have increased homing and/or local proliferation of T cells, notably of IL-17-expressing CD4 T cells, as previously described in the lungs²⁸ or the intestine²⁹ for CD4 and CD8 TRM cells, respectively. Mechanisms of T cell homing within the mammary tissue are probably determinant and merits to be further examined in order to define the best modalities to settle locally more efficient T cell immunity and to improve the immunization protocol efficacy above the results of our proof of concept study.

When studying the response to immunization, blood transcriptome analysis allowed us to identify the presence of differential genes in IM and IMM, while very few DEGs were identified in the CONT group (Fig. 4b). At B7, activation of functions related to monocytes, neutrophils, and T cells were present in the immunized groups. Also, the modules related to inflammatory response and interferon pathway were activated in the IM group, indicating a stronger systemic inflammatory response in the group that received two vaccine injections in adjuvant. PCA showed that the difference of response was already present at B7 time point with upregulated functions remembering what was seen at time 40 hpi, like activation of neutrophil-related functions and TLR signaling. In contrast, profiles were indistinguishable at the peak of infection (12 hpi). This suggests that the cured group of IMM cows had a response at the time of the booster injection that is predictive of more rapid healing upon infection. A genetic control of the response to immunization cannot be excluded at this point to explain this difference.

To summarize we provided some evidence that the two routes of immunization and two different forms of *E. coli* antigens gave different results of immune activation and protective immunity against mammary *E. coli* infection (Fig. 7). The local route of immunization provided some mechanisms of *E. coli* mastitis control above what was observed in adjuvant-only receiving cows. Systemic immunization was helpful to generate a circulating pool of antigen-specific activated/memory T cells that may be mobilized after mammary infection. However, this recruitment occurred later in the course of mastitis and does not seem to be as protective as was the local immunization. Conversely, the antigenic booster enabled the local establishment of TRM cells that are poised to respond from the beginning of the infection. We provide a proof of principle that mammary type 3 immunity may be instrumental in protection against mastitis. Eliciting type 3 immunity in the mammary gland is likely to help the development of new commercial vaccines.

METHODS

Animals and ethical statement

All experimental procedures and clinical and laboratory records were previously described in Herry et al.¹⁵. All procedures involving animals were approved by the Ethics Committee of Val de Loire (France), DGRI's agreement APAFIS#813-2015061109103810v2. Animal studies were compliant with all applicable provisions established by the European directive 2010/63/UE. All methods were applied by approved staff members in accordance with the relevant standard operating procedures approved by the above-mentioned Ethics Committee. All animals used in this study were handled in strict accordance with good clinical practice and all efforts were made to minimize suffering. In brief, 18 Holstein-Frisian heifers in total were synchronized and inseminated at 15 month-old on average. During the last trimester of the gestation, 12 heifers were immunized using IM injections with heat-killed *E. coli* P4 emulsified in Montanide ISA 61VG[®] (Seppic, France) oil adjuvant, as recommended by the manufacturer. Two months later, half of the heifers (IM group, $n = 6$) got a second injection as a booster. The others received 50 μ g of protein concentrate prepared from a *E. coli* P4 supernate in two quarters (IMM group, $n = 6$). Cows of the control group (CONT, $n = 6$) were injected intramuscularly with adjuvant only twice at the same dates. Cows were then challenged on average at 49 days in milk in one healthy quarter by infusion of a *E. coli* P4 bacterial suspension (10^3 bacteria).

Forty days after the first challenge, a new mammary infusion with *E. coli* P4 was performed using the same conditions as before, in order to collect tissue samples and to assess the local response. Six previously challenged cows (3 CONT, 3 IMM) were inoculated in one healthy (SCC < 50,000 cells/mL), and not previously challenged quarter. Clinical data were not collected. SCCs were assessed and *E. coli* bacteria were re-isolated in order to check that inoculation was effective at inducing a mastitis. Cattle were euthanized 16 h after challenge and just before tissue harvest.

CD4 T cell isolation

Mammary tissue specimens from challenged and control quarters were harvested in the parenchyma region. Mononuclear cells were isolated from 2 g of mammary tissue. Briefly, mammary gland homogenates were incubated with collagenase type IV and DNase in RPMI for 30 min at 37 °C under agitation. Cells were recovered through Percoll gradient centrifugation and stained with PE-labeled CD45 (CC1, Bio-Rad) and FITC-labeled CD4 (CC8, Bio-Rad) mAbs, before cell sorting using a MoFlo Astrios EQ (Beckman Coulter, Brea, CA, USA).

RNA extraction

Blood samples were collected in EDTA tubes from the jugular vein at several time points, put on ice, and immediately treated with lysis buffer before storage at -80 °C. Total RNA was extracted from these samples with Nucleospin® RNA Blood kit (Macherey-Nagel). The amount of RNA was determined by measuring 260 nm absorbance using a NanoDrop ND-1000 (NanoDrop Technologies, USA). RNA quality was checked by measuring RIN with a Bioanalyzer (Agilent Technologies, USA). Measured values ranged between 7.8 and 9.3. RNA from isolated CD4 T cells was extracted using Nucleospin® XS kit (Macherey-Nagel).

Library preparation

RNA-seq was performed at the GeT-PlaGe core facility, INRA Toulouse. RNA-seq libraries have been prepared according to Illumina's protocols using the Illumina TruSeq Stranded mRNA sample prep kit to analyze mRNA. Briefly, mRNA was selected using poly-T beads. Then, RNA was fragmented, and double stranded cDNA were generated before adaptors were ligated. Eleven cycles of PCR were applied to amplify libraries. Library quality was assessed using a Fragment Analyzer and libraries were quantified by QPCR using the Kapa Library Quantification Kit. RNA-seq experiments were performed on an Illumina HiSeq3000 using a paired-end read length of 2 × 150 pb with the Illumina HiSeq3000 sequencing kits.

RNA-seq data processing and analysis

Read quality was assessed with FastQC software. Reads were aligned in comparison to the bovine reference genome (UMD3.1.87) with STAR software³⁰. Count table was generated with Cufflinks/RSEM^{31,32}. Identification of differentially expressed genes (DEGs) was performed with DESeq2 package within the R environment (R Core Team (2013)) following standard normalization procedures. False discovery rate was estimated with Benjamini and Hochberg correction. Expression changes were considered significant when both q value < 0.01 and a $\text{AbsLog}_2\text{Fold-Change} > 1$, except when otherwise indicated. DEGs were determined with group comparison at each time point. For control group analysis, the design matrix was defined by $\text{design} = \sim\text{time}$. For group comparison analysis, the design matrix was calculated by $\text{design} = \sim\text{time} + \text{group} + \text{time}/\text{group}$.

Determination of cytokine production

Mammary tissue was weighed to normalization before grounding using ceramic beads (Bio-Rad) in a mixture containing Triton X-100 and a anti-proteases cocktail (Sigma). Plasma, milk, and mammary tissue concentrations for 15 cytokines (IL-1 α , IL-1 β , IL-1RA, IL-2, IL-4, IL-6, IL-17a, IFN γ , CCL-2, CCL-3, CCL-4, CXCL8, CXCL10, and TNF α) were determined using a Custom bovine cytokines Milliplex xMAP assay (Merck Millipore). Data were recorded on a MAGPIX flow cytometer using Xponent software (Luminex).

Modular analysis

To identify immune functions activated at each time point of the transcriptome analysis, a modular analysis was performed with tmod

package using normalized count table and modules created by Chaussabel et al.³³ and Li et al.³⁴ as previously described.

Weighted correlation network analysis (WGCNA)

WGCNA was used to identify co-expressed genes associated with mammary and systemic clinical scores. The WGCNA R package was used for analysis³⁵. The minimum module size was set to 30, with a merge cut height of 0.25. The soft-thresholding power as a function of the scale-free topology index was defined at six. In this construction, all pairwise comparisons were conducted between control, IM, and IMM groups. Modules with correlation ($r \geq 0.5$) and p value (≤ 0.05) were extracted for further inquisition. WGCNA module annotation was carried out with ToppFun³⁶. A hypergeometric test with Bonferroni correction was applied to select statistically significant terms.

Data visualization

PCA and volcano plot were prepared with DESeq2 rlog-normalized RNA-seq data only. For PCA, the top 500 most variable genes were used. The EnhancedVolcano R package was used for volcano plot³⁷ and the pcaExplorer³⁸ was used for PCA. The heatmap package was used for heatmap visualization³⁹. The clustering was based on Euclidean distance. Boxplot graphs were drawn using tmod R package^{33,34}. To visualize gene expression and networks, Ingenuity Pathways Analysis (IPA, Qiagen) software was used to create network and to examine genes functions. Gene set enrichment analysis was represented in a dendrogram created with the IDEP R script⁴⁰. The network analysis was carried out with IPA Software (Qiagen). Barplot and dot plot representation were performed with GraphPad Prism version 6.01 for Windows (GraphPad Software, La Jolla California USA).

Statistical methods

All data are presented as the mean value \pm SD for $n = 6$ cows per experimental group, with individual values showed as a dot. Statistical analyses were performed using R software by applying Kruskal-Wallis, Mann-Whitney, or repeated measures Wilcoxon tests. The statistical significance was considered with a threshold p value below 0.05, except otherwise indicated.

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY

Sequence data that support the findings of this study have been deposited in Gene Expression Omnibus (GEO) with the accession number GSE159283 and GSE159286.

Received: 18 May 2020; Accepted: 23 October 2020;

Published online: 24 November 2020

REFERENCES

- Ruegg, P. L. A 100-year review: mastitis detection, management, and prevention. *J. Dairy Sci.* **100**, 10381–10397 (2017).
- Martin, P., Barkema, H. W., Brito, L. F., Narayana, S. G. & Miglior, F. Symposium review: novel strategies to genetically improve mastitis resistance in dairy cattle. *J. Dairy Sci.* **101**, 2724–2736 (2018).
- Suojala, L., Kaartinen, L. & Pyörälä, S. Treatment for bovine *Escherichia coli* mastitis—an evidence-based approach. *J. Vet. Pharmacol. Ther.* **36**, 521–531 (2013).
- Krömker, V. & Leimbach, S. Mastitis treatment—reduction in antibiotic usage in dairy cows. *Reprod. Domest. Anim.* **52**(Suppl. 3), 21–29 (2017).
- Sordillo, L. M. Mammary gland immunobiology and resistance to mastitis. *Vet. Clin. North Am. Food Anim. Pract.* **34**, 507–523 (2018).
- Wilson, D. J. et al. Comparison of J5 vaccinates and controls for incidence, etiologic agent, clinical severity, and survival in the herd following naturally occurring cases of clinical mastitis. *J. Dairy Sci.* **90**, 4282–4288 (2007).
- Steele, N. M. et al. The effect of J5 bacterins on clinical, behavioral, and antibody response following an *Escherichia coli* intramammary challenge in dairy cows at peak lactation. *J. Dairy Sci.* **102**, 11233–11249 (2019).

8. Lycke, N. Recent progress in mucosal vaccine development: potential and limitations. *Nat. Rev. Immunol.* **12**, 592–605 (2012).
9. Schenkel, J. M. et al. T cell memory. Resident memory CD8 T cells trigger protective innate and adaptive immune responses. *Science* **346**, 98–101 (2014).
10. Stary, G. et al. VACCINES. A mucosal vaccine against Chlamydia trachomatis generates two waves of protective memory T cells. *Science* **348**, aaa8205 (2015).
11. Shin, H. & Iwasaki, A. A vaccine strategy that protects against genital herpes by establishing local memory T cells. *Nature* **491**, 463–467 (2012).
12. Lascelles, A. K. The immune system on the ruminant mammary gland and its role in the control of mastitis. *J. Dairy Sci.* **62**, 154–167 (1979).
13. Finch, J. M., Hill, A. W., Field, T. R. & Leigh, J. A. Local vaccination with killed *Streptococcus uberis* protects the bovine mammary gland against experimental intramammary challenge with the homologous strain. *Infect. Immun.* **62**, 3599–3603 (1994).
14. Pomeroy, B. et al. Intramammary immunization with ultraviolet-killed *Escherichia coli* shows partial protection against late gestation intramammary challenge with a homologous strain. *J. Dairy Sci.* **99**, 9014–9026 (2016).
15. Herry, V. et al. Local immunization impacts the response of dairy cows to *Escherichia coli* mastitis. *Sci. Rep.* **7**, 3441 (2017).
16. Bonnefont, C. M. D. et al. Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. *BMC Genomics* **12**, 208 (2011).
17. Bouvier-Muller, J. et al. Whole blood transcriptome analysis reveals potential competition in metabolic pathways between negative energy balance and response to inflammatory challenge. *Sci. Rep.* **7**, 2379 (2017).
18. Sbardella, A. P. et al. RNA sequencing differential gene expression analysis of isolated perfused bovine udders experimentally inoculated with *Streptococcus agalactiae*. *J. Dairy Sci.* **102**, 1761–1767 (2019).
19. Mitterhuemer, S. et al. *Escherichia coli* infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genomics* **11**, 138 (2010).
20. Burvenich, C., Van Merris, V., Mehrzad, J., Diez-Fraile, A. & Duchateau, L. Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet. Res.* **34**, 521–564 (2003).
21. McMaster, S. R. et al. Pulmonary antigen encounter regulates the establishment of tissue-resident CD8 memory T cells in the lung airways and parenchyma. *Mucosal Immunol.* **11**, 1071–1078 (2018).
22. Matsuzaki, G. & Umemura, M. Interleukin-17 as an effector molecule of innate and acquired immunity against infections. *Microbiol. Immunol.* **51**, 1139–1147 (2007).
23. McAleer, J. P. & Kolls, J. K. Mechanisms controlling Th17 cytokine expression and host defense. *J. Leukoc. Biol.* **90**, 263–270 (2011).
24. Porcherie, A. et al. IL-17A is an important effector of the immune response of the mammary gland to *Escherichia coli* infection. *J. Immunol.* **196**, 803–812 (2016).
25. Blum, S. E., Heller, E. D., Jacoby, S., Krifucks, O. & Leitner, G. Comparison of the immune responses associated with experimental bovine mastitis caused by different strains of *Escherichia coli*. *J. Dairy Res.* **84**, 190–197 (2017).
26. Roussel, P. et al. Investigating the contribution of IL-17A and IL-17F to the host response during *Escherichia coli* mastitis. *Vet. Res.* **46**, 56 (2015).
27. Rainard, P. et al. Antigen-specific mammary inflammation depends on the production of IL-17A and IFN- γ by bovine CD4⁺ T lymphocytes. *PLoS ONE* **10**, e0137755 (2015).
28. Strutt, T. M. et al. IL-15 supports the generation of protective lung-resident memory CD4 T cells. *Mucosal Immunol.* **11**, 668–680 (2018).
29. Bergsbaken, T., Bevan, M. J. & Fink, P. J. Local inflammatory cues regulate differentiation and persistence of CD8⁺ tissue-resident memory T cells. *Cell Rep.* **19**, 114–124 (2017).
30. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
31. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
32. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
33. Chaussabel, D. et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* **29**, 150–164 (2008).
34. Li, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat. Immunol.* **15**, 195–204 (2014).
35. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
36. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
37. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.4.0. <https://github.com/kevinblighe/EnhancedVolcano> (2019).
38. Marini, F. & Binder, H. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics* **20**, 331 (2019).
39. Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.12. <https://cran.r-project.org/web/packages/pheatmap> (2019).
40. Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018).

ACKNOWLEDGEMENTS

This work was essentially supported by a grant from Agence Nationale de la Recherche (ANR) ANIHWA in the project KOLimastR (ANR-13-ANWA-0003). N.C. received a fellowship from Apis-Gen (MASTIM project) and Occitanie region (Transcriptovac project).

AUTHOR CONTRIBUTIONS

G.F. and P.R. designed the experiment; N.C., S.W., and B.G. performed experiments; S.M. analyzed sequencing data and produced table counts; N.C. and G.F. analyzed data; N.C., P.R., and G.F. wrote the manuscript. All authors contributed to the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interest.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41541-020-00258-4>.

Correspondence and requests for materials should be addressed to G.F.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

2.3. Etude 2 : Etude d'un protocole de préparation sur les paramètres sanitaires, immunitaires et les performances zootechniques de veaux dans des ateliers d'engraissement

En préparation pour soumission à *Vaccines*

2.3.1. *Objectif et protocole*

Les jeunes bovins qui sont engraisés dans des ateliers spécialisés en vue de leur abattage à l'âge relativement précoce de 18 mois, sont prédisposés aux pneumonies. Le changement du milieu de vie et le regroupement d'animaux d'origines diverses dans un même bâtiment sont propices à l'apparition de maladies respiratoires, qui ont des répercussions négatives sur la croissance et l'engraissement. Par ailleurs, le traitement de ces infections respiratoires conduit à l'administration d'antibiotiques, malgré l'objectif affiché de réduire leur utilisation.

L'objectif de cette étude est de déterminer les effets produits par un nouveau protocole de préparation des animaux avant leur entrée en atelier d'engraissement dans le but de réduire la fréquence des broncho-pneumonies. Ce protocole de préparation vise à protéger les bovins vis-à-vis des infections par le *BRSV*, le *BPI3V* et *Mannheimia.haemolytica* par la vaccination.

Pour cet essai, deux groupes de 20 animaux ont été recrutés au sein d'un ensemble plus large de 168 jeunes bovins. Le premier groupe a suivi le protocole de préparation avant l'entrée en atelier d'engraissement alors que le second qui a servi de groupe témoin a été élevé au pâturage pendant la même période. L'étude a été réalisée dans 2 ateliers d'engraissement, hébergeant la moitié de l'effectif de chaque groupe. L'ensemble des bovins provenaient de 5 élevages naisseurs différents.

Afin d'évaluer les effets de la préparation, différentes mesures ont été réalisées :

- Des enregistrements zootechniques : poids à différentes étapes de leur croissance pour le calcul du gain moyen quotidien (GMQ), du poids carcasse et de la note de conformation
- Des mesures de la réponse immunitaire générale et spécifique :
 - La réponse anticorps vis-à-vis de cinq pathogènes majeurs (*BRSV*, *BPI3V*, *BAV*, *M. haemolytica*, *M. bovis*)
 - La réponse cellulaire via le criblage par analyse multiplexe après stimulation avec différents agents pathogènes ou des ligands activateurs des leucocytes

- Des informations cliniques : divers signes synthétisés avec un score clinique chez le naisseur et après entrée dans l'atelier d'engraissement, ainsi que le score lésionnel des poumons à la fin de l'étude.

L'ensemble de ces mesures a été réalisé suivant une cinétique avec 5 temps majeurs (P0, date de début de la préparation, D0, date de l'entrée en bâtiment d'engraissement, D30, D60 et D180 correspondant à la date de la fin de l'étude).

Des informations sur des interventions vaccinales avant le recrutement dans l'étude ont également été prises en compte *a posteriori*, après le constat de différences dans le statut immunitaire des animaux au moment de l'analyse.

2.3.2. Résultats

2.3.2.1. Des différences avant la préparation ?

La vaccination dans le jeune âge et dans l'élevage naisseur, a un impact sur la réponse humorale, ainsi que sur l'état clinique de l'animal aux différentes périodes où le risque infectieux augmente. Ce critère doit donc être inclus dans les études de ce type. Le taux d'anticorps vis-à-vis de *M. haemolytica* n'est pas altéré alors que ceux contre *BPIV3* et *BRSV* sont très augmentés chez les animaux vaccinés.

2.3.2.2. Un effet positif de la préparation lors de l'entrée en engrais ?

Du fait de la quantité et de la complexité des données à analyser, des étapes de réduction de données puis de classification ont permis de catégoriser les animaux en trois groupes avec chacun des caractéristiques propres comme nous le verrons dans la partie suivante. L'analyse factorielle multiple utilisée pour agréger l'ensemble de l'information (profil immunitaire avec cytokines, anticorps, le poids, les scores clinique et lésionnel, le statut en oligo-éléments) en deux dimensions a permis de classer les animaux en 4 groupes en fonction de la vaccination dans le jeune âge et de la préparation. La classification via la méthode HCPC qui a été appliquée par la suite a permis de déterminer trois groupes, le premier correspondant aux animaux vaccinés qu'ils aient suivi la préparation ou non, le second aux animaux non vaccinés, parmi lesquels deux sous-groupes, préparés et vaccinés se distinguent.

2.3.2.3. La classification, une méthode valable pour l'étude en élevage ?

Du fait de la quantité et de la complexité des données, des étapes de réduction de données puis de classification ont permis de catégoriser les animaux en 3 groupes avec chacun leurs caractéristiques comme nous le verrons dans la partie suivante. L'analyse factorielle multiple utilisée pour résumer l'ensemble de l'information (cytokine, anticorps, poids, score, oligo-

éléments) en deux dimensions a permis de classifier les animaux en 4 groupes en fonction de l'existence ou non d'une vaccination dans le jeune âge et de la préparation. La classification via la méthode HCPC qui a suivi a permis de déterminer 3 groupes, le premier correspondant aux animaux vaccinés qu'ils soient préparés ou non, le second aux animaux non vaccinés mais préparés et le dernier aux animaux non vaccinés et non préparés.

2.3.2.4. Quelles caractéristiques pour chacun des groupes ?

Chacun des groupes possède des caractéristiques qui lui sont propres, avec un effet non attendu concernant les animaux vaccinés et préparés. En effet, ce groupe présente les taux d'anticorps vis-à-vis des valences vaccinales (*BRSV*, *BPIV3* et *M. haemolytica*) les plus faibles à la fin de l'étude, le GMQ moyen le plus faible et en conséquence un poids carcasse plus faible ainsi que le score lésionnel le plus élevé. En revanche, c'est l'opposé pour les animaux préparés qui n'avaient pas été préalablement vaccinés dans l'élevage naisseur ; ils possèdent les taux d'anticorps, le GMQ et le poids carcasse les plus élevés, et ne présentent pas de lésion pulmonaire à l'abattoir. Pour le dernier groupe, ils sont dans la moyenne entre les deux précédents, mais possèdent des taux d'anticorps et un statut en oligo-éléments plus élevés ou favorables, au début de l'étude. Les animaux ayant été vaccinés et préparés ont une réponse inflammatoire plus élevée à l'entrée en atelier d'engraissement après stimulation avec *BRSV* et *M.haemolytica* par rapport aux animaux non vaccinés indiquant un effet de la vaccination, réactivé ou non par la préparation.

Au contraire, cette réponse, 60 jours après l'entrée en engraissement, est moins forte après les mêmes stimulations chez les vaccinés qu'ils aient été ou non préparés, indiquant une baisse de la réponse induite par les différents protocoles.

2.3.2.5. Des relations fortes entre les différents types de variables ?

Une étude de corrélation utilisant le package R *Mixomics*, a révélé des liens forts entre différents types de variables quel que soit le groupe, et notamment la cuprémie à l'arrivée en atelier d'engraissement qui est corrélée au poids carcasse ainsi qu'à la concentration en haptoglobine à l'entrée en préparation. Les taux plasmatiques de cuivre et de glutathion à l'entrée en atelier, élevés chez les préparés, sont corrélés à un GMQ plus grand au début de la période d'engrais ainsi qu'à une meilleure réponse en anticorps vis-à-vis de *BRSV*.

2.3.3. Discussion

Cette étude a permis de montrer que la vaccination contre les trois principaux agents de maladies respiratoires a un effet protecteur relativement faible, et qu'il est sans effet sur les performances de croissance ou la santé de l'animal, lorsque les conditions d'hébergement sont

mauvaises. De plus certaines variables tels que la concentration en cuivre sont corrélées avec un meilleur résultat final et une meilleure santé. L'utilisation de la préparation doit se faire dans des conditions qui auront été préalablement évaluées comme favorables, ce qui signifie qu'une gestion de l'hygiène et des conduites d'élevages est nécessaire dans ce cadre. La vaccination seule ne prévient pas l'apparition des troubles sanitaires habituellement observés dans ce contexte.

Integrative analysis of vaccination against bovine respiratory disease in a preconditioning program for beef cattle

Nathan Cebron¹, Blandine Gausserès¹, Julie Cournet¹, Christian Tasca¹, Clément Boudenne¹, Lucie Dutheil¹, Elise Vanbergue², and Gilles Foucras^{1*}

1. IHAP, Université de Toulouse, ENVT, INRAE, 31076 Toulouse, France
2. Idele, Saint Gilles

*Correspondence: gilles.foucras@envt.fr

Abstract

Pre-conditioning young bulls (JB) before entry is considered a promising way to reduce the frequency of respiratory disorders during fattening. A case-control study was set up in five cow-calf farms and two fattening units to compare preconditioned calves (PREC) with control calves (CTRL) in the French context (Vanbergue et al., 2021). Contrary to expectations, this protocol did not reduce the incidence of infections, nor improve the performance and health of the PREC bulls. A detailed study on immune status involving 40 animals was carried out. An integrative study taking into account growth results, clinical and lesional scores, and analysis of the immune response is presented. This study showed that animals vaccinated prior to preconditioning had a high lesion score, a lower than average carcass weight and a low immune response to *BRSV* and *M. haemolytica*. On the contrary, the animals that fared best were those that were prepared at the time they entered the facility and had not been vaccinated in the breeder farm. Preconditioning and vaccination do not have the expected positive effect when they are carried out under unfavourable housing conditions.

Keywords: preconditioning, immunity, BRD, vaccination, cattle

Introduction

Bovine Respiratory diseases (BRD) account for 75-80% of the health problems observed in the weeks following the introduction of young calves in feedlots [1]. These are multifactorial infectious diseases caused by a synergistic infection between several pathogens of bacterial

(*Mannheimia haemolytica*, *Pasteurella multocida*, *Histophilus somni* and *Mycoplasma bovis*) and viral (*BRSV*, *BPIV3*, *B CoV*; *BVDV*) types.

The susceptibility of young bulls to BRD at this stage is explained by physiological and individual factors (low respiratory capacity of cattle, breed, genetics, conformation [2], by breeding and environmental factors (season, atmosphere of the rearing building, feed) [1] and distance between breeding farm and fattening facilities.

Abrupt weaning, transport, repeated handling, mixing of animals from different origins and social stress are all factors that favour the development of BRD [1], [3]–[6]. These factors have an impact on the exposure of young bulls to respiratory pathogens and lead to a decrease in the systemic and respiratory immune response.

In addition to an impact on animal welfare and performances [7], BRD require curative antibiotic treatment when disorders develop. In a context of reducing antibiotic use, alternatives are sought through changes in the way the animals are run or within the sector organisation.

In France compared to North America where the protocols are more complete, these programmes are based on the implementation of vaccination protocols against respiratory pathogens (*BRSV*, *BPIV3* and *Mannheimia haemolytica*) in breeder farms. Several reviews report a positive effect of these pre-conditioning programmes in the United States on morbidity, mortality and performance of young bulls [8]–[10].

As North American breeding conditions are different from those implemented in France, the objective of the WelHBeeF project is to evaluate the relevance and results of such programmes in the French context (Vanbergue et al, 2021). The preconditioning protocol used here included: weaning 50 days before departure to the sorting facility, reinforcement of immunity by a trivalent vaccination (*BRSV*, *BPIV3*, *Mannheimia haemolytica*) and a balanced diet with trace elements (copper, zinc, selenium), and habituation to the stall. This protocol did not reduce the incidence of infectious bronchopneumonia and did not improve zootechnical performances. To understand this lack of effect, an integrative analysis of 40 animals was carried out, taking into account various information at different follow-up dates (immune, zootechnical, score).

Materials and Methods

Experimental protocol

In this study, 40 young bulls were followed in two fattening units to evaluate their response to a vaccine program and to determine the impact of a preconditioning programme on their general

responsiveness and against specific respiratory infections. This study is part of a large trial on 168 young to evaluate the effects and benefit of preconditioning that was recently reported (Vanbergue et al, 2021). Preconditioned animals followed a vaccination protocol against respiratory pathogens (*BRSV*, *BPIV3* and *M. haemolytica*) with a commercial vaccine (Bovalto 3, Boehringer Ingelheim, Lyon, France). Operators were blind to the vaccine status of each animal before inclusion in the study.

Bulls also received a deworming treatment at sorting. The ration was supplemented with adequate amounts of trace minerals and vitamins to cover animal needs.

Experimental data

Experimental design has been previously reported and is provided for information in the **supplementary figure 4** with all recordings (Vanbergue, 2021). Animal data like body weight (BW), average daily gain (ADG) were previously reported (Vanbergue, 2021). Clinical scores were collected at 0 and 15 days later during preconditioning (P0, P15), and at 1, 4 and 15 days after arrival (D0, D4, D15). Post-mortem observations were performed at slaughterhouse to determine lung lesions that were reported as a lung lesion score, on a scale between 0 and 52, according to the extent and severity of the lesions.

Laboratory data

Several biochemical parameters were assessed in relation with trace element status, inflammatory response and immune status against BRD pathogens.

Copper and Zinc concentrations were determined by chromatography on plasma samples collected on Lithium heparinate Monovette syringes (Starstedt). Selenium status was determined on the basis of the glutathione peroxidase activity in blood erythrocyte pellet.

For inflammatory status, haptoglobin concentration was determined by ELISA (BioX, Rochefort, France) on plasma samples. Results are expressed as the percentage of a ratio to positive control.

Assessment of immunity against BRD pathogens and general responsiveness

Antibody response

Serum samples were collected at P0, D0, D30, D60 and D180 (day before slaughtering) on dry tubes. Antibody titres against *BRSV*, *BPIV3*, *BAV*, *M. haemolytica*, *M. bovis* were determined

using a commercially available kit (BIOK 369, BioX, Rochefort, France). Results are expressed as a percentage relative to the sample to positive control ratio.

Cell-mediated response assessment using cytokine release assay

Blood samples were collected on arrival, and 30 and 60 days later. Briefly, blood samples were collected in six pre-filled Monovette syringes (Sarstedt), with different ligands (LPS, Concanavalin A) or whole antigens from respiratory pathogens (*BRSV*, *M. haemolytica*, and *M. bovis*). Heat-killed bacteria are a kind gift of Aude Ferran and Eric Baranovski, respectively. A non-stimulated condition with vehicle only was also included. Syringes were incubated at 38.5°C for 24 hours before centrifugation at 2500 rpm for 15 min at 4°C. Plasma was collected and stored at -80°C until analysis. Cytokine production was measured with a custom bovine cytokines/chemokines MilliPlex xMAP assay (SPRUS706, Merck Millipore). Data were recorded on a MAGPIX flow cytometer using Xponent software (Luminex).

Data analysis and statistical methods

Cytokines data were first transformed by log and adjusted with a regression for normalization using *Cytomod* R package. Clusterisation was performed for each time and each stimulus. For each condition, the 15 cytokines were grouped in 2 to 5 clusters. Cytokine classification was determined with a Pearson correlation. The values of cytokine MFI or concentrations for one cluster were added to obtain the value of this cluster. These results were used in the multivariate analysis instead of data from individual conditions.

Dot plot representation was performed with GraphPad prism version 7.0 for Windows (GraphPad Software, La Jolla California USA). All variables were presented with mean values + SD for n = 10 or 20 calf per condition. The individual values are shown as a dot. Statistical analysis was performed with R by applying Mann-whitney test or Kruskal-Wallis test depending on the analysis. A threshold below 0.05 was considered as significant.

Four categories of data (Clinical and lesion scores, Weight, Antibody response, Cell-mediated response) were defined to perform a multiple factor analysis (MFA) with *FactomineR* R package. A classification with Hierarchical Clustering on Principal Components based on k-means and Agglomerative Hierarchical Clustering (AHC) after MFA was performed. The optimal number of clusters was not predetermined and automatically calculated by the software.

Correlation analysis between variables was performed with *Mixomics* R package. For this study, the framework DIABLO of *Mixomics* was used. The optimal number of components was determined with the function `ncomp`. A correlation cut-off value of 0.7 was used.

Results

Experimental design

In total, forty animals that originated from five different cow-calf herds were recruited randomly on criteria presented in **Table 1**. They were subdivided into two groups that were directed to two fattening units after sorting. Group allocations between PREC and CTRL was performed in the breeding herd. All animals were of the Charolais breed, and no difference of age and weight (**Figure 1A**) was noticed after group allocation. At sorting, haptoglobin concentration was very low and similar between the groups (**Figure 1B**). At P1, clinical scores were low and did not differ despite the fact that a few individuals had weak clinical respiratory signs (**Figure 1C**). Antibody responses against *BRSV*, *BPIV3* and *M. haemolytica* were determined at the end of the study and were blind to the operators. They were similar between the experimental groups. However, half of the animals had an antibody response in each group against *BRSV* and *BPIV3* (**Figure 1D and 1E**). No difference for *M. haemolytica* was present although a few animals had antibodies against the bacteria (**Figure 1F**). Information on previous vaccination or infection before inclusion in the study was not available at sorting and was obtained during data analysis, showing that these cattle were indeed vaccinated against *BRSV* and *BPIV3* using different protocols, as indicated in **Table 1**. This information was considered later on for data analysis.

Group comparison after preconditioning

At arrival, mean BW was not significantly different between the groups (**Figure 2A**), as was the mean clinical score at D1 (**Figure 2C**). Conversely, mean haptoglobin concentration was significantly increased in the CTRL group (172 ± 319 $\mu\text{g/mL}$), but was low (9 ± 16 $\mu\text{g/mL}$) in the PREC cattle indicating a great difference of the inflammatory status (Mann-whitney, $p < 0.01$) (**Figure 2B**). Cupremia was the same in both groups (**Figure 2D**), but zincemia (10 ± 2.2 and 13.3 ± 2.7 in CTRL and PREC, respectively, **Figure 2E**) and Selenium status-indicative glutathione peroxidase activity (316 ± 109 and 433 ± 60 U/gHb, for CTRL and PREC, respectively, **Figure 2F**) were significantly higher in the PREC group (Mann-Whitney, p -

value < 0.001). At arrival, antibody concentrations against respiratory pathogens included in the vaccine were significantly higher in PREC, with a value of the S/P ratio multiplied by 2.6 for *BRSV*, 4.2 for *BPIV3* and 2 for *M. haemolytica* antibody, indicating a clear seroconversion in the PREC bulls against the vaccine antigens (**Figure 2G-I**), while the antibody response has not changed in the CTRL.

Effects of preconditioning on immune status and growth

The effect of preconditioning was evaluated at five critical dates, just after arrival and during the first two months. Growth was evaluated through calculation of the average daily gain (ADG). During preconditioning, ADG was higher in the PREC group, but this difference was not present any more at D30. At D60, the difference was even the opposite with a more rapid growth in the CTRL group. Mean ADG was the same in both groups during the last four months, (**Figure 3A**), and during the entire study; mean body and carcass weight were thus not different at slaughter (**Figure 3B** and **3C**). As expected, the clinical scores increased in some animals during the first two weeks after arrival, but no difference was observed between PREC and CTRL (**Figure 3E**). At slaughter, careful examination of the lungs enabled establishment of a lesion score based on diffusion and intensity of the lesions. Conversely to expectations, the mean lesion score was higher in PREC than in CTRL cattle (**Figure 3F**). Antibody response against *BRSV*, *BPIV3* and *M. haemolytica* was increased in the PREC group at D0 (**Figure 2G-I** and **Figure 3G-I**). At D30 and D60, no difference was detected between the groups. At D180, antibody responses against pathogens except *BPIV3* were higher in the CTRL group compared to the PREC group (**Figure 3G-I**). In parallel, the cellular response was studied using quantification of cytokine release with multiplex bead-based method upon stimulation with general immune ligands like LPS and ConA, or specific antigens for *BRSV* and *M. haemolytica*, the two most frequent respiratory pathogens in cattle. Rather than comparing individual conditions, data were clustered using CYTOMOD module. The result of clustering is shown in **Table S1**. In the unstimulated condition, the cellular response was similar between groups at D0 and D30 (**Figure 4A** and **4B**). At D60, cluster 1 expression was significantly increased in the CTRL group. This cluster was correlated with recruitment functions as it contains mostly chemokines (**Figure 4C**). For *M. haemolytica* stimulation, no difference was present at D0 (**Figure 4D**) but at D30, cluster 1 value that correlated with inflammatory response was increased in PREC group (**Figure 4E**), while, at D60, cluster 3 value was increased in CTRL group (**Figure 4F**). For the *BRSV* condition, no difference was shown at D0 and D60 (**Figure 4G** and **4I**) between

groups. At D30, values of clusters 2 and 4 were increased in the CTRL group (**Figure 4H**). The first cluster was related with the interferon response whereas the second was correlated with recruitment and inflammatory response. Despite a high variability amongst cattle, a difference of the cell responsiveness assessed through cytokine secretion is detected between the groups.

Unsupervised analysis identifies subgroups based on preconditioning and former vaccination

To compare the groups and determine contributing variables using an unbiased approach, a multivariate analysis and a classification method were applied. All data were grouped into four datasets. The first dataset contains antibody responses against *BRSV*, *BPIV3* and *M. haemolytica* at all times points (P0, D0, D30, D60, D180). The second dataset includes clinical scores (P1, P15, D1, D4, D15), lung lesion scores and trace element statuses at J0. The third dataset contains growth performances as mean ADG, weight (P0, D0, D30, D60 and D180) and carcass weight. The last dataset includes clusters from the aggregative process with CYTOMOD R package of the cytokine release values against *BRSV* and *M. haemolytica* antigens at D0 and D30. A list of all variables is presented in the **Table S3**. With these datasets, a multiple factor analysis (MFA) was performed showing that animals are separated by the vaccination status before inclusion (VAC vs NO-VAC) in the study for the first component (MFA axis: 16.2%), and by preconditioning (PREC vs CTRL) for the second component (MFA axis: 13.6%) (**Figure 5A**). On the forty animals, one animal was excluded from the analysis due to high inflammatory cytokine expression indicating that the animal was ill at arrival (data not shown). Each group was characterized by several variables of each dataset. NO VAC-PREC group is characterized by higher zinc plasma concentration, higher glutathion peroxidase activity, and higher antibody responses at D0. VAC PREC group is characterized by higher clinical and lung lesion scores at D1 and D15. NO VAC CTRL group has higher antibody responses at D60 and D180 (**Figure 5B**). Next, we performed a hierarchical clustering on principal components (HCPC) analysis to classify animals with an unsupervised method. This method was based on the k-means method and Agglomerative Hierarchical Clustering (AHC). The optimal number of clusters was determined by the software, and three clusters were determined by the HCPC method (**Figure 5C**). The first cluster (C1) contains 18 animals, with half of them belonging to the VAC PREC group and half to the VAC CTRL group; these animals were characterized by their vaccine status before the beginning of the study. The second cluster (C2) contains 11 animals, 10 of them from the NO VAC CTRL group and the remaining

to the VAC CTRL group. The last cluster (C3) contains 10 animals of which 9 were from the NO VAC PREC group and 1 to the VAC PREC group (**Figure 5D**).

Characterisation of the identified clusters according to preconditioning and vaccination

After this analysis, a cluster characterisation was performed on all variables to determine which factors are different amongst the clusters. At D0, ADG was significantly higher in NO VAC PREC cluster at D0 but progressively declined in this group compared to the other groups (**Figure 6A**). Carcass weight was higher in the NO VAC CTRL cluster compared to the VAC cluster (**Figure 6B**). Lung lesion score was lower in the NO VAC CTRL group compared to the others groups (**Figure 6C**). Copper concentration at D0 was lower in the VAC CTRL and PREC compared to the other groups (**Figure 6D**) while zinc plasma concentration and GPX activity were higher in the NO VAC PREC group at D0 (**Figure 6E** and **6F**). *BRSV* and *BPIV3* antibody titres at P0 were higher in VAC bulls compared to the animals in other groups. At D0, antibody concentrations against the three pathogens were more elevated in the NO VAC PREC cluster than in the other clusters. Antibody concentrations of the NO VAC CTRL were lower against all pathogens. At D30, values were the same compared to D0, but the concentration was higher in the VAC than the No VAC PREC group against all three pathogens. At D60, *BRSV* antibody concentration was lower in the VAC cluster, and at D180 the antibody concentrations against *BRSV* and *M. haemolytica* were higher in the NO VAC CTRL group compared to the others groups (**Figure 6G-I**).

For the cytokine response at D0, *M. haemolytica* stimulation, cytokine cluster 3 containing most of the inflammatory cytokines was more expressed in VAC compared to NO VAC CTRL group (**Figure 7A**). Many clusters had higher values in the VAC group at D30 compared to the other groups like cluster 1 (inflammation response), cluster 2 (cellular recruitment), cluster 3 (IL-17 response) and cluster 5 (Interferon response) (**Figure 7B**). At D60, cytokine cluster 1 linked to cellular recruitment was more expressed in VAC compared to NO VAC CTRL group. Cytokine cluster 3 at D60 was more activated in NO VAC CTRL group compared to the others and was correlated with inflammatory response (**Figure 7C**).

After *BRSV* stimulation at D0, cytokine cluster 1 related to inflammation and cellular recruitment was more activated in NO VAC CTRL compared to the VAC group (**Figure 7D**). At D30 after the same stimulation, cytokine cluster 1 correlated with inflammation response and was more expressed in the VAC group whereas cluster 3 expression related to lymphocyte

response was increased in the NO VAC CTRL group. Cytokine cluster 4 (cellular recruitment) was lowly expressed in the NO VAC PREC group compared to the other groups (**Figure 7E**). Cytokine cluster 1 (cellular recruitment) and cluster 3 (lymphocyte response and inflammation response) were less expressed in VAC (**Figure 7F**). At D0, the cellular response in the control condition with no stimulation, lymphocyte (ConA) and innate response (LPS) activation, no difference was present amongst the groups compared to D30 (**Figure S1**).

To further describe the VAC group and interrogate if preconditioning had an effect on the immune response of these animals, a MFA was performed on VACC cattle using the same variables (**Figure S2A**). The analysis clearly separated CTRL and PREC bulls amongst the VAC cattle, indicating that this set of variables enables a clear separation according to preconditioning. Only one outlier (4440) with high carcass, D180 BW, and high antibody levels against *M. haemolytica* at D180 was removed (**Figure S2B**). A HCPC classification was performed and 3 clusters were determined (**Figure S2C**). The first cluster contains 11 animals, all CTRL but one. The second cluster was formed by 8 PREC animals. The last cluster corresponds to individual 4440 (**Figure S2D**).

Several descriptive variables enabled separation between CTRL and PREC. ADG at D0 was increased in the PREC group but the difference was then absent for the rest of the study (**Figure S3A**). Haptoglobin concentration wasn't different at P0 and D30 but it increased at J0 in CTRL group (**Figure S3B**). Carcass weight was the same between groups (**Figure S3C**).

The clinical score in Farm and Fattening unit were not different between the groups (**Figure S3D** and **S3E**). Lung lesion score was increased in the PREC group (**Figure S3F**). Zinc and copper concentrations were the same between groups but glutathion activity was increased in the PREC group (**Figure S3G-I**). At D0, antibody concentrations against *BRSV*, *BPIV3*, *M. haemolytica* was increased in the PREC group compared to the CTRL group due to preconditioning. At the end of the study (D180), *M. haemolytica* and BRSV-specific antibody concentrations were increased in the CTRL group. At the other time points, no difference between the groups were identified (**Figure S3J-L**).

Correlation analysis

A correlation analysis was performed with the framework DIABLO of Mixomics R package to determine the links between variables. Six variable datasets were prepared with Antibody, Cytokines, Growth records, Score, Trace elements and Haptoglobin, and a circosplot analysis was performed with a correlation cut-off set to 0.7. Many correlations were shown in blue for negative correlation and red for positive correlation (**Figure 8A**). To go further in the analysis,

a network analysis was also prepared with the Network function of Mixomics package. Relations between variables are shown in red color for a positive correlation and green color for a negative correlation. Each dataset was shown in specific color (Lightyellow : Zootechnical performance, Purple : Trace-elements, Lightblue : haptoglobin concentration, Lightgreen : Cytokines response, Red : Antibody response, Lightpink : Score). With Mixomics analysis, 3 networks were determined. Copper concentration was correlated with Growth performance like Carcass weight or Weight at D180, Haptoglobin at D0 and the clinical score at P1. Zinc concentration was anti-correlated with haptoglobin concentrations at P0 and ADG at D30 but correlated with clinical score at P15, BRSV antibody concentrations at D0, ADG at D60 and cellular response against RSV at D30 (cellular recruitment) (**Figure 8B**), suggesting that the response to the virus is protective against occurrence of health events and decreased growth.

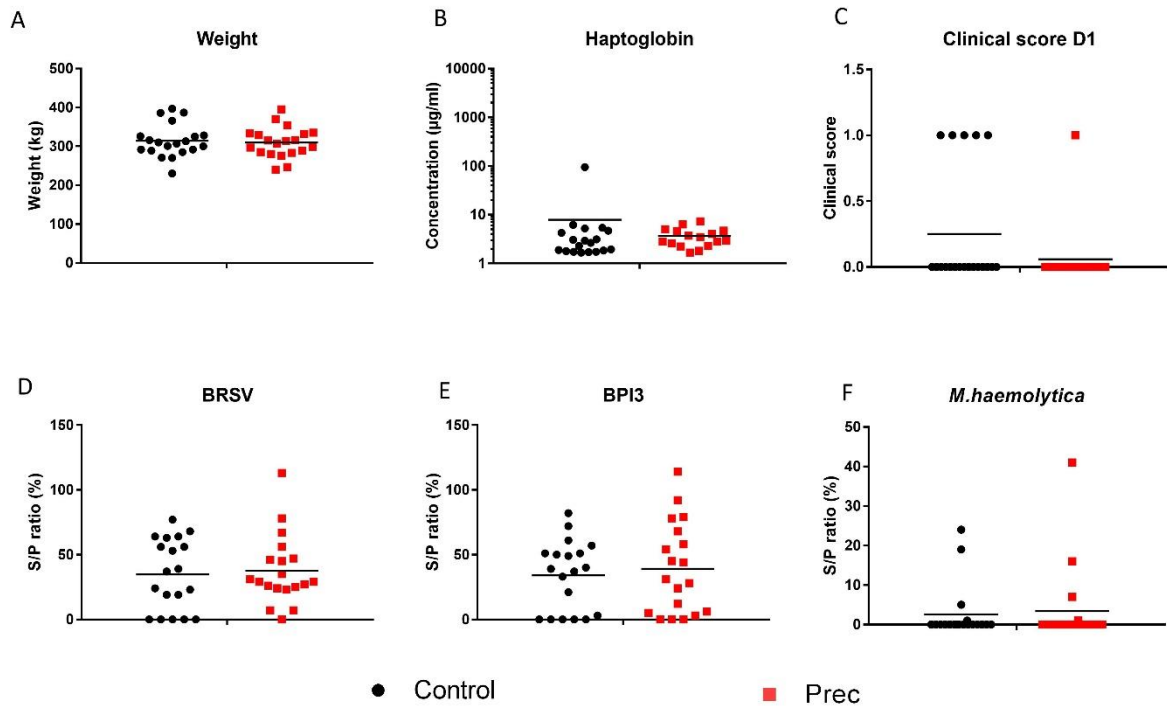


Figure 1. Characteristics of animals in each group before preparation. All data were represented as dot plot with mean. Weight was expressed in Kg (A). Haptoglobin concentration determined by ELISA was expressed in $\mu\text{g/ml}$ (B). The clinical score was based on respiratory sign observation (Cough, throwing) (C). Antibody response against *BRSV* (D), *BPIV3* (E) and *M. haemolytica* (F) was determined by ELISA and expressed in percentage compared to a positive control. The statistics values were calculated with a Mann-Whitney test to compare these two groups.

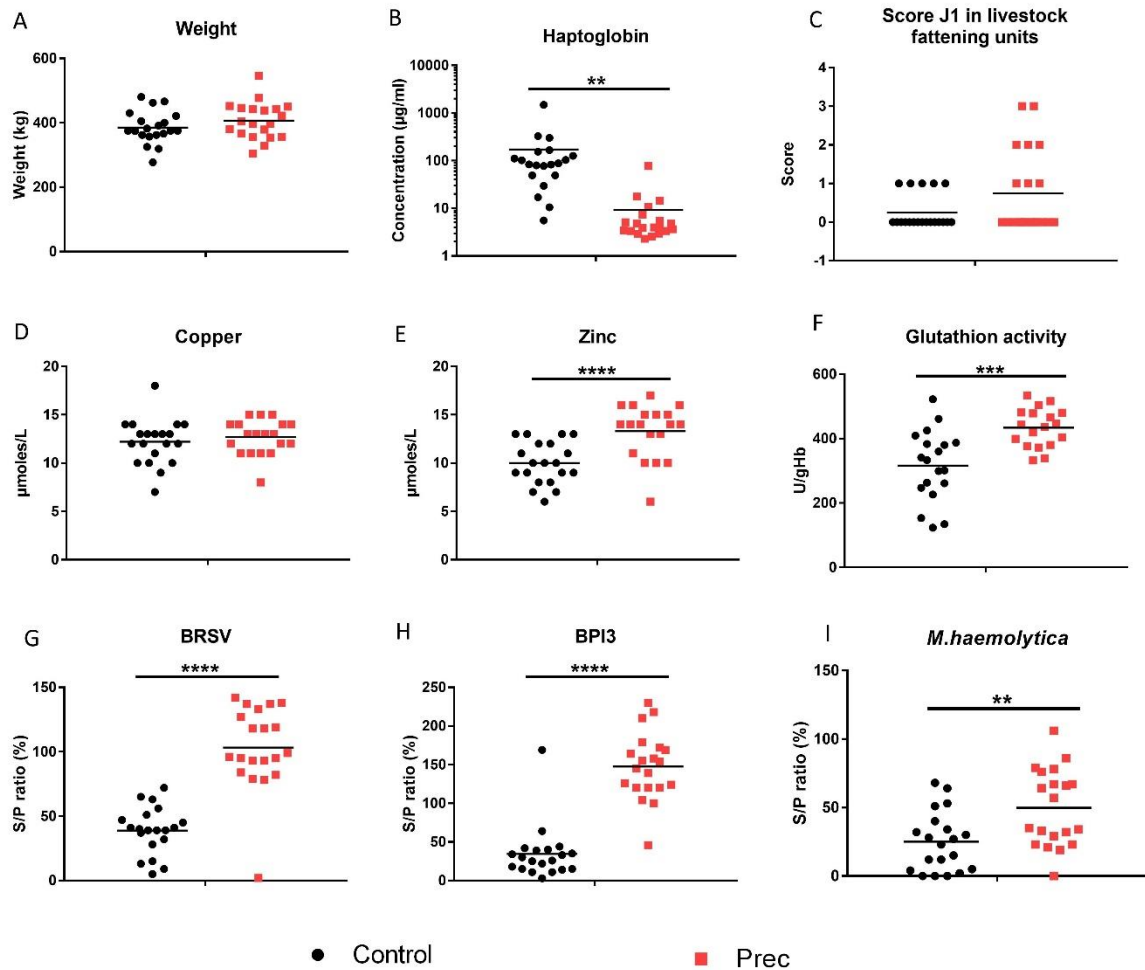


Figure 2. Characteristics of animals in each group before the entry in fattening farm. All data were represented as dot plot with mean. Weight was expressed in Kg (A). Haptoglobin concentration determined by ELISA was expressed in µg/ml (B). The clinical score was based on respiratory sign observation (Cough, throwing) (C). Trace-elements concentration like Copper (D) and Zinc (E) were expressed in µmoles/L and glutathion activity (F) in U/gHb. Antibody response against *BRSV* (G), *BPIV3* (H) and *M. haemolytica* (I) was determined by ELISA and expressed as the percentage compared to a positive control. The statistics values were calculated with a Mann-Whitney to compare these two groups.

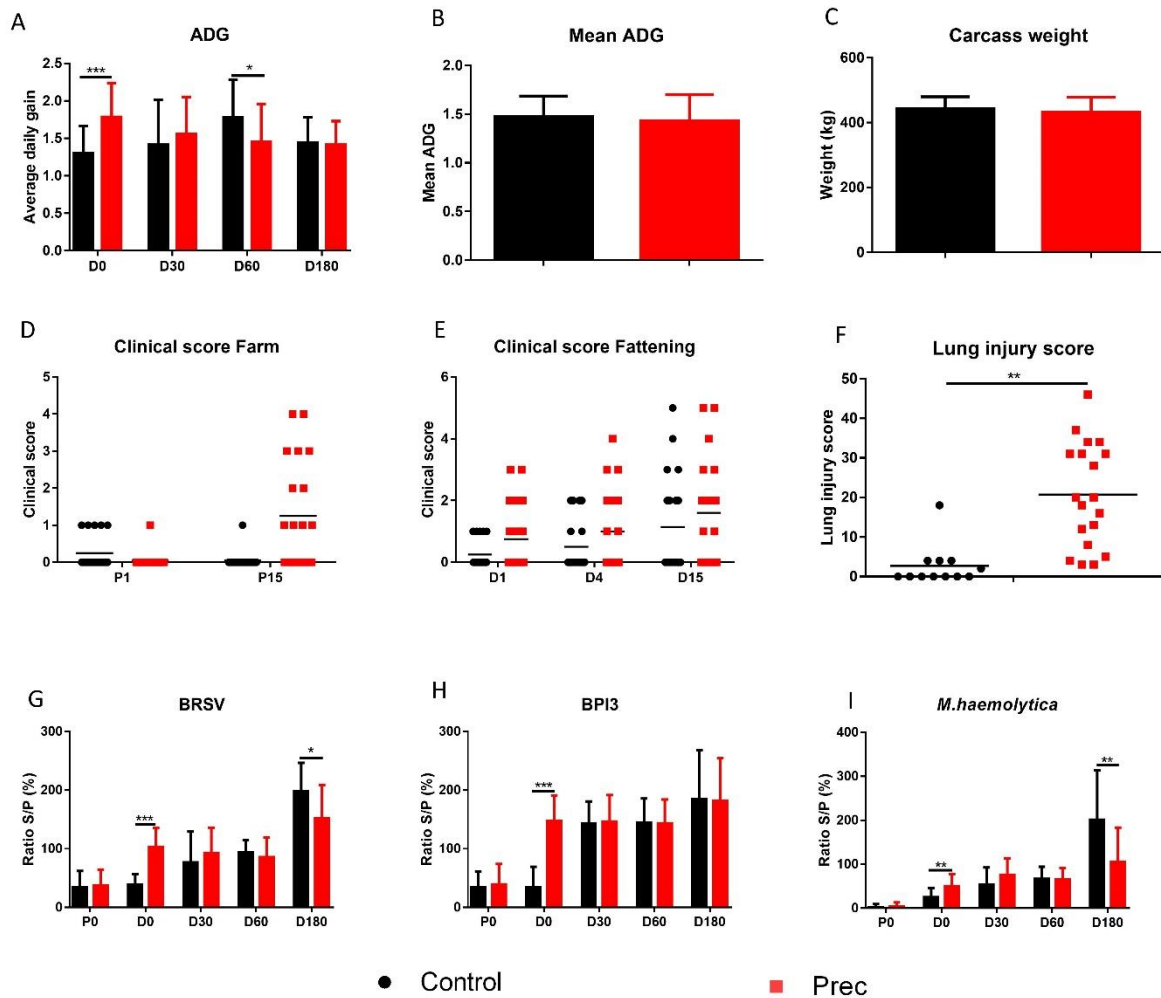


Figure 3. Effect of preparation on antibody response and zootechnical performance. ADG for each date was calculated in function of days with the before time and gained weight in this range (A). Mean ADG corresponding to mean gain per day during all experiment (B). The carcass weight was determined at slaughterhouse at the end of the study and expressed in Kg (C). These data were represented in bar plot. Clinical score and lung score were represented as dot plot with mean. Clinical score at farm (D) was determined at P1 and P15 and fattening (E) at D1, D4 and D15. These scores were based on respiratory sign observation (Cough, throwing). Lung injury score was determined at the end of study and determined the mathematical formula detailed in protocol (F). Antibody response against *BRSV* (G), *BPIV3* (H) and *M. haemolytica* (I) was determined by ELISA and expressed in percentage compared to a positive control. Antibody response was determined against each pathogen at P0, D0, D30, D60 and D180. The statistics values were calculated with a Mann-Whitney test to compare these two groups at each date.

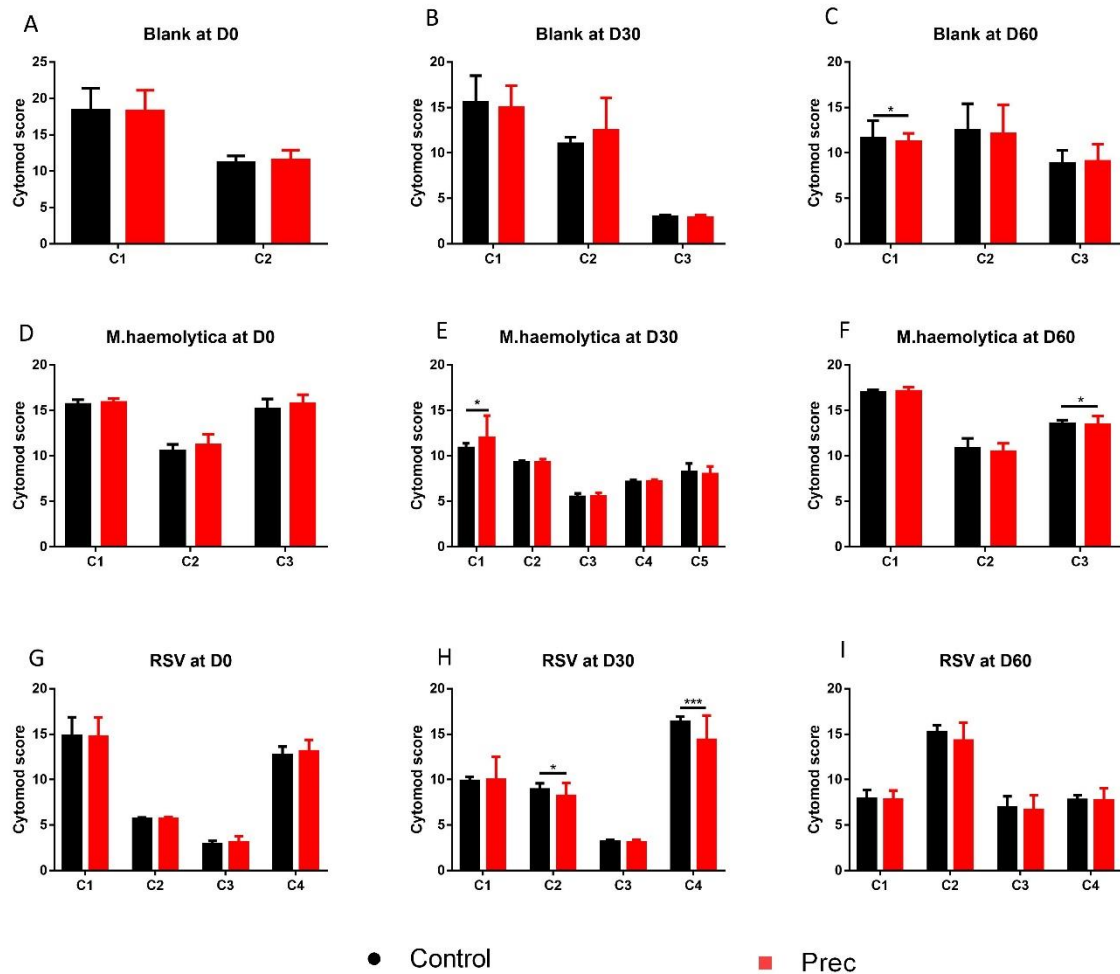


Figure 4. Cytokines response at D0 and D30. These concentrations were determined by milliplex and aggregated by CYTOMOD to obtain a cytomod score to summarize the information. This quantification after blank stimulation was determined at D0 (A), D30 (B) and D60 (C). A stimulation with *M. haemolytica* was performed and the quantification was determined at D0 (D), D30 (E) and D60 (F). Finally, a stimulation with RSV was also performed at these three times (G-I). The statistics values were calculated with a Mann-Whitney test at each date and for each cluster.

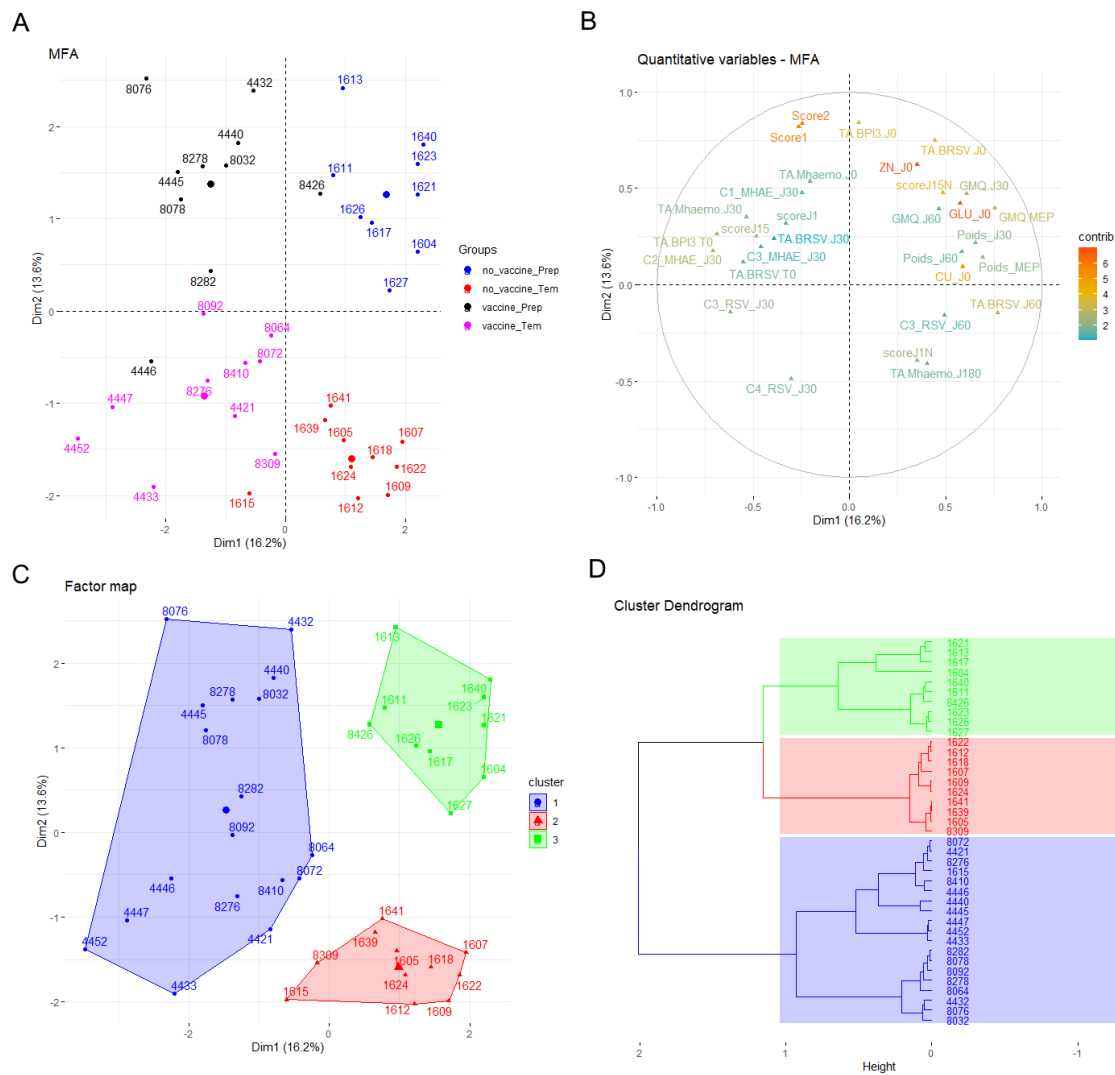


Figure 5. Multivariable analysis on forty animals. **(A)** Multiple factor analysis used zootechnical data, immunology data and score data. The color representation was based on preparation status and vaccine status before preparation. **(B)** Representation of variables contribution. The variable color was based on their contribution in the MFA. **(C)** Factor map of HCPC classification. This method used the result of MFA to classify animals with K-means and Agglomerative hierarchical clustering. **(D)** Dendrogram of each cluster displayed by color.

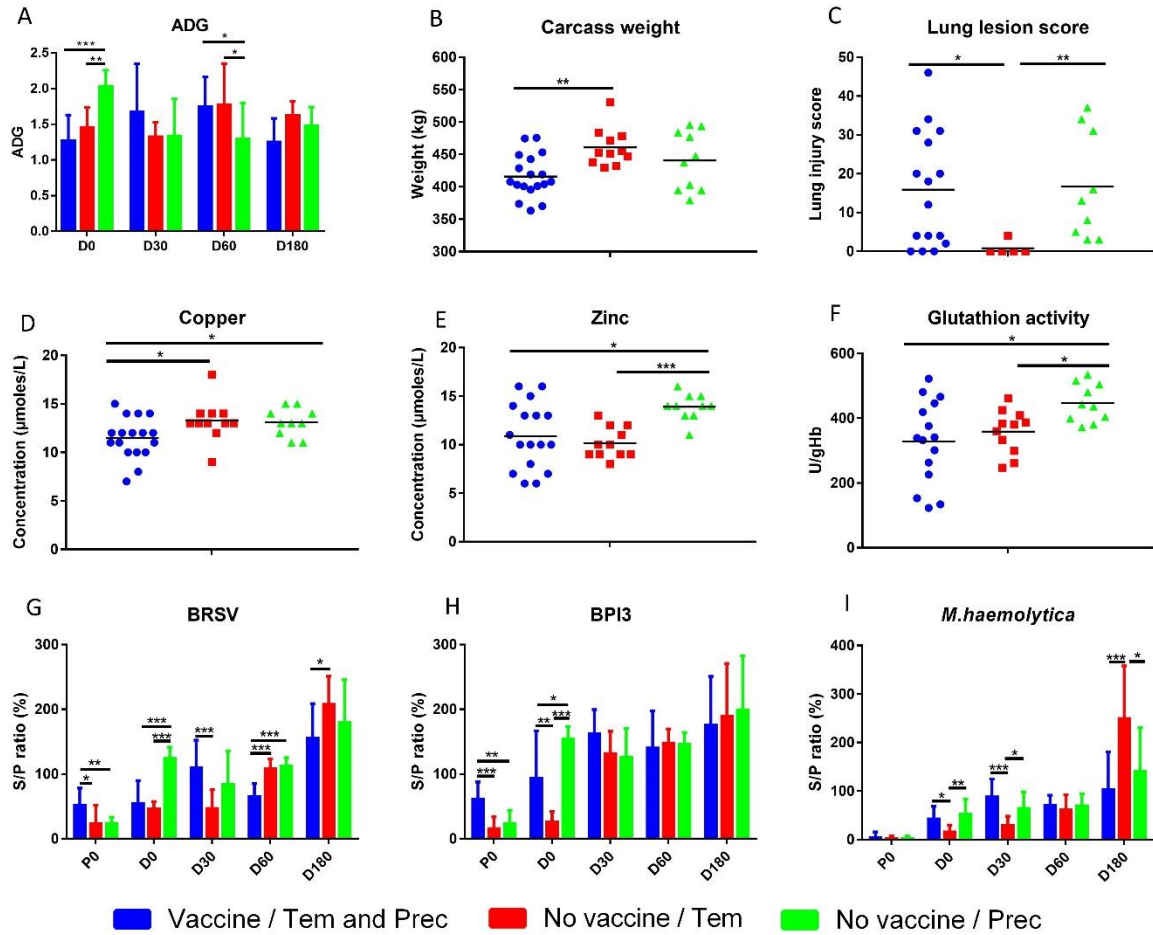


Figure 6. Antibody, growth performance and trace-elements characterization of clusters determined by HCPC. Zootechnical characterization like ADG (A), Carcass weight (B) and lung lesion score (C) were represented in bar plot for the first variable and in dot plot with mean values or each cluster for two other graphs. Dot plot with mean were used to represented Copper concentration (D) and Zinc concentration (E). These concentrations were expressed in $\mu\text{moles/L}$. Glutathion activity was shown at dot plot with mean and expressed in U/gHB (F). Barplot of antibody response against *BRSV* (G), against *BPIV3* (H) and against *M. haemolytica* (I). The characterization was realized at 5 times (P0, D0, D30, D60 and D180). The antibody response was expressed in percentage compared to a positive control. The statistics values were calculated with a Kruskal-Wallis at each date and for each cluster.

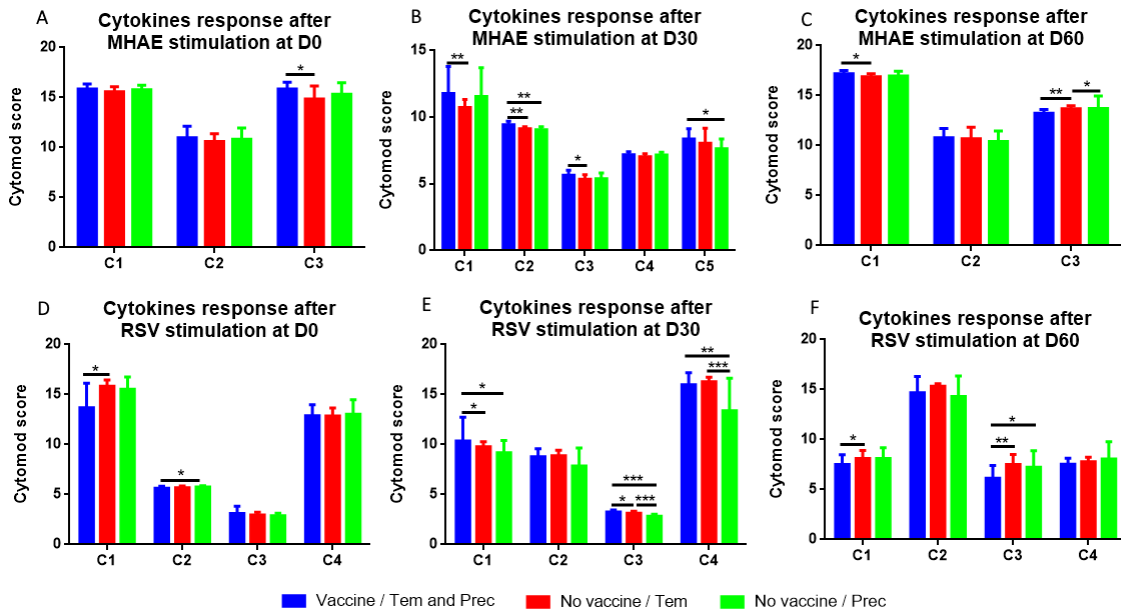


Figure 7. Cellular characterization after specific stimulation of clusters determined by HCPC. Barplot of cytokines response after *M.haemolytica* stimulation at D0 (A), D30 (B) and D60 (C). After RSV stimulation at D0 (D) and D30 (E) and D60 (F). The statistics values were calculated with a Kruskal-Wallis test at each date and for each cluster.

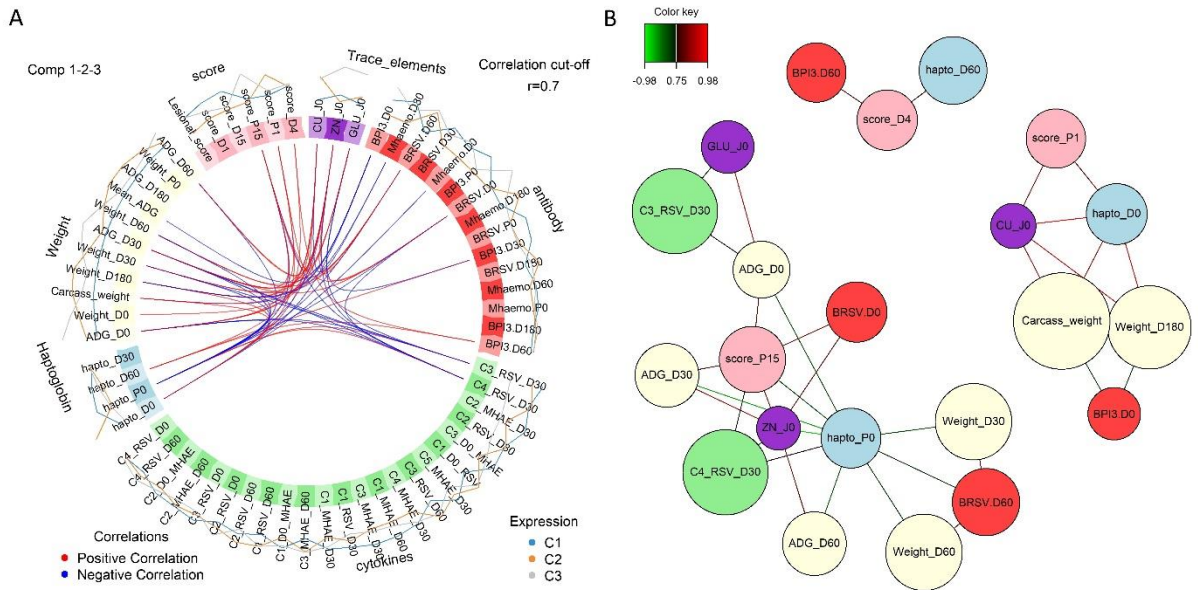


Figure 8. Association between variables. For this study, Mixomics package was used to calculate correlation between variables. Each dataset was shown in specific color (Lightyellow : Zootechnical performance, Purple : Trace-elements, Lightblue : haptoglobin concentration, Lightgreen : Cytokines response, Red : Antibody response, Lightpink : Score). Many correlations with cut-off equal to 0.7 were shown in blue for negative correlation and red for positive correlation in (A) or in green for negative correlation and red for positive correlation in (B).

Table 1. Table containing characteristics of animals number, Feed, Vaccination, Deworming for each breeder farms. The associated fattening farms was shown.

Breeder farms	Headcount	Feed	Vaccination	Deworming	Fattening farms
Farm 1	20	RustiCroiss	No	No	Fattening 1
Farm 2	8	RustiCroiss	RispovalRS 9d	Cysectin	Fattening 2
		Cerifibre	et 3/4month	RispovalRS	Cysectine
Farm 3	6	Fermier –	10d + Bovilis	Cydertine LA	
		7kg/JB/d	Bovigrip à 1,5		
Farm 4	4	Bufflo -		Cysectine	Fattening 2
		2,2kg/JB/d	RispovalRS		
		Foin –	10d		
Farm 5	2	1,5kg/JB/d			
		ToniCroiss–	RS-BVD-PI3 à	Repidose or	Fattening 2
		5kg/JB/d	8d et 3 month	Cysectin	

Discussion

In this study, we show that respiratory signs after arrival and lung lesions at slaughter were not prevented by a preconditioning program including a vaccination schedule against the three main pathogens associated with BRD in fattening cattle. Long standing effects of vaccination with a combination of intranasal and systemic injections was still detectable at sorting and arrival. Antibody and cellular responses were different amongst cattle indicating that a memory response development was present. Conversely, evaluation of the general responsiveness of the innate and adaptive lymphocytic response using LPS and ConA did not reveal any difference between PREC and CTRL animals. Despite a clear response with increased production of inflammatory cytokines with LPS or cytokines associated with lymphocyte activation, no difference could be observed between the groups, indicating that this kind of assay will not provide useful information on the immune status of the animals and their capacity to respond to an inflammatory challenge. These results show that the assays are not informative on the global immune responsiveness of the animals. Weaning, commingling, and transportation to the fattening unit are considered stress factors inducing an increase of cortisolemia usually associated with immune depression. Cytokine profiling in the null condition will neither be helpful to evaluate the immune activation in these animals although some of them had clearly heightened cytokine release probably linked to an ongoing inflammation or pneumonia.

Another observation of our study is the development of pneumonia and lung lesion in bulls that were preconditioned in facilities that were not adapted for young bull housing and that presented high risk factors for BRD development. Despite animals were from the same cow-calf herd, weaning, commingling and housing were associated with occurrence of respiratory signs, that were absent in the group remaining on pasture. These changes in herd management were sufficient to promote development of BRD. Later on, the bulls preconditioned in farm facilities that were not specifically adapted for this usage, had lung lesions at slaughter that correlated with respiratory signs during the preconditioning phase. Altogether these observations suggest that the risk of BRD development has moved from the period just after arrival at the fattening unit to the preconditioning period in farms. Indeed, risk factors for BRD predisposition in preconditioning facilities should be investigated and removed in the barn before animals can be housed for preconditioning in order to avoid or reduce the risk for BRD at this phase when animals are not yet protected. An alternative to housing will be to condition cattle on pasture when equipment for handling and feeding are available. This will avoid commingling and housing animals, two factors that are known to promote BRD development.

Another lesson learnt from our study is the limited efficacy of vaccination to prevent BRD occurrence when facing unfavorable conditions. Indeed, even if vaccination was performed earlier during the young age (between 1 to 3 months of age), immunity to *BRSV*, *BPIV3* and *M. haemolytica* did not totally prevent clinical signs and lesion development in young bulls. Several reasons may explain these observations. First, vaccination before three months of age was not sufficient to totally prevent the disease when facing the pathogens four to six months later. Despite the presence of a detectable amount of specific antibodies in vaccinated animals, this response was not fully protective. This is in accordance with previous data showing that the vaccination against BRD was short-lived and that repeated injection at six-months intervals will provide higher protection against BRD. Second, although their immunogenicity cannot be discussed, the type of immune response induced by the available commercial vaccines is not sufficient or appropriate to prevent lesion development of the respiratory tract. Third, the array of respiratory pathogens may not be covered by the vaccine. All vaccines contain at least one of the following pathogen like *BRSV*, *BPIV3* and *M. haemolytica*. However, when animals with clinical signs were sampled, other pathogens like *BCoV* and *Pasteurella multocida* or *Histophilus somni* were detected. The contribution of *Mycoplasma*, like *M. bovis*, to BRD development cannot be excluded according to the serological results and heightened antibody titres that were detected before slaughtering. The cellular response did not show up during the first two months after arrival, and was not assessed lately at the end of the fattening period. Fourth, the repetition of vaccine injection might be detrimental to the development of a protective response. A questioning observation is the fact that animals with the highest number of vaccine injections are those with the lowest antibody titres at the end of the trial and with the most severe lesions. Several hypotheses can be made to explain this unexpected result. Vaccination was implemented in several cow-calf herds that provided young bulls for the study. The use of vaccination in beef cattle where bulls are sold at weaning might be motivated by the high frequency of BRD early in life. In that case, vaccination may be a last resort for preventing BRD in a context of high prevalence of risk factors and occurrence of the diseases. Indeed, there is a correlation between vaccination and poor protection not because vaccines are poorly efficient but because they are used in a context of high BRD prevalence. Another possibility is that animals that were vaccinated before recruitment to the study are from herds where technical skills are globally poorer. A difference of genetic background cannot be excluded amongst herds where insemination is done by mounting. However, this possibility is lower since animals are from five different herds with mixing of the origins and heritability for BRD is low. The cellular response was not assessed at sorting before the preconditioning programme.

The modalities of the cell response evaluation are through cytokine release with a combined effect on innate and adaptive cells. The use of a library of peptide megapools for the main/dominant antigens of viruses and bacteria may be preferred, but some may not yet be identified for some pathogens, in order to activate T lymphocytes (CD4/CD8) in the absence of precise MHC information in the bovine species.

One caveat is the small number of cattle that were evaluated in this study. It will be helpful to increase the size of the studied groups to reduce individual and herd effects. This preliminary study has set up the conditions for a larger study or investigations on a large scale which will be necessary to ascertain our observations. It remains to be determined if our observations can be extended to other vaccines against BRD and other rearing systems (different breeds, different seasons, ...). Since vaccinated animals were from a single herd, there could be a confounding factor between the response to vaccination at the young age and later responsiveness that cannot be fully separated and serve as a confounding effect.

Références

- [1] S. Assié, N. Bareille, F. Beaudeau, et H. Seegers, « Management- and housing-related risk factors of respiratory disorders in non-weaned French Charolais calves », *Prev. Vet. Med.*, vol. 91, n° 2-4, p. 218-225, oct. 2009, doi: 10.1016/j.prevetmed.2009.06.003.
- [2] C. a. S. Rossi, R. Compiani, G. Baldi, et M. Bonfanti, « Determination and assessment of BRD risk factors in newly received beef cattle », *Large Anim. Rev.*, vol. 19, n° 2, p. 65-72, mai 2013.
- [3] L. Mounier, H. Dubroeuq, S. Andanson, et I. Veissier, « Variations in meat pH of beef bulls in relation to conditions of transfer to slaughter and previous history of the animals1 », *J. Anim. Sci.*, vol. 84, n° 6, p. 1567-1576, juin 2006, doi: 10.2527/2006.8461567x.
- [4] S. Assié, H. Seegers, et F. Beaudeau, « Incidence of respiratory disorders during housing in non-weaned Charolais calves in cow-calf farms of Pays de la Loire (Western France) », *Prev. Vet. Med.*, vol. 63, n° 3-4, p. 271-282, mai 2004, doi: 10.1016/j.prevetmed.2004.01.014.
- [5] N. Cernicchiaro, B. J. White, D. G. Renter, A. H. Babcock, L. Kelly, et R. Slattery, « Associations between the distance traveled from sale barns to commercial feedlots in the United States and overall performance, risk of respiratory disease, and cumulative mortality in feeder cattle during 1997 to 20091 », *J. Anim. Sci.*, vol. 90, n° 6, p. 1929-1939, juin 2012, doi: 10.2527/jas.2011-4599.
- [6] L. Herve, N. Bareille, B. Cornette, P. Loiseau, et S. Assié, « To what extent does the composition of batches formed at the sorting facility influence the subsequent growth performance of young beef bulls? A French observational study », *Prev. Vet. Med.*, vol. 176, p. 104936, mars 2020, doi: 10.1016/j.prevetmed.2020.104936.
- [7] N. Cernicchiaro, D. G. Renter, B. J. White, A. H. Babcock, et J. T. Fox, « Associations between weather conditions during the first 45 days after feedlot arrival and daily respiratory disease risks in autumn-placed feeder cattle in the United States », *J. Anim. Sci.*, vol. 90, n° 4, p. 1328-1337, avr. 2012, doi: 10.2527/jas.2011-4657.
- [8] G. C. Duff et M. L. Galyean, « Board-invited review: recent advances in management of highly stressed, newly received feedlot cattle », *J. Anim. Sci.*, vol. 85, n° 3, p. 823-840, mars 2007, doi: 10.2527/jas.2006-501.
- [9] F. Thrift et T. Thrift, « REVIEW: Update on preconditioning beef calves prior to sale by cow-calf producers1 », *Prof. Anim. Sci.*, vol. 27, avr. 2011, doi: 10.15232/S1080-7446(15)30452-6.
- [10] B. K. Wilson, C. J. Richards, D. L. Step, et C. R. Krehbiel, « Best management practices for newly weaned calves for improved health and well-being », *J. Anim. Sci.*, vol. 95, n° 5, p. 2170-2182, mai 2017, doi: 10.2527/jas.2016.1006.

Supplementary data

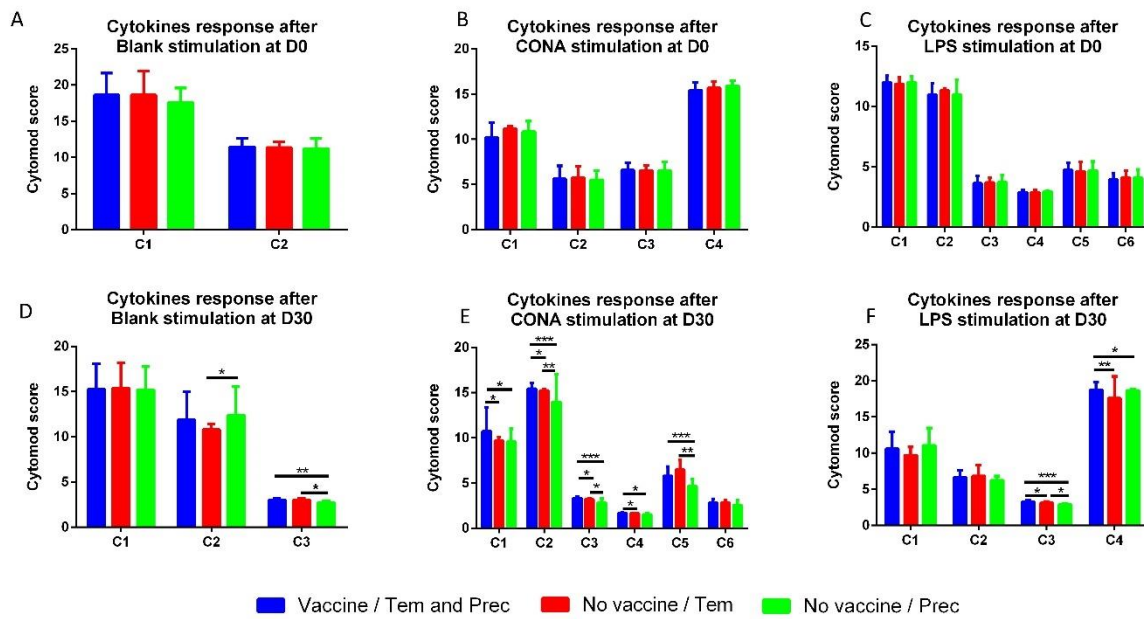


Figure S1. Cellular caractérisation after no or no specifics stimulation of clusters determined by HCPC. Barplot of cytomod score after blank stimulation at D0 (**A**) and D30 (**D**), after concanavalin a (ConA) stimulation at D0 (**B**) and D30 (**E**) and after LPS stimulation at D0 (**C**) and D30 (**F**). The statistics values were calculated with a Kruskal-Wallis at each time and for each cluster.

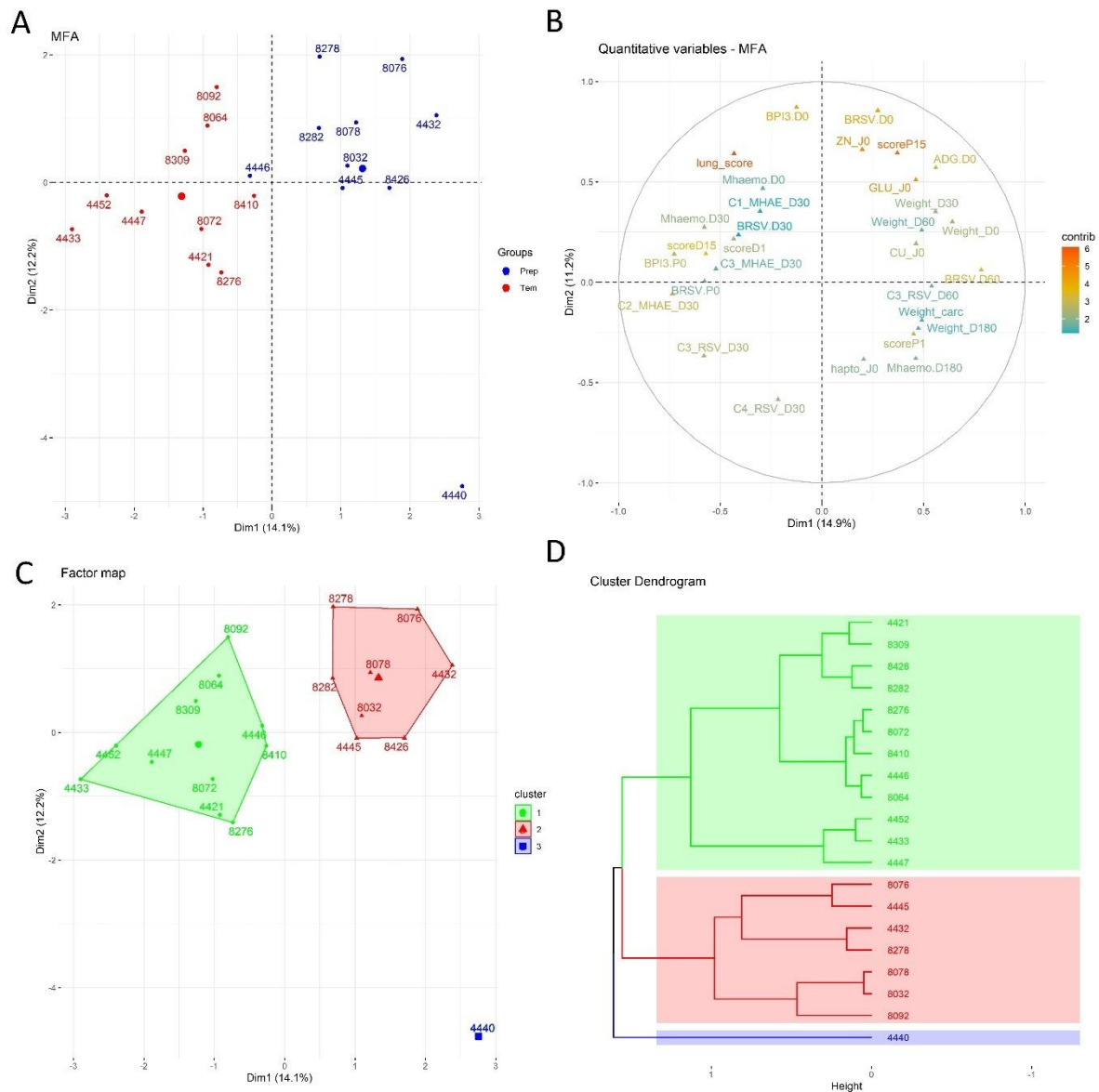


Figure S2 : Multivariable analysis on vaccine animals. **(A)** Multiple factor analysis used zootechnical data, immunology data and score data. The color representation was based on preparation status. **(B)** Representation of variables contribution. The variable color was based on their contribution in the MFA. **(C)** Factor map of HCPC classification. This method used the result of MFA to classified animals with K-means and Agglomerative hierarchical clustering. **(D)**. Dendrogram of each cluster displayed by color.

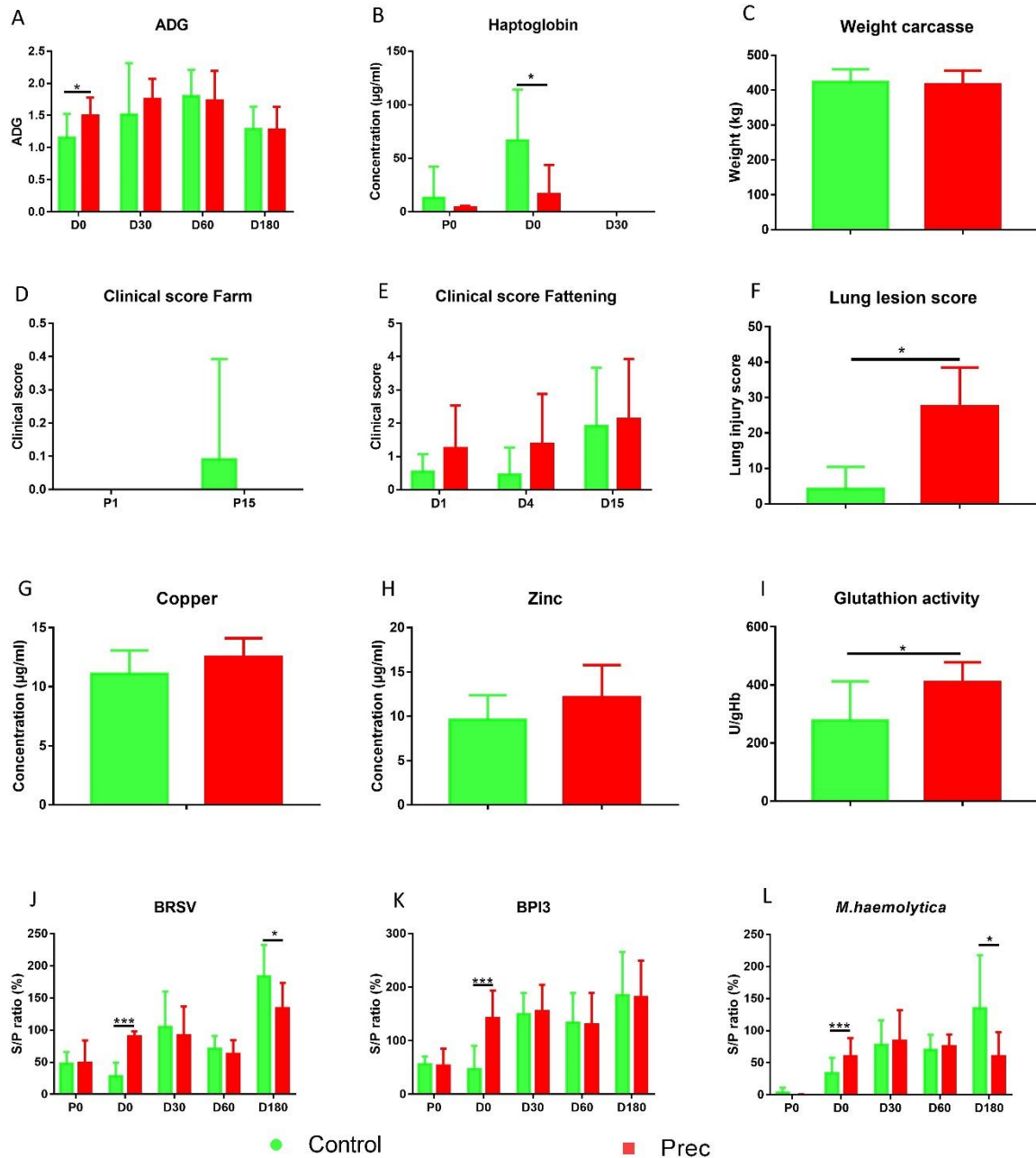


Figure S3 : Antibody, growth performance and trace-elements characterization of clusters in vaccine group determined by HCPC. ADG for each date was calculated in function of days with the before time and gained weight in this range (A). Haptoglobin concentration was expressed in concentration (B). The carcass weight was determined at slaughterhouse at the end of the study and expressed in Kg (C). These data were represented in bar plot. Clinical score and lung score were represented as barplot. Clinical score at farm (D) was determined at P1 and P15 and fattening (E) at D1, D4 and D15. These scores were based on respiratory sign observation (Cough, throwing). Lung injury score was determined at the end of study and determined the mathematical formula detailed in protocol (F). Barplot were used to represented Copper

concentration (**G**) and Zinc concentration (**H**). These concentrations were expressed in $\mu\text{moles/L}$. Gluthation activity was shown at dot plot with mean and expressed in U/gHB (**I**). Antibody response against *BRSV* (**J**), *BPIV3* (**K**) and *M.haemolytica* (**L**) was determined by ELISA and expressed in percentage compared to a positive control. Anti-body response was determined against each pathogen at P0, D0, D30, D60 and D180. The statistics values were calculated with a t-test to compare these two groups at each date.

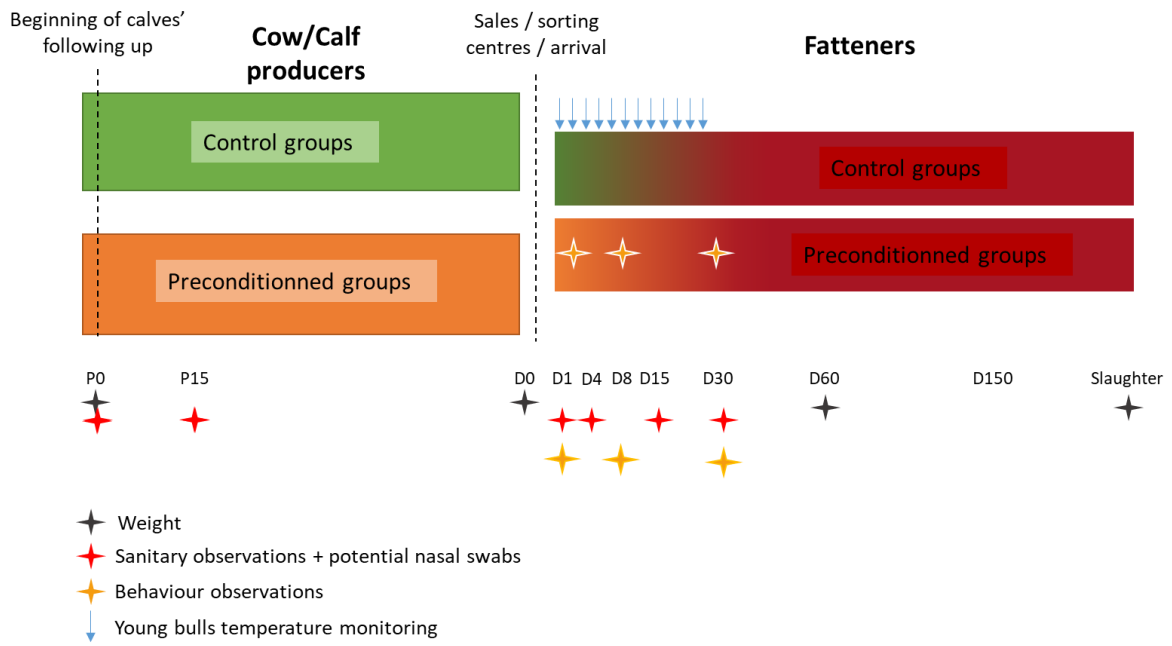


Figure S4 : Experimental scheme set up to evaluate the impact of a preconditioning program on young bulls' respiratory diseases incidence (Vanbergue and al, 2021)

	<i>Mh D0</i>	<i>Mh D30</i>	<i>Mh D60</i>	<i>RSV D0</i>	<i>RSV D30</i>	<i>RSV D60</i>
Cluster 1	IL6	IL1A	CCL3	IL1b	IL1A	CXCL10
	IL2	IL6	CCL2	IL1RA	IL6	IFNg
	IL4	IL2	IL8	CXCL8	IL2	IL4
	CXCL8	IL10	CCL4	CCL3	IL10	IL17A
	CCL3	TNF	IL17A	CCL4	TNF	
	IL10					
Cluster 2	IL17A	IL4	CXCL10	IL4	IFN	CCL2
	IFN	CXCL8	IFNg	CCL2	CXCL10	CCL3
	CXCL10	CCL3	IL1b		CCL2	CCL4
	CCL4		IL1a			IL8
Cluster 3	IL1A	IL17A	IL2	IL17A	IL17A	IL1a
	IL1B	CCL4	IL6	IFN	IL4	IL1b
	IL1RA		IL10			IL1RA
	CCL2		IL4			
	TNF		IL1RA			
		TNFa				
Cluster 4		IL1RA		IL1a	IL1B	IL2
		CCL2		IL6	IL1RA	IL6
				IL2	CXCL8	IL10
				CXCL10	CCL3	TNFa
				IL10	CCL4	
				TNF		
Cluster 5		IL1B				
		IFN				
		CXCL10				

Table S1 : Table of cytokine clusters for each specific stimulation and time. These clusters were determined with CYTOMOD python function. Cytokines were correlated with pearson correlation based on their expression.

	BL D0	BL D30	CONA D0	CONA D30	LPS D0	LPS D30
Cluster 1	IL1B	IL1B	CXCL8	IL1a	IL1RA	IL1A
	IL1RA	IL1RA	CCL3	IL6	IL6	IL6
	IL6	IFN	CCL4	IL2	CCL2	IL2
	IL17A	CXCL8		IL10	IL10	IL10
	IFN	CCL2		TNF		TNF
	CXCL8	CCL3				
	CCL2	CCL4				
	CCL3					
CCL4						
Cluster 2	IL1A	IL1A	IL1B	CXCL8	CXCL8	IL1B
	IL2	IL6	IL17A	CCL2	CCL3	IFN
	IL4	IL2	IFN	CCL3	CCL4	CXCL10
	CXCL10	CXCL10		CCL4		
	IL10	IL10				
	TNF	TNF				
Cluster 3		IL17A	IL1A	IL1RA	IL1A	IL17A
		IL4	IL2		IL2	IL4
Cluster 4			IL1RA	IL4	IL17A	IL1RA
			IL6		IL4	CXCL8
			IL4			CCL2
			CCL2			CCL3
			IL10			CCL4
Cluster 5				IL1b	IFN	
				IL17A	CXCL10	
				IFN		
Cluster 6				CXCL10	IL1B	
					TNF	

Table S2 : Table of cytokine clusters for each non specific stimulation and time. These clusters were determined with CYTOMOD python function. Cytokines were correlated with pearson correlation based on their expression.

	Variables number	Characteristics	Time used
Dataset 1	6	Trace-elements and haptoglobine response	At J0 for trace elements and P0, J0 and J30 for haptoglobin
Dataset 2	6	Clinicals score and injury score	At P1, P15, D1, D4, D15 and D180
Dataset 3	11	Zootechnical performance	At P0, D0, D30, D60 and D180
Dataset 4	15	Antibody response	At P0, D0, D30, D60 and D180
Dataset 5	23	Cellular response (<i>M.haemolytica</i> and <i>RSV</i>)	At D0, D30 and D60

Table S3 : Table described datasets used for multiple factor analysis.

3. Discussion

L'ensemble de mon travail s'inscrit dans le cadre général des études pour améliorer les connaissances sur les mécanismes de protection vis-à-vis des maladies infectieuses chez les animaux de rente comme les bovins, et d'évaluer les effets ou les limites de la vaccination pour la faire évoluer vers de nouvelles modalités ou de nouvelles conditions d'application dont nous aurons prouvées qu'elles étaient plus efficaces à protéger l'animal et à réduire l'utilisation des antibiotiques.

3.1. La vaccination, encore beaucoup de limites ?

Comme indiqué dans l'introduction bibliographique, les vaccins vis-à-vis des mammites et notamment contre celles provoquées par *E. coli*, sont encore peu efficaces. L'absence relative de progrès tient en partie à l'absence de connaissances sur les mécanismes immunitaires protecteurs vis-à-vis de l'inflammation causée par cet agent pathogène [401]. De plus, la majorité des vaccins qui ont été développés jusqu'alors contre cette maladie sont administrés par voie systémique dans le but d'augmenter la production d'anticorps. Ce procédé permet d'immuniser l'animal au niveau systémique, via une activation du système immunitaire dans les nœuds lymphatiques drainant le site d'immunisation (région de l'encolure dans la majorité des cas), mais la nature de cette réponse ou sa localisation dans le temps au moment de l'infection pourraient ne pas être suffisantes. En effet, l'une des limites de la réponse induite est le délai d'activation de la réponse immunitaire au moment de l'infection et de l'arrivée des cellules propres à l'éliminer au site infectieux. En effet, les lymphocytes éduqués sont présents dans le sang et dans les principaux centres lymphoïdes de l'organisme entre lesquels les cellules circulent. Après l'infection dans un organe ou un tissu, il faut plusieurs heures avant que les cellules soient recrutées en nombre et qu'elles se multiplient ou agissent en produisant des cytokines dans le cas des lymphocytes [402]. Durant ce laps de temps, la bactérie pathogène a largement le temps de proliférer, de libérer différents facteurs toxiques ou phlogogènes à l'origine de la mammite. Aussi, une nouvelle voie d'immunisation par injection locale de l'antigène à travers le canal du trayon directement dans la citerne de la glande a été envisagée. Néanmoins, l'immunisation locale est encore très peu utilisée sauf vis-à-vis des infections virales respiratoires chez les bovins, alors que la majorité des vaccins sont encore administrés par voie intramusculaire ou sous-cutanée. La vaccination par voie nasale fonctionne bien pour les pathogènes viraux mais ce n'est pas encore le cas pour les bactéries. Pour les BPI, ce constat

est à nuancer car en l'absence d'infection virale préalable ou concomitante, le développement de BPI bactériennes est réduit [26].

La majorité des vaccins commercialisés à destination de l'homme ou des animaux ont pour objectif la production d'anticorps, synonyme de protection. Ce corrélat est loin d'être parfait surtout vis-à-vis des bactéries intracellulaires ou même de certains virus, même si les anticorps permettent d'empêcher leur entrée dans les cellules. Pour ce qui concerne les bactéries extracellulaires comme *E. coli*, les anticorps sont un peu plus efficaces. De nombreuses études ont montré dans le cadre des mammites, la réponse cellulaire médiée par les lymphocytes T CD4+ est indispensable pour la neutralisation et la destruction du pathogène [6], [403], [404]. De nombreuses études pour l'amélioration des vaccins vis-à-vis des mammites ont déjà été réalisées [405], [406] avec, dans la quasi-totalité des cas, des vaccins administrés par voie intramusculaire ou sous-cutanée.

Notre étude se distingue par l'utilisation de méthodes d'analyse encore peu fréquentes, voire qui n'avaient pas été utilisées pour l'étude de la réponse vaccinale dans l'espèce bovine.

Les études actuelles que ce soit en santé humaine ou en médecine vétérinaire envisagent de plus en plus souvent une immunisation locale. Cette voie d'immunisation est notamment testée dans le cadre du développement d'un vaccin vis-à-vis de la COVID-19 [407].

Le protocole d'immunisation, en particulier le moment de sa réalisation ou l'intervalle entre les injections, sont des critères importants à prendre en compte pour l'efficacité de la vaccination. Dans le cas des maladies respiratoires des bovins, la première injection se fait souvent entre 3 et 6 mois. De plus, le moment de la réalisation du protocole doit être judicieusement choisi car les animaux pour des raisons économiques et pratiques ne peuvent pas recevoir des injections très fréquentes. Les conditions et le moment de la vaccination, notamment pour les BPI sont primordiaux [408]. En effet, en cas de contexte de réponse défavorable, la vaccination pourrait même être délétère, comme le suggère certains résultats de notre étude.

Pour obtenir un vaccin efficace, il est nécessaire que celui-ci cible la population cellulaire permettant le développement d'une bonne défense. Or en santé animale, toutes les connaissances concernant la protection vis-à-vis des agents pathogènes sont encore très imparfaites, et c'est notamment le cas de la mammite causée par *E. coli*. Des études conduites dans un modèle d'infection expérimentale chez la souris, ont montré que les lymphocytes T CD4+ via la production d'IL-17A permettaient le recrutement de neutrophiles et *in fine* l'élimination de la bactérie [6].

Notre étude a permis d'apporter des connaissances sur la réponse immunitaire et notamment la réponse liée à l'IL-17 susceptible de protéger la glande mammaire vis-à-vis d'une mammites à *E. coli*, par le fait que nous avons ciblé la population cellulaire responsable de cette production. Des travaux complémentaires sont nécessaires pour compléter ces résultats et valider cette approche.

3.2. La transcriptomique et la protéomique ciblée, comme moyens d'étude de la réponse immunitaire

Lors de l'étude des mécanismes immunitaires induits par la vaccination, nous avons vu que l'étude du transcriptome et du protéome ciblée (sécrétion de cytokines et de chimiokines) du sang étaient informatifs sur le déroulement de la mammites. Ils nous ont permis d'identifier et d'étudier des mécanismes mis en jeu au niveau systémique par la vaccination. L'avantage de l'étude du sang est la facilité de réalisation d'une étude cinétique pour étudier les variations de l'activité immunitaire dans le temps chez le même animal. Néanmoins, cette méthode a une limite : cette investigation ne concerne pas les mécanismes intervenant au niveau local, dans les tissus pour des cellules résidentes qui ne circulent plus ou bien en trop petit nombre pour être détectées. Par l'étude du compartiment sanguin, il est difficile, voire impossible d'étudier les réponses locales. Or lors de la production et la collecte des échantillons en 2016, avant le début de mon travail, les méthodes transcriptomiques étaient moins développées qu'aujourd'hui. La quantité d'ARN nécessaire excluait la réalisation de biopsies. Par ailleurs les questions éthiques et biologiques à cause des effets secondaires d'une biopsie chez l'animal conservé vivant empêchent la répétition des biopsies sur des pas de temps très proches, comme ceux que nous avons identifiés comme informatifs, soit environ deux fois en moins de 48 heures. Concernant l'étude des cytokines à moyen débit à l'aide de la technologie Luminex, elle était impossible à l'époque puisque la trousse destinée à cette espèce n'a été créée que plus récemment. (2018).

L'une des limites de notre étude concerne l'analyse locale de la réponse. En effet, peu d'animaux ont été utilisés pour examiner le transcriptome des lymphocytes T CD4+. Cette étape nécessite obligatoirement l'euthanasie de l'animal, pour prélever un échantillon de tissu mammaire, ce qui cause des contraintes financières et éthiques. Une augmentation de l'effectif aurait été utile pour une étude plus robuste. De plus, cette expérience a été conduite uniquement sur les groupes témoin et intramammaire, alors que l'analyse de la réponse du groupe immunisé

par voie intramusculaire aurait été utile pour connaître la nature et l'activité des lymphocytes présents ou recrutés pendant les 16 premières heures de la mammité.

L'analyse du transcriptome sanguin via la technique de RNA-seq apporte de nombreuses informations sur le contexte immunitaire général systémique. Néanmoins cette méthode a des limites, il n'est pas possible de déterminer quel type cellulaire est impliqué ou celui dont le profil est modifié, s'agissant d'une analyse globale des gènes exprimés. Pour combler cette lacune, des méthodes de déconvolution pour connaître la composition cellulaire telle que CIBERSORT ont été développées avec d'assez bons résultats. Mais cette impression est superficielle car si l'on compare la description obtenue à partir des données de déconvolution à une analyse réalisée en cytométrie de flux par exemple, il existe des divergences. A l'époque de la réalisation des essais, la technique fondée sur l'analyse des cellules isolées ou single-cell RNA-seq n'était qu'à ses débuts et n'était pas aussi développée et accessible qu'elle ne l'est aujourd'hui. Cette méthode aurait permis de connaître la proportion des différents types cellulaires ainsi que leur profil transcriptomique. De nombreuses méthodes existent et chacune possède ses avantages et ses défauts. En effet, des études de « benchmarking » entre les différentes méthodes ont été réalisées [359], [360]. Par exemple, la méthode xCell est peu fiable dans la détermination des populations de monocytes et de macrophages au contraire de la méthode MCP-counter. Il n'existe pas de méthodes fiables fonctionnant sur tous les types cellulaires étudiés dans la discipline de l'immunologie.

Il est possible de coupler la technologie de la cytométrie en flux et la méthode RNA seq pour étudier le transcriptome d'une population cellulaire en particulier. C'est ce qui a été réalisé dans l'étude des lymphocytes présents dans le parenchyme mammaire. Néanmoins, il est difficile d'étudier de nombreuses populations simultanément car il n'est pas possible de trier plusieurs populations cellulaires différentes en parallèle, à partir d'un même échantillon. Pour l'étude des réponses locales, les nouvelles méthodes tels que le single cell RNA-seq et surtout la transcriptomique spatiale sont très pertinentes et utiles. En effet, via la première, il est possible d'étudier la quasi-totalité des populations immunitaires, de connaître les sous-types et d'approcher leurs fonctions. En utilisant les bases de données, il est possible d'identifier des marqueurs spécifiques des principales populations. Néanmoins, la limite de cette approche concerne la localisation car il n'est pas possible de connaître la localisation spatiale lors de l'analyse et donc de déterminer les interactions entre les cellules. La transcriptomique spatiale permet de résoudre cette lacune et permet de comprendre si une zone particulière du tissu est différemment concernée et de voir où se trouvent les différents types de cellules immunitaires.

Dans des prochaines études, l'utilisation de cette méthode innovante pourrait apporter des informations précieuses.

L'avantage de l'ensemble de ces méthodes est l'approche non supervisée et donc sans sélection *a priori* des cellules, des gènes et des fonctions permettant une analyse globale de l'ensemble des mécanismes immunitaires qui sont développés à la faveur d'une infection. L'une des limites de ces méthodes concerne l'étude des voies de signalisation des cytokines. En effet, la majorité des cytokines et des chimiokines partagent des voies de signalisation communes. Il est donc difficile de déterminer avec précision quelle voie est activée malgré la disponibilité d'informations de mieux en mieux documentées et précises dans les bases de données qui sont nécessaires à l'analyse fonctionnelle des données du séquençage. De plus, de nombreuses cytokines sont très peu exprimées au niveau transcriptomique, comme l'IL-17A, et la sensibilité analytique peut être insuffisante pour les détecter ou identifier des différences. C'est pourquoi l'analyse des fonctions biologiques surpassent l'analyse gène à gène. Elle est devenue aujourd'hui incontournable.

L'analyse des fonctions biologiques est de plus en plus aisée car le développement des méthodes d'analyse comme nous les avons décrites dans l'introduction est important. De nombreuses bases de données tels que *Réactome* prennent en compte l'espèce d'étude, y compris *Bos taurus*. De plus certains logiciels tels que *Biomart* permettent de déterminer les orthologues d'un gène en utilisant l'espèce murine ou humaine en cas d'absence d'annotation dans l'espèce étudiée. Le développement de l'analyse modulaire a permis une étude plus ciblée des fonctions immunitaires dans le cadre de diverses infections. Cette méthode, développé par Damien Chaussabel, est très utilisée pour la résolution de questions de santé humaine [384] mais elle est aussi applicable aux études de l'espèce bovine. L'analyse modulaire ne prend pas en compte les gènes différentiels, mais la table de comptage normalisée comme déterminée classiquement par des logiciels comme *DESeq2* ou *EdgeR* [388]. Elle a l'avantage d'étudier globalement toute l'information transcriptomique et ne pas se limiter aux quelques gènes différentiels. Cette méthode est appliquée majoritairement dans des contextes infectieux ou inflammatoires [389]. Contrairement aux méthodes basées sur l'enrichissement de fonctions, celle-ci contient des modules plus précis à propos des fonctions immunitaires par rapport aux méthodes utilisant les bases de données GSEA. Néanmoins, cette méthode a pour inconvénient d'avoir été développée pour l'homme et l'analyse de données obtenues à partir d'échantillons de sang et non pas de tissu. Néanmoins, l'utilisation de ces méthodes en parallèle peut être

avantageuse, notamment dans le cas où une fonction est identifiée de façon récurrente avec plusieurs de ces méthodes [409].

Concernant l'analyse ciblée de la sécrétion de médiateurs immunitaires comme les cytokines et les chimiokines, la méthode fondée sur la technologie Luminex nous a permis l'étude de 15 molécules en parallèle sur un même échantillon de faible volume. Cette méthode peut être couplée à l'étude transcriptomique permettant ainsi d'avoir des informations aux deux niveaux, transcriptome et protéome sécrété. Tout comme la transcriptomique, il est possible d'étudier de nombreux mécanismes immunitaires via le dosage de cytokines spécifiques. La trousse utilisée dans notre étude contient 15 cytokines permettant l'étude de l'inflammation, du recrutement cellulaire et de la réponse lymphocytaire. De plus, après adaptation des conditions de dosage qui ont généralement été conçues pour le sang, il est possible de doser ces produits dans différents types d'échantillons tels que le lait, ou des broyats de tissu mammaire ou pulmonaire. Cette technologie peut donc être appliquée aussi bien aux études systémiques que locales.

L'ensemble de ces méthodes requièrent des compétences en bio-informatique et en bio-statistiques pour être utilisées et développées. L'ensemble de ces méthodes est très utilisé en santé humaine, mais elles ne l'ont été que trop peu dans les études sur les espèces de rente, même si elles se développent. La plupart des fonctions d'analyse sont disponibles dans l'environnement R qui demande un certain apprentissage, mais qui est très flexible outre le fait d'être en libre accès par rapport à des logiciels commerciaux souvent très accessibles pour les néophytes mais dont les fonctionnalités sont beaucoup plus réduites.

Avec le développement des bases de données grâce aux méthodes dites à haut débit tels que les technologies omiques, l'analyse des fonctions biologiques est de plus en plus précise. Dans les espèces humaines et murines, les processus biologiques et les voies de signalisation sont très bien décrits, alors que cela n'est pas le cas dans les espèces domestiques malgré des différences du point de vue de la composition ou des fonctions immunitaires. Néanmoins, de nombreux mécanismes immunitaires sont conservés entre les mammifères voire entre l'ensemble des espèces vertébrés, facilitant leur analyse au niveau transcriptomique [410].

3.3. Intégration des données multidimensionnelles

L'étude de données de nature très diverse, telles que les données cliniques, transcriptomiques ou protéomiques permet d'investiguer de façon très spécifique des mécanismes qui participent à une fonction biologique, comme l'élimination d'un agent pathogène et le rétablissement de

l'homéostasie lorsque ce dernier a été éliminé. Grâce à l'utilisation de méthodes de corrélation telle que la suite WGCNA, il est possible d'identifier des fonctions corrélées avec des données externes comme un score clinique, lésionnel ou une performance zootechnique dans le but d'identifier des déterminismes communs, voire même de suggérer un lien causal qui pourra être démontré à l'aide de données complémentaires. Dans notre étude, nous avons établi que les fonctions liées à l'interféron en lien avec la vaccination étaient fortement corrélées à l'augmentation du score clinique mammaire. Ce lien n'a jamais été décrit auparavant et il n'est pas facile d'en établir l'origine ou les mécanismes. Cela suggère qu'une trop forte réponse liée aux interférons (dont la nature reste à déterminer) pourrait être délétère pour l'animal dans ce cadre. L'utilisation de ces méthodes permet donc d'obtenir de nouvelles informations qui n'étaient pas disponibles auparavant et ainsi ouvrir de nouvelles pistes d'investigation pour le développement de mesures préventives ou curatives plus efficaces.

Bien que ces méthodes possèdent de nombreux avantages, elles ont aussi certains défauts tels que la complexité de l'analyse informatique. Elles ne permettent pas non plus de s'affranchir de l'interprétation biologique qui reste souvent au stade d'hypothèse et demande des travaux complémentaires par la réalisation d'essais expérimentaux. Même si ces approches permettent des investigations sans *a priori* et en profondeur, elles ne remplacent pas les essais biologiques qui restent nécessaires pour confirmer ou infirmer telle ou telle hypothèse. Les résultats fournis sont quelquefois difficiles à déchiffrer et à interpréter.

Avec l'avènement des méthodes à haut débit permettant l'analyse des différents compartiments et systèmes biologiques, la discipline pan-omique a émergé. Elle consiste à étudier simultanément les différentes échelles du vivant et en concomitance des données transcriptomiques, génomiques et protéomiques, mais aussi génotypiques et phénotypiques. Elle s'avère très utile dans les domaines de l'immunologie et de l'infectiologie [411]. Elle est aujourd'hui très utilisée dans le cadre de l'étude du microbiote et de son effet dans différents contextes immunitaires et métaboliques [411], [412]. Cette méthodologie a déjà été utilisée dans le cadre d'études vaccinales chez l'homme [413]. Avec le développement du *machine learning*, de nombreuses méthodes permettant ces analyses ont vu le jour tels que *Mixomics* [400] ou encore *omicade4* [414]. Néanmoins, cette méthode a au moins deux inconvénients. Dans certains cas, il est difficile d'interpréter d'un point de vue biologique les résultats obtenus ; Cette critique a été formulée par les termes de « Low imput, high throughout no output biology » [415] par Sydney Brenner. Le second concerne les compétences et le matériel nécessaire. En effet, pour ce type d'analyse, des connaissances en

bioinformatiques/biostatistiques mais aussi de très bonnes bases en immunologie dans notre exemple sont nécessaires. De plus, la quantité de matériel biologique et la qualité nécessaire sont souvent élevées notamment en transcriptomique et en protéomique, ce qui peut limiter l'utilisation d'échantillons collectés dans des conditions dites de terrain. La puissance informatique peut aussi se révéler être une limite car ces méthodes utilisent des algorithmes nécessitant une mémoire de travail relativement grande, des processeurs puissants et des réseaux de bon débit. L'ensemble de ces inconvénients peut limiter l'utilisation des méthodes multi-omiques malgré leur fort potentiel [416].

3.4. Réduction des données expérimentales

Les méthodes expérimentales actuelles sont de plus en plus souvent basées sur l'utilisation de données de grandes dimensions. En effet, de nombreuses données peuvent être collectées et pour les animaux d'élevage, certaines le sont même de façon automatisée à l'aide d'équipements installés directement sur les animaux (podomètre, bolus ...) ou à leur proximité (caméra, capteur GPS, ...). Des mesures en ligne tels que les CCS permettant de qualifier la qualité du lait et l'absence d'infection peuvent être enregistrées chaque fois que la vache se présente dans le robot de traite par exemple. La quantité d'informations issues de l'animal est de plus en plus volumineuse et les méthodes d'analyses de ces données doivent évoluer. En effet, comme nous avons pu le voir dans l'étude sur la préparation des jeunes bovins à l'engrais, il est possible d'obtenir de nombreuses informations de natures très différentes (immunité, alimentation, croissance). L'analyse variable par variable était la règle jusqu'alors pour l'étude de ces données, mais l'étude intégrative permet d'aller plus loin pour tirer des informations utiles et au final, exploitables.

L'analyse variable par variable n'est pas une solution adéquate à de nombreuses situations car les variables sont souvent liées entre elles. Les méthodes d'analyse en composantes principales constituent une méthodologie d'étude qui résume l'information sur les animaux par exemple, sur quelques dimensions permettant de classifier les animaux comme cela a été fait dans l'une de nos études. En effet l'analyse factorielle multiple permet de catégoriser les variables d'études et d'attribuer une même contribution pour chacun des jeux de variables.

Ces méthodes dites multivariées ont pour avantage de pouvoir prendre en compte le lien entre les variables. En effet, dans notre étude, nous avons pu voir que les performances zootechniques pouvaient être fortement liées à la concentration plasmatique du cuivre de l'animal. Or dans la littérature [417], le lien entre performances zootechniques et concentration du cuivre est plutôt négatif. Nous avons aussi pu voir que les liens entre les données de différents compartiments

étaient présents et qu'il est nécessaire de les prendre en compte, ce qui est possible dans les analyses multivariées.

L'étude de l'immunité humorale via le dosage des anticorps vis-à-vis du ou des pathogènes étudiés est possible grâce aux diverses trousse ELISA qui ont été développées pour leur détection et quantification. Ce sont des informations facilement acquérables mais elles ne représentent pas l'ensemble de la réponse immunitaire. Jusqu'à peu, il n'y avait pas de méthode permettant l'analyse de la réponse cellulaire. La trousse développée au laboratoire a permis de combler ce manque. De plus, ce profilage peut facilement être couplé avec une méthode de stimulation du sang dans un système de tube-seringue (monovette) comme cela a été mis en place au laboratoire. Cette association de méthodes peut permettre d'étudier très facilement la réponse immunitaire dans le sang, après stimulation avec différentes catégories de ligands ou des antigènes plus ou moins complexes (microbes, protéines recombinantes, ...), pour ensuite doser les cytokines produites après la stimulation. La mesure des cytokines sécrétées permet d'approcher mieux leur fonction biologique par rapport à l'étude du transcriptome seul.

3.5. Et après ?!

L'étude sur la vaccination vis-à-vis des mammites causées par *E.coli* a permis de déterminer approximativement les mécanismes locaux permettant une protection. Une étude plus poussée sur l'étude locale serait une suite logique pour approfondir les connaissances sur ces mécanismes protecteurs en lien avec la voie IL-17. Une étude globale de la réponse immunitaire locale à l'aide d'une analyse en single cell RNA-seq couplé à des dosages cytokiniques serait bénéfique. Cette seconde étude ainsi que les résultats déjà obtenus pourrait servir de base au développement de nouvelles méthodes de vaccination vis-à-vis des mammites à *E .coli*, voire vis-à-vis des autres agents pathogènes mammaires.

Dans le cadre de l'étude sur les BPI, une conduite d'élevage exemplaire semble plus efficace qu'une préparation faite en amont de l'entrée en atelier d'engraissement. Néanmoins, les méthodes actuellement utilisées pour la prédiction des maladies ou de leur évolution sont encore à améliorer. Le recours à des méthodes multidimensionnelles comme celles vues dans l'étude 2 pourraient être très utiles dans ce contexte. En effet, elles peuvent avoir un intérêt dans la prédiction via l'utilisation de données évolutives tels que les performances zootechniques, les scores cliniques ou encore la réponse immunitaire vis-à-vis du pathogène étudié.

4. Bibliographie

- [1] Y. H. Schukken, D. J. Wilson, F. Welcome, L. Garrison-Tikofsky, et R. N. Gonzalez, « Monitoring udder health and milk quality using somatic cell counts », *Vet. Res.*, vol. 34, n° 5, p. 579-596, oct. 2003, doi: 10.1051/vetres:2003028.
- [2] M. Green, J. Huxley, et A. Bradley, « A rational approach to dry cow therapy : 1. Udder health priorities during the dry period », *In Pract.*, vol. 24, nov. 2002, doi: 10.1136/inpract.24.10.582.
- [3] C. Burvenich, V. Van Merris, J. Mehrzad, A. Diez-Fraile, et L. Duchateau, « Severity of *E. coli* mastitis is mainly determined by cow factors », *Vet. Res.*, vol. 34, n° 5, p. 521-564, sept. 2003, doi: 10.1051/vetres:2003023.
- [4] J. Hogan et K. Larry Smith, « Coliform mastitis », *Vet. Res.*, vol. 34, n° 5, p. 507-519, oct. 2003, doi: 10.1051/vetres:2003022.
- [5] A. J. Bradley et M. J. Green, « A study of the incidence and significance of intramammary enterobacterial infections acquired during the dry period », *J. Dairy Sci.*, vol. 83, n° 9, p. 1957-1965, sept. 2000, doi: 10.3168/jds.S0022-0302(00)75072-7.
- [6] A. Porcherie *et al.*, « IL-17A Is an Important Effector of the Immune Response of the Mammary Gland to *Escherichia coli* Infection », *J. Immunol.*, vol. 196, n° 2, p. 803-812, janv. 2016, doi: 10.4049/jimmunol.1500705.
- [7] Y. Schukken *et al.*, « The “other” gram-negative bacteria in mastitis: *Klebsiella*, *serratia*, and more », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 28, n° 2, p. 239-256, juill. 2012, doi: 10.1016/j.cvfa.2012.04.001.
- [8] M. G. Ribeiro *et al.*, « Peracute bovine mastitis caused by *Klebsiella pneumoniae* », *Arq. Bras. Med. Veterinária E Zootec.*, vol. 60, n° 2, p. 485-488, avr. 2008, doi: 10.1590/S0102-09352008000200031.
- [9] J. A. Hertl, Y. H. Schukken, F. L. Welcome, L. W. Tauer, et Y. T. Gröhn, « Pathogen-specific effects on milk yield in repeated clinical mastitis episodes in Holstein dairy cows », *J. Dairy Sci.*, vol. 97, n° 3, p. 1465-1480, mars 2014, doi: 10.3168/jds.2013-7266.
- [10] R. J. Erskine, S. Wagner, et F. J. DeGraves, « Mastitis therapy and pharmacology », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 19, n° 1, p. 109-138, vi, mars 2003, doi: 10.1016/s0749-0720(02)00067-1.
- [11] I. K. Mullarky, C. Su, N. Frieze, Y. H. Park, et L. M. Sordillo, « *Staphylococcus aureus* agr genotypes with enterotoxin production capabilities can resist neutrophil bactericidal activity », *Infect. Immun.*, vol. 69, n° 1, p. 45-51, janv. 2001, doi: 10.1128/IAI.69.1.45-51.2001.
- [12] P. Trinidad, S. C. Nickerson, et T. K. Alley, « Prevalence of intramammary infection and teat canal colonization in unbred and primigravid dairy heifers », *J. Dairy Sci.*, vol. 73, n° 1, p. 107-114, janv. 1990, doi: 10.3168/jds.S0022-0302(90)78652-3.
- [13] B.-A. Tenhagen, I. Hansen, A. Reinecke, et W. Heuwieser, « Prevalence of pathogens in milk samples of dairy cows with clinical mastitis and in heifers at first parturition », *J. Dairy Res.*, vol. 76, n° 2, p. 179-187, mai 2009, doi: 10.1017/S0022029908003786.
- [14] R. N. Zadoks, B. E. Gillespie, H. W. Barkema, O. C. Sampimon, S. P. Oliver, et Y. H. Schukken, « Clinical, epidemiological and molecular characteristics of *Streptococcus uberis* infections in dairy herds », *Epidemiol. Infect.*, vol. 130, n° 2, p. 335-349, avr. 2003, doi: 10.1017/s0950268802008221.
- [15] « Molecular Epidemiology of *Streptococcus uberis* Clinical Mastitis in Dairy Herds: Strain Heterogeneity and Transmission - PubMed ». <https://pubmed.ncbi.nlm.nih.gov/26491180/> (consulté le mai 07, 2021).

- [16] M. H. Milne *et al.*, « Treatment of persistent intramammary infections with *Streptococcus uberis* in dairy cows », *Vet. Rec.*, vol. 157, n° 9, p. 245-250, août 2005, doi: 10.1136/vr.157.9.245.
- [17] R. N. Zadoks *et al.*, « Cow- and quarter-level risk factors for *Streptococcus uberis* and *Staphylococcus aureus* mastitis », *J. Dairy Sci.*, vol. 84, n° 12, p. 2649-2663, déc. 2001, doi: 10.3168/jds.s0022-0302(01)74719-4.
- [18] G. Martinez, J. Harel, R. Higgins, S. Lacouture, D. Daignault, et M. Gottschalk, « Characterization of *Streptococcus agalactiae* isolates of bovine and human origin by randomly amplified polymorphic DNA analysis », *J. Clin. Microbiol.*, vol. 38, n° 1, p. 71-78, janv. 2000.
- [19] A. Khan, « Basic facts of mastitis in dairy animals: A review », *Pak. Vet. J.*, vol. 26, oct. 2006.
- [20] L. K. Fox et J. M. Gay, « Contagious mastitis », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 9, n° 3, p. 475-487, nov. 1993, doi: 10.1016/s0749-0720(15)30615-0.
- [21] A. Wilkinson, *16 National Mastitis Council Regional Meeting Proceedings (2003) TO SEAL OR NOT TO SEAL: INTERNAL TEAT SEALANT STRATEGIES*.
- [22] V. Joshi, V. K. Gupta, U. Dimri, R. S. K. Mandal, et D. K. Sharma, « Evaluating serum lipid profile in bacterial bovine respiratory disease (BRD) affected calves », *Intas Polivet*, vol. 16, n° 2, p. 187-189, juill. 2015.
- [23] G. C. Duff et M. L. Galyean, « Board-invited review: recent advances in management of highly stressed, newly received feedlot cattle », *J. Anim. Sci.*, vol. 85, n° 3, p. 823-840, mars 2007, doi: 10.2527/jas.2006-501.
- [24] A. Prohl, C. Ostermann, M. Lohr, et P. Reinhold, « The bovine lung in biomedical research: visually guided bronchoscopy, intrabronchial inoculation and in vivo sampling techniques », *J. Vis. Exp. JoVE*, n° 89, juill. 2014, doi: 10.3791/51557.
- [25] M. de la Torre, D. Gonzalez-Rivas, R. Fernández-Prado, M. Delgado, E. M. Fieira, et A. Centeno, « Uniportal video-assisted thoracoscopic lobectomy in the animal model », *J. Thorac. Dis.*, vol. 6, n° Suppl 6, p. S656-S659, oct. 2014, doi: 10.3978/j.issn.2072-1439.2014.10.03.
- [26] P. Cusack, N. McMeniman, et I. Lean, « The medicine and epidemiology of bovine respiratory disease in feedlots », *Aust. Vet. J.*, vol. 81, p. 480-7, sept. 2003, doi: 10.1111/j.1751-0813.2003.tb13367.x.
- [27] M. D. Welsh, B. M. Adair, et J. C. Foster, « Effect of BVD virus infection on alveolar macrophage functions », *Vet. Immunol. Immunopathol.*, vol. 46, n° 3, p. 195-210, juin 1995, doi: 10.1016/0165-2427(94)05366-Z.
- [28] B. Muylkens, J. Thiry, P. Kirten, F. Schynts, et E. Thiry, « Bovine herpesvirus 1 infection and infectious bovine rhinotracheitis », *Vet. Res.*, vol. 38, n° 2, p. 181-209, 2007, doi: 10.1051/vetres:2006059.
- [29] D. W. V. Banie Penzhorn, *Infectious Diseases of Livestock*. 2005.
- [30] Bradford P. Smith DVM, *Large Animal Internal Medicine*, 3rd Edition. 2001.
- [31] R. W. Fulton *et al.*, « Bovine viral diarrhoea viral infections in feeder calves with respiratory disease: interactions with *Pasteurella* spp., parainfluenza-3 virus, and bovine respiratory syncytial virus », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 64, n° 3, p. 151-159, juill. 2000.
- [32] R. W. Fulton *et al.*, « Transmission of bovine viral diarrhoea virus 1b to susceptible and vaccinated calves by exposure to persistently infected calves », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 69, n° 3, p. 161-169, juill. 2005.
- [33] R. W. Fulton *et al.*, « Bovine viral diarrhoea virus (BVDV) 1b: predominant BVDV subtype in calves with respiratory disease », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 66, n° 3, p. 181-190, juill. 2002.

- [34] J. R. Campbell, « Effect of bovine viral diarrhea virus in the feedlot », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 20, n° 1, p. 39-50, mars 2004, doi: 10.1016/j.cvfa.2003.11.003.
- [35] J. F. Evermann et J. F. Ridpath, « Clinical and epidemiologic observations of bovine viral diarrhea virus in the northwestern United States », *Vet. Microbiol.*, vol. 89, n° 2-3, p. 129-139, oct. 2002, doi: 10.1016/s0378-1135(02)00178-5.
- [36] R. W. Fulton *et al.*, « Evaluation of diagnostic tests used for detection of bovine viral diarrhea virus and prevalence of subtypes 1a, 1b, and 2a in persistently infected cattle entering a feedlot », *J. Am. Vet. Med. Assoc.*, vol. 228, n° 4, p. 578-584, févr. 2006, doi: 10.2460/javma.228.4.578.
- [37] B.W.J. Mahy, *Encyclopedia of Virology 3rd edition*. 2008.
- [38] E. L. Shroyer et B. C. Easterday, « Studies on the pathogenesis of infectious bovine rhinotracheitis following aerosol exposure », *Cornell Vet.*, vol. 58, n° 3, p. 442-461, juill. 1968.
- [39] D. E. Anderson et D. M. Rings, *Current veterinary therapy*. St. Louis, Mo.; [Edinburgh: Saunders Elsevier, 2009.
- [40] Port Huron, *Compendium of veterinary products : CVP.*, 9th éd. 2006.
- [41] L. N. Potgieter, M. D. McCracken, F. M. Hopkins, R. D. Walker, et J. S. Guy, « Experimental production of bovine respiratory tract disease with bovine viral diarrhea virus », *Am. J. Vet. Res.*, vol. 45, n° 8, p. 1582-1585, août 1984.
- [42] B. W. Brodersen et C. L. Kelling, « Alteration of leukocyte populations in calves concurrently infected with bovine respiratory syncytial virus and bovine viral diarrhea virus », *Viral Immunol.*, vol. 12, n° 4, p. 323-334, 1999, doi: 10.1089/vim.1999.12.323.
- [43] R. W. Fulton *et al.*, « Evaluation of health status of calves and the impact on feedlot performance: assessment of a retained ownership program for postweaning calves », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 66, n° 3, p. 173-180, juill. 2002.
- [44] R. J. Schroeder et M. D. Moys, « An acute upper respiratory infection of dairy cattle », *J. Am. Vet. Med. Assoc.*, vol. 125, n° 933, p. 471-472, déc. 1954.
- [45] S. Kapil et R. J. Basaraba, « Infectious bovine rhinotracheitis, parainfluenza-3, and respiratory coronavirus », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 13, n° 3, p. 455-469, nov. 1997, doi: 10.1016/s0749-0720(15)30308-x.
- [46] P. Aubry, L. D. Warnick, C. L. Guard, B. W. Hill, et M. F. Witt, « Health and performance of young dairy calves vaccinated with a modified-live Mannheimia haemolytica and Pasteurella multocida vaccine », *J. Am. Vet. Med. Assoc.*, vol. 219, n° 12, p. 1739-1742, déc. 2001, doi: 10.2460/javma.2001.219.1739.
- [47] S. K. Highlander, « Molecular genetic analysis of virulence in Mannheimia (pasteurella) haemolytica », *Front. Biosci. J. Virtual Libr.*, vol. 6, p. D1128-1150, sept. 2001, doi: 10.2741/highland.
- [48] M. Harper, J. D. Boyce, et B. Adler, « Pasteurella multocida pathogenesis: 125 years after Pasteur », *FEMS Microbiol. Lett.*, vol. 265, n° 1, p. 1-10, déc. 2006, doi: 10.1111/j.1574-6968.2006.00442.x.
- [49] S. M. Dabo, J. D. Taylor, et A. W. Confer, « Pasteurella multocida and bovine respiratory disease », *Anim. Health Res. Rev.*, vol. 8, n° 2, p. 129-150, déc. 2007, doi: 10.1017/S1466252307001399.
- [50] O. Angen *et al.*, « Respiratory disease in calves: microbiological investigations on trans-tracheally aspirated bronchoalveolar fluid and acute phase protein response », *Vet. Microbiol.*, vol. 137, n° 1-2, p. 165-171, mai 2009, doi: 10.1016/j.vetmic.2008.12.024.
- [51] D. C. DeRosa, G. D. Mechor, J. J. Staats, M. M. Chengappa, et T. R. Shryock, « Comparison of Pasteurella spp. simultaneously isolated from nasal and transtracheal swabs from cattle with clinical signs of bovine respiratory disease », *J. Clin. Microbiol.*, vol. 38, n° 1, p. 327-332, janv. 2000.

- [52] C. W. Booker, P. T. Guichon, G. K. Jim, O. C. Schunicht, R. J. Harland, et P. S. Morley, « Seroepidemiology of undifferentiated fever in feedlot calves in western Canada », *Can. Vet. J. Rev. Veterinaire Can.*, vol. 40, n° 1, p. 40-48, janv. 1999.
- [53] J. L. Caswell et M. Archambault, « Mycoplasma bovis pneumonia in cattle », *Anim. Health Res. Rev.*, vol. 8, n° 2, p. 161-186, déc. 2007, doi: 10.1017/S1466252307001351.
- [54] W. J. Byrne, R. McCormack, N. Brice, J. Egan, B. Markey, et H. J. Ball, « Isolation of Mycoplasma bovis from bovine clinical samples in the Republic of Ireland », *Vet. Rec.*, vol. 148, n° 11, p. 331-333, mars 2001, doi: 10.1136/vr.148.11.331.
- [55] V. Tancin, « THE EFFECT OF SOMATIC CELL ON MILK YIELD AND MILK FLOW AT QUARTER LEVEL », *Vet. Ir Zootech.*, vol. 66, p. 69-72, juin 2014.
- [56] M. Cobirka, V. Tancin, et P. Slama, « Epidemiology and Classification of Mastitis », *Anim. Open Access J. MDPI*, vol. 10, n° 12, nov. 2020, doi: 10.3390/ani10122212.
- [57] J. B. C. Fernandes, L. G. Zanardo, N. N. Galvão, I. A. Carvalho, L. A. Nero, et M. A. S. Moreira, « Escherichia coli from clinical mastitis: serotypes and virulence factors », *J. Vet. Diagn. Investig. Off. Publ. Am. Assoc. Vet. Lab. Diagn. Inc.*, vol. 23, n° 6, p. 1146-1152, nov. 2011, doi: 10.1177/1040638711425581.
- [58] M. G. Netea, C. van der Graaf, J. W. M. Van der Meer, et B. J. Kullberg, « Toll-like receptors and the host defense against microbial pathogens: bringing specificity to the innate-immune system », *J. Leukoc. Biol.*, vol. 75, n° 5, p. 749-755, mai 2004, doi: 10.1189/jlb.1103543.
- [59] P. Rainard, « The complement in milk and defense of the bovine mammary gland against infections », *Vet. Res.*, vol. 34, n° 5, p. 647-670, oct. 2003, doi: 10.1051/vetres:2003025.
- [60] D. D. Bannerman, M. J. Paape, J.-W. Lee, X. Zhao, J. C. Hope, et P. Rainard, « Escherichia coli and Staphylococcus aureus elicit differential innate immune responses following intramammary infection », *Clin. Diagn. Lab. Immunol.*, vol. 11, n° 3, p. 463-472, mai 2004, doi: 10.1128/CDLI.11.3.463-472.2004.
- [61] D. D. Bannerman, M. J. Paape, J. P. Goff, K. Kimura, J. D. Lippolis, et J. C. Hope, « Innate immune response to intramammary infection with Serratia marcescens and Streptococcus uberis », *Vet. Res.*, vol. 35, n° 6, p. 681-700, déc. 2004, doi: 10.1051/vetres:2004040.
- [62] C. Riollet, P. Rainard, et B. Poutrel, « Differential induction of complement fragment C5a and inflammatory cytokines during intramammary infections with Escherichia coli and Staphylococcus aureus », *Clin. Diagn. Lab. Immunol.*, vol. 7, n° 2, p. 161-167, mars 2000, doi: 10.1128/cdli.7.2.161-167.2000.
- [63] D. E. Shuster, M. E. Kehrli, P. Rainard, et M. Paape, « Complement fragment C5a and inflammatory cytokines in neutrophil recruitment during intramammary infection with Escherichia coli », *Infect. Immun.*, vol. 65, n° 8, p. 3286-3292, août 1997, doi: 10.1128/IAI.65.8.3286-3292.1997.
- [64] D. Legrand, E. Ellass, A. Pierce, et J. Mazurier, « Lactoferrin and host defence: an overview of its immuno-modulating and anti-inflammatory properties », *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.*, vol. 17, n° 3, p. 225-229, juin 2004, doi: 10.1023/b:biom.0000027696.48707.42.
- [65] P. Rainard, « Bacteriostatic activity of bovine milk lactoferrin against mastitic bacteria », *Vet. Microbiol.*, vol. 11, n° 4, p. 387-392, avr. 1986, doi: 10.1016/0378-1135(86)90068-4.
- [66] P. Rainard, « Binding of bovine lactoferrin to Streptococcus agalactiae », *FEMS Microbiol. Lett.*, vol. 77, n° 1-3, p. 235-239, nov. 1992, doi: 10.1016/0378-1097(92)90162-h.

- [67] N. Orsi, « The antimicrobial activity of lactoferrin: current status and perspectives », *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.*, vol. 17, n° 3, p. 189-196, juin 2004, doi: 10.1023/b:biom.0000027691.86757.e2.
- [68] L. Sanchez *et al.*, « Synthesis of lactoferrin and transport of transferrin in the lactating mammary gland of sheep », *J. Dairy Sci.*, vol. 75, n° 5, p. 1257-1262, mai 1992, doi: 10.3168/jds.S0022-0302(92)77875-8.
- [69] P. Rainard, « Experimental mastitis with *Escherichia coli*: kinetics of bacteriostatic and bactericidal activities », *Ann. Rech. Veterinaires Ann. Vet. Res.*, vol. 14, n° 1, p. 1-11, 1983.
- [70] J. T. Hancock *et al.*, « Antimicrobial properties of milk: dependence on presence of xanthine oxidase and nitrite », *Antimicrob. Agents Chemother.*, vol. 46, n° 10, p. 3308-3310, oct. 2002, doi: 10.1128/aac.46.10.3308-3310.2002.
- [71] P. D. Eckersall *et al.*, « Acute phase proteins in serum and milk from dairy cows with clinical mastitis », *Vet. Rec.*, vol. 148, n° 2, p. 35-41, janv. 2001, doi: 10.1136/vr.148.2.35.
- [72] U. Grönlund, C. Hallén Sandgren, et K. Persson Waller, « Haptoglobin and serum amyloid A in milk from dairy cows with chronic sub-clinical mastitis », *Vet. Res.*, vol. 36, n° 2, p. 191-198, avr. 2005, doi: 10.1051/vetres:2004063.
- [73] N. Y. Yount *et al.*, « Cloning and expression of bovine neutrophil beta-defensins. Biosynthetic profile during neutrophilic maturation and localization of mature peptide to novel cytoplasmic dense granules », *J. Biol. Chem.*, vol. 274, n° 37, p. 26249-26258, sept. 1999, doi: 10.1074/jbc.274.37.26249.
- [74] M. E. Selsted *et al.*, « Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils », *J. Biol. Chem.*, vol. 268, n° 9, p. 6641-6648, mars 1993.
- [75] D. M. E. Bowdish, D. J. Davidson, et R. E. W. Hancock, « A re-evaluation of the role of host defence peptides in mammalian immunity », *Curr. Protein Pept. Sci.*, vol. 6, n° 1, p. 35-51, févr. 2005, doi: 10.2174/1389203053027494.
- [76] M. Dürr et A. Peschel, « Chemokines meet defensins: the merging concepts of chemoattractants and antimicrobial peptides in host defense », *Infect. Immun.*, vol. 70, n° 12, p. 6515-6517, déc. 2002, doi: 10.1128/iai.70.12.6515-6517.2002.
- [77] S. van Wetering, G. S. Tjabringa, et P. S. Hiemstra, « Interactions between neutrophil-derived antimicrobial peptides and airway epithelial cells », *J. Leukoc. Biol.*, vol. 77, n° 4, p. 444-450, avr. 2005, doi: 10.1189/jlb.0604367.
- [78] D. L. Jensen et R. J. Eberhart, « Total and differential cell counts in secretions of the nonlactating bovine mammary gland », *Am. J. Vet. Res.*, vol. 42, n° 5, p. 743-747, mai 1981.
- [79] J. S. McDonald et A. J. Anderson, « Total and differential somatic cell counts in secretions from noninfected bovine mammary glands: the peripartum period », *Am. J. Vet. Res.*, vol. 42, n° 8, p. 1366-1368, août 1981.
- [80] L. M. Sordillo, S. C. Nickerson, R. M. Akers, et S. P. Oliver, « Secretion composition during bovine mammary involution and the relationship with mastitis », *Int. J. Biochem.*, vol. 19, n° 12, p. 1165-1172, 1987, doi: 10.1016/0020-711x(87)90098-x.
- [81] N. A. Mullan, E. A. Carter, et K. A. Nguyen, « Phagocytic and bactericidal properties of bovine macrophages from non-lactating mammary glands », *Res. Vet. Sci.*, vol. 38, n° 2, p. 160-166, mars 1985.
- [82] J. MacMicking, Q. W. Xie, et C. Nathan, « Nitric oxide and macrophage function », *Annu. Rev. Immunol.*, vol. 15, p. 323-350, 1997, doi: 10.1146/annurev.immunol.15.1.323.
- [83] J. W. Blum *et al.*, « Tumor necrosis factor-alpha and nitrite/nitrate responses during acute mastitis induced by *Escherichia coli* infection and endotoxin in dairy cows », *Domest. Anim. Endocrinol.*, vol. 19, n° 4, p. 223-235, nov. 2000, doi: 10.1016/s0739-7240(00)00079-5.

- [84] L. Bouchard, S. Blais, C. Desrosiers, X. Zhao, et P. Lacasse, « Nitric oxide production during endotoxin-induced mastitis in the cow », *J. Dairy Sci.*, vol. 82, n° 12, p. 2574-2581, déc. 1999, doi: 10.3168/jds.S0022-0302(99)75512-8.
- [85] K.-I. Komine *et al.*, « Induction of nitric oxide production mediated by tumor necrosis factor alpha on staphylococcal enterotoxin C-stimulated bovine mammary gland cells », *Clin. Diagn. Lab. Immunol.*, vol. 11, n° 1, p. 203-210, janv. 2004, doi: 10.1128/cdli.11.1.203-210.2004.
- [86] I. Politis, B. W. McBride, J. H. Burton, X. Zhao, et J. D. Turner, « Secretion of interleukin-1 by bovine milk macrophages », *Am. J. Vet. Res.*, vol. 52, n° 6, p. 858-862, juin 1991.
- [87] P. Rainard et C. Riollet, « Innate immunity of the bovine mammary gland », *Vet. Res.*, vol. 37, n° 3, p. 369-400, mai 2006, doi: 10.1051/vetres:2006007.
- [88] L. M. Sordillo, M. Campos, et L. A. Babiuk, « Antibacterial activity of bovine mammary gland lymphocytes following treatment with interleukin-2 », *J. Dairy Sci.*, vol. 74, n° 10, p. 3370-3375, oct. 1991, doi: 10.3168/jds.S0022-0302(91)78526-3.
- [89] O. Wellnitz et D. E. Kerr, « Cryopreserved bovine mammary cells to model epithelial response to infection », *Vet. Immunol. Immunopathol.*, vol. 101, n° 3-4, p. 191-202, oct. 2004, doi: 10.1016/j.vetimm.2004.04.019.
- [90] H. Okada *et al.*, « Detection of interleukin-1 and interleukin-6 on cryopreserved bovine mammary epithelial cells in vitro », *J. Vet. Med. Sci.*, vol. 59, n° 7, p. 503-507, juill. 1997, doi: 10.1292/jvms.59.503.
- [91] R. J. Harmon et C. W. Heald, « Migration of polymorphonuclear leukocytes into the bovine mammary gland during experimentally induced *Staphylococcus aureus* mastitis », *Am. J. Vet. Res.*, vol. 43, n° 6, p. 992-998, juin 1982.
- [92] M. Rambeaud, R. A. Almeida, G. M. Pighetti, et S. P. Oliver, « Dynamics of leukocytes and cytokines during experimentally induced *Streptococcus uberis* mastitis », *Vet. Immunol. Immunopathol.*, vol. 96, n° 3-4, p. 193-205, déc. 2003, doi: 10.1016/j.vetimm.2003.08.008.
- [93] B. C. Taylor, J. D. Dellinger, J. S. Cullor, et J. L. Stott, « Bovine milk lymphocytes display the phenotype of memory T cells and are predominantly CD8+ », *Cell. Immunol.*, vol. 156, n° 1, p. 245-253, juin 1994, doi: 10.1006/cimm.1994.1169.
- [94] Y. Zhao *et al.*, « Shifted T Helper Cell Polarization in a Murine *Staphylococcus aureus* Mastitis Model », *PLOS ONE*, vol. 10, n° 7, p. e0134797, juill. 2015, doi: 10.1371/journal.pone.0134797.
- [95] S. Bougarn, P. Cunha, F. B. Gilbert, A. Harmache, G. Foucras, et P. Rainard, « Staphylococcal-associated molecular patterns enhance expression of immune defense genes induced by IL-17 in mammary epithelial cells », *Cytokine*, vol. 56, n° 3, p. 749-759, déc. 2011, doi: 10.1016/j.cyto.2011.09.020.
- [96] A. Lopez *et al.*, « The pulmonary clearance of *Pasteurella haemolytica* in calves infected with bovine virus diarrhea or *Mycoplasma bovis* », *Can. J. Comp. Med. Rev. Can. Med. Comp.*, vol. 46, n° 3, p. 302-306, juill. 1982.
- [97] M. R. Ackermann, R. Derscheid, et J. A. Roth, « Innate immunology of bovine respiratory disease », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 26, n° 2, p. 215-228, juill. 2010, doi: 10.1016/j.cvfa.2010.03.001.
- [98] J. L. Caswell, « Failure of respiratory defenses in the pathogenesis of bacterial pneumonia of cattle », *Vet. Pathol.*, vol. 51, n° 2, p. 393-409, mars 2014, doi: 10.1177/0300985813502821.
- [99] A. U. N'jai, J. Rivera, D. N. Atapattu, K. Owusu-Ofori, et C. J. Czuprynski, « Gene expression profiling of bovine bronchial epithelial cells exposed in vitro to bovine

- herpesvirus 1 and Mannheimia haemolytica », *Vet. Immunol. Immunopathol.*, vol. 155, n° 3, p. 182-189, sept. 2013, doi: 10.1016/j.vetimm.2013.06.012.
- [100] M. S. Marin, S. Quintana, C. Faverín, M. R. Leunda, A. C. Odeón, et S. E. Pérez, « Toll-like receptor activation and expression in bovine alpha-herpesvirus infections », *Res. Vet. Sci.*, vol. 96, n° 1, p. 196-203, févr. 2014, doi: 10.1016/j.rvsc.2013.11.015.
- [101] K. Taha-Abdelaziz *et al.*, « Bactericidal activity of tracheal antimicrobial peptide against respiratory pathogens of cattle », *Vet. Immunol. Immunopathol.*, vol. 152, n° 3-4, p. 289-294, avr. 2013, doi: 10.1016/j.vetimm.2012.12.016.
- [102] K. A. Brogden, G. Nordholm, et M. Ackermann, « Antimicrobial activity of cathelicidins BMAP28, SMAP28, SMAP29, and PMAP23 against *Pasteurella multocida* is more broad-spectrum than host species specific », *Vet. Microbiol.*, vol. 119, n° 1, p. 76-81, janv. 2007, doi: 10.1016/j.vetmic.2006.08.005.
- [103] M. Al-Haddawi, G. B. Mitchell, M. E. Clark, R. D. Wood, et J. L. Caswell, « Impairment of innate immune responses of airway epithelium by infection with bovine viral diarrhoea virus », *Vet. Immunol. Immunopathol.*, vol. 116, n° 3-4, p. 153-162, avr. 2007, doi: 10.1016/j.vetimm.2007.01.007.
- [104] R. Osman, P. Gonzalez-Cano, R. Brownlie, et P. J. Griebel, « Induction of interferon and interferon-induced antiviral effector genes following a primary bovine herpesvirus-1 (BHV-1) respiratory infection », *J. Gen. Virol.*, vol. 98, n° 7, p. 1831-1842, juill. 2017, doi: 10.1099/jgv.0.000825.
- [105] B. Bossert, S. Marozin, et K.-K. Conzelmann, « Nonstructural proteins NS1 and NS2 of bovine respiratory syncytial virus block activation of interferon regulatory factor 3 », *J. Virol.*, vol. 77, n° 16, p. 8661-8668, août 2003, doi: 10.1128/jvi.77.16.8661-8668.2003.
- [106] K. C. Eberle, J. L. McGill, T. A. Reinhardt, et R. E. Sacco, « Parainfluenza Virus 3 Blocks Antiviral Mediators Downstream of the Interferon Lambda Receptor by Modulating Stat1 Phosphorylation », *J. Virol.*, vol. 90, n° 6, p. 2948-2958, déc. 2015, doi: 10.1128/JVI.02502-15.
- [107] R. A. Palomares, H. G. Walz, et K. V. Brock, « Expression of type I interferon-induced antiviral state and pro-apoptosis markers during experimental infection with low or high virulence bovine viral diarrhoea virus in beef calves », *Virus Res.*, vol. 173, n° 2, p. 260-269, mai 2013, doi: 10.1016/j.virusres.2013.02.010.
- [108] A. A. Alkheraif *et al.*, « Type 2 BVDV Npro suppresses IFN-1 pathway signaling in bovine cells and augments BRSV replication », *Virology*, vol. 507, p. 123-134, juill. 2017, doi: 10.1016/j.virol.2017.04.015.
- [109] J. J. Rivera-Rivas, D. Kisiela, et C. J. Czuprynski, « Bovine herpesvirus type 1 infection of bovine bronchial epithelial cells increases neutrophil adhesion and activation », *Vet. Immunol. Immunopathol.*, vol. 131, n° 3-4, p. 167-176, oct. 2009, doi: 10.1016/j.vetimm.2009.04.002.
- [110] P. E. Sudaryatma *et al.*, « Bovine respiratory syncytial virus infection enhances *Pasteurella multocida* adherence on respiratory epithelial cells », *Vet. Microbiol.*, vol. 220, p. 33-38, juill. 2018, doi: 10.1016/j.vetmic.2018.04.031.
- [111] C. Senthilkumaran *et al.*, « Increased annexin A1 and A2 levels in bronchoalveolar lavage fluid are associated with resistance to respiratory disease in beef calves », *Vet. Res.*, vol. 44, p. 24, avr. 2013, doi: 10.1186/1297-9716-44-24.
- [112] M. A. Breider, R. D. Walker, F. M. Hopkins, T. W. Schultz, et T. L. Bowersock, « Pulmonary lesions induced by *Pasteurella haemolytica* in neutrophil sufficient and neutrophil deficient calves », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 52, n° 2, p. 205-209, avr. 1988.
- [113] N. A. Aulik, K. M. Hellenbrand, H. Klos, et C. J. Czuprynski, « Mannheimia haemolytica and its leukotoxin cause neutrophil extracellular trap formation by bovine

- neutrophils », *Infect. Immun.*, vol. 78, n° 11, p. 4454-4466, nov. 2010, doi: 10.1128/IAI.00840-10.
- [114] S. Gondaira, H. Higuchi, K. Nishi, H. Iwano, et H. Nagahata, « Mycoplasma bovis escapes bovine neutrophil extracellular traps », *Vet. Microbiol.*, vol. 199, p. 68-73, févr. 2017, doi: 10.1016/j.vetmic.2016.12.022.
- [115] K. M. Hellenbrand, K. M. Forsythe, J. J. Rivera-Rivas, C. J. Czuprynski, et N. A. Aulik, « Histophilus somni causes extracellular trap formation by bovine neutrophils and macrophages », *Microb. Pathog.*, vol. 54, p. 67-75, janv. 2013, doi: 10.1016/j.micpath.2012.09.007.
- [116] B. Cortjens, R. de Jong, J. G. Bonsing, J. B. M. van Woensel, A. F. G. Antonis, et R. A. Bem, « Local dornase alfa treatment reduces NETs-induced airway obstruction during severe RSV infection », *Thorax*, vol. 73, n° 6, p. 578-580, juin 2018, doi: 10.1136/thoraxjnl-2017-210289.
- [117] R. G. Schaut, J. F. Ridpath, et R. E. Sacco, « Bovine Viral Diarrhea Virus Type 2 Impairs Macrophage Responsiveness to Toll-Like Receptor Ligation with the Exception of Toll-Like Receptor 7 », *PLoS One*, vol. 11, n° 7, p. e0159491, 2016, doi: 10.1371/journal.pone.0159491.
- [118] E. Trigo, H. D. Liggitt, J. F. Evermann, R. G. Breeze, L. Y. Huston, et R. Silflow, « Effect of in vitro inoculation of bovine respiratory syncytial virus on bovine pulmonary alveolar macrophage function », *Am. J. Vet. Res.*, vol. 46, n° 5, p. 1098-1103, mai 1985.
- [119] M. A. Risalde *et al.*, « Pathogenic mechanisms implicated in the intravascular coagulation in the lungs of BVDV-infected calves challenged with BHV-1 », *Vet. Res.*, vol. 44, p. 20, mars 2013, doi: 10.1186/1297-9716-44-20.
- [120] S. Ayalew, E. R. Blackwood, et A. W. Confer, « Sequence diversity of the immunogenic outer membrane lipoprotein PlpE from Mannheimia haemolytica serotypes 1, 2, and 6 », *Vet. Microbiol.*, vol. 114, n° 3-4, p. 260-268, mai 2006, doi: 10.1016/j.vetmic.2005.11.067.
- [121] L. J. Berghaus, L. B. Corbeil, R. D. Berghaus, W. V. Kalina, R. A. Kimball, et L. J. Gershwin, « Effects of dual vaccination for bovine respiratory syncytial virus and Haemophilus somnus on immune responses », *Vaccine*, vol. 24, n° 33-34, p. 6018-6027, août 2006, doi: 10.1016/j.vaccine.2006.03.077.
- [122] F. P. Maunsell et C. Chase, « Mycoplasma bovis: Interactions with the Immune System and Failure to Generate an Effective Immune Response », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 35, n° 3, p. 471-483, nov. 2019, doi: 10.1016/j.cvfa.2019.08.003.
- [123] J.-F. Toussaint *et al.*, « Genetic immunisation of cattle against bovine herpesvirus 1: glycoprotein gD confers higher protection than glycoprotein gC or tegument protein VP8 », *Vet. Res.*, vol. 36, n° 4, p. 529-544, août 2005, doi: 10.1051/vetres:2005015.
- [124] M. A. Jutila, J. Holderness, J. C. Graff, et J. F. Hedges, « Antigen-independent priming: a transitional response of bovine gammadelta T-cells to infection », *Anim. Health Res. Rev.*, vol. 9, n° 1, p. 47-57, juin 2008, doi: 10.1017/S1466252307001363.
- [125] J. L. McGill, R. A. Rusk, M. Guerra-Maupome, R. E. Briggs, et R. E. Sacco, « Bovine Gamma Delta T Cells Contribute to Exacerbated IL-17 Production in Response to Co-Infection with Bovine RSV and Mannheimia haemolytica », *PLoS One*, vol. 11, n° 3, p. e0151083, 2016, doi: 10.1371/journal.pone.0151083.
- [126] N. L. Mathy, J.-P. D. Mathy, R. P. Lee, J. Walker, S. Lofthouse, et E. N. T. Meeusen, « Pathological and immunological changes after challenge infection with Pasteurella multocida in naive and immunized calves », *Vet. Immunol. Immunopathol.*, vol. 85, n° 3-4, p. 179-188, mars 2002, doi: 10.1016/s0165-2427(01)00427-5.
- [127] T. J. Vanden Bush et R. F. Rosenbusch, « Characterization of the immune response to Mycoplasma bovis lung infection », *Vet. Immunol. Immunopathol.*, vol. 94, n° 1-2, p. 23-33, juill. 2003, doi: 10.1016/s0165-2427(03)00056-4.

- [128] J. J. Endsley, M. J. Quade, B. Terhaar, et J. A. Roth, « BHV-1-Specific CD4+, CD8+, and gammadelta T cells in calves vaccinated with one dose of a modified live BHV-1 vaccine », *Viral Immunol.*, vol. 15, n° 2, p. 385-393, 2002, doi: 10.1089/08828240260066305.
- [129] R. M. Silflow, P. M. Degel, et A. G. Harmsen, « Bronchoalveolar immune defense in cattle exposed to primary and secondary challenge with bovine viral diarrhoea virus », *Vet. Immunol. Immunopathol.*, vol. 103, n° 1-2, p. 129-139, janv. 2005, doi: 10.1016/j.vetimm.2004.09.008.
- [130] G. Taylor, L. H. Thomas, S. G. Wyld, J. Furze, P. Sopp, et C. J. Howard, « Role of T-lymphocyte subsets in recovery from respiratory syncytial virus infection in calves », *J. Virol.*, vol. 69, n° 11, p. 6658-6664, nov. 1995, doi: 10.1128/JVI.69.11.6658-6664.1995.
- [131] R. L. Levings et J. A. Roth, « Immunity to bovine herpesvirus 1: II. Adaptive immunity and vaccinology », *Anim. Health Res. Rev.*, vol. 14, n° 1, p. 103-123, juin 2013, doi: 10.1017/S1466252313000054.
- [132] A. R. Woolums *et al.*, « Cytotoxic T lymphocyte activity and cytokine expression in calves vaccinated with formalin-inactivated bovine respiratory syncytial virus prior to challenge », *Comp. Immunol. Microbiol. Infect. Dis.*, vol. 27, n° 1, p. 57-74, janv. 2004, doi: 10.1016/S0147-9571(03)00036-5.
- [133] S. De Vliegher *et al.*, « Management practices and heifer characteristics associated with early lactation somatic cell count of Belgian dairy heifers », *J. Dairy Sci.*, vol. 87, n° 4, p. 937-947, avr. 2004, doi: 10.3168/jds.S0022-0302(04)73238-5.
- [134] B. R.I, N. S.c, O. W.e, et W. J.I, « Udder microflora in nonlactating heifers. », *Agri Pract.*, 1987, Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://agris.fao.org/agris-search/search.do?recordID=US19870109549>
- [135] J. R. Roberson, L. K. Fox, D. D. Hancock, J. M. Gay, et T. E. Besser, « Ecology of *Staphylococcus aureus* isolated from various sites on dairy farms », *J. Dairy Sci.*, vol. 77, n° 11, p. 3354-3364, nov. 1994, doi: 10.3168/jds.S0022-0302(94)77277-5.
- [136] J. R. Roberson, L. K. Fox, D. D. Hancock, J. M. Gay, et T. E. Besser, « Sources of intramammary infections from *Staphylococcus aureus* in dairy heifers at first parturition », *J. Dairy Sci.*, vol. 81, n° 3, p. 687-693, mars 1998, doi: 10.3168/jds.S0022-0302(98)75624-3.
- [137] O W. Schalm, « *Streptococcus agalactiae* in the udders of heifers at parturition traced to sucking among calves », *Cornell Vet.*, vol. 32, n° 1, p. 49-60, avr. 1942.
- [138] L. Bassel, D. Kelton, A. Godkin, K. Leslie, et K. Lissemore, « Risk Factors for Intramammary Infection at First Calving in Ontario Dairy Heifers », *Am. Assoc. Bov. Pract. Proc. Annu. Conf.*, p. 177-178, sept. 2003, doi: 10.21423/aabppro20035316.
- [139] M. G. Lopez-Benavides, J. H. Williamson, S. J. Lacy-Hulbert, et R. T. Cursons, « Heifer teats sprayed in the dry period with an iodine teat sanitizer have reduced *Streptococcus uberis* teat-end contamination and less *Streptococcus uberis* intra-mammary infections at calving », *Vet. Microbiol.*, vol. 134, n° 1-2, p. 186-191, févr. 2009, doi: 10.1016/j.vetmic.2008.09.027.
- [140] A. J. Heinrichs, S. S. Costello, et C. M. Jones, « Control of heifer mastitis by nutrition », *Vet. Microbiol.*, vol. 134, n° 1-2, p. 172-176, févr. 2009, doi: 10.1016/j.vetmic.2008.09.025.
- [141] S. McDougall, L. Young, et F. M. Annis, « Production and health of pasture-fed dairy cattle following oral treatment with the ionophore lasalocid », *J. Dairy Sci.*, vol. 87, n° 9, p. 2967-2976, sept. 2004, doi: 10.3168/jds.S0022-0302(04)73428-1.
- [142] T. F. Duffield, A. R. Rabiee, et I. J. Lean, « A meta-analysis of the impact of monensin in lactating dairy cattle. Part 3. Health and reproduction », *J. Dairy Sci.*, vol. 91, n° 6, p. 2328-2341, juin 2008, doi: 10.3168/jds.2007-0801.

- [143] A.-K. Nyman, U. Emanuelson, A. H. Gustafsson, et K. Persson Waller, « Management practices associated with udder health of first-parity dairy cows in early lactation », *Prev. Vet. Med.*, vol. 88, n° 2, p. 138-149, févr. 2009, doi: 10.1016/j.prevetmed.2008.08.005.
- [144] W. P. Weiss, J. S. Hogan, D. A. Todhunter, et K. L. Smith, « Effect of vitamin E supplementation in diets with a low concentration of selenium on mammary gland health of dairy cows », *J. Dairy Sci.*, vol. 80, n° 8, p. 1728-1737, août 1997, doi: 10.3168/jds.S0022-0302(97)76105-8.
- [145] W. P. Weiss, J. S. Hogan, K. L. Smith, et K. H. Hoblet, « Relationships among selenium, vitamin E, and mammary gland health in commercial dairy herds », *J. Dairy Sci.*, vol. 73, n° 2, p. 381-390, févr. 1990, doi: 10.3168/jds.S0022-0302(90)78684-5.
- [146] V. Krömker et J. Friedrich, « Teat canal closure in non-lactating heifers and its association with udder health in the consecutive lactation », *Vet. Microbiol.*, vol. 134, n° 1-2, p. 100-105, févr. 2009, doi: 10.1016/j.vetmic.2008.09.002.
- [147] YrjöT. Gröhn, H. N. Erb, C. E. McCulloch, et H. S. Saloniemi, « Epidemiology of mammary gland disorders in multiparous Finnish Ayrshire cows », *Prev. Vet. Med.*, vol. 8, n° 4, p. 241-252, mai 1990, doi: 10.1016/0167-5877(90)90082-S.
- [148] P. J. Rajala et Y. T. Gröhn, « Disease occurrence and risk factor analysis in Finnish Ayrshire cows », *Acta Vet. Scand.*, vol. 39, n° 1, p. 1-13, 1998.
- [149] A. K. Rathore, « Relationships between teat shape, production and mastitis in Friesian cows », *Br. Vet. J.*, vol. 132, n° 4, p. 389-392, août 1976, doi: 10.1016/s0007-1935(17)34638-9.
- [150] A. J. Seykora et B. T. McDaniel, « Udder and teat morphology related to mastitis resistance: a review », *J. Dairy Sci.*, vol. 68, n° 8, p. 2087-2093, août 1985, doi: 10.3168/jds.S0022-0302(85)81072-9.
- [151] N. Sharma, Singh, NK, S. OP, V. Pandey, et P. Verma, « Oxidative Stress and Antioxidant Status during Transition Period in Dairy Cows », *Asian Australas. J. Anim. Sci.*, vol. 24, p. 479-484, avr. 2011, doi: 10.5713/ajas.2011.10220.
- [152] A. Bradley et M. Green, « Use and interpretation of somatic cell counts in the dairy cow », *In Pract.*, vol. 27, juin 2005, doi: 10.1136/inpract.27.6.310.
- [153] W. N. Plastringe, « Bovine Mastitis: A Review », *J. Dairy Sci.*, vol. 41, n° 9, p. 1141-1181, sept. 1958, doi: 10.3168/jds.S0022-0302(58)91071-3.
- [154] A. Akyuz, S. Boyaci, et A. Çayli, « Determination of Critical Period for Dairy Cows Using Temperature Humidity Index », *J. Anim. Vet. Adv. - J ANIM VET ADV*, vol. 9, p. 1824-1827, déc. 2010, doi: 10.3923/javaa.2010.1824.1827.
- [155] « Weather effect on bacterial mastitis in dairy cows. » <https://www.cabdirect.org/cabdirect/abstract/20093196539> (consulté le mai 07, 2021).
- [156] A. I. Donaldson, « Factors influencing the dispersal, survival and deposition of airborne pathogens of farm animals », *Vet. Bull.*, 1978, Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://agris.fao.org/agris-search/search.do?recordID=US201301279386>
- [157] A. Lago, S. M. McGuirk, T. B. Bennett, N. B. Cook, et K. V. Nordlund, « Calf respiratory disease and pen microenvironments in naturally ventilated calf barns in winter », *J. Dairy Sci.*, vol. 89, n° 10, p. 4014-4025, oct. 2006, doi: 10.3168/jds.S0022-0302(06)72445-6.
- [158] W. C. Losinger et A. J. Heinrichs, « Management variables associated with high mortality rates attributable to respiratory tract problems in female calves prior to weaning », *J. Am. Vet. Med. Assoc.*, vol. 209, n° 10, p. 1756-1759, nov. 1996.
- [159] M. H. Mars, C. J. Brusckke, et J. T. van Oirschot, « Airborne transmission of BHV1, BRSV, and BVDV among cattle is possible under experimental conditions », *Vet. Microbiol.*, vol. 66, n° 3, p. 197-207, avr. 1999, doi: 10.1016/s0378-1135(99)00009-7.

- [160] M. H. Mars, M. C. de Jong, C. van Maanen, J. J. Hage, et J. T. van Oirschot, « Airborne transmission of bovine herpesvirus 1 infections in calves under field conditions », *Vet. Microbiol.*, vol. 76, n° 1, p. 1-13, sept. 2000, doi: 10.1016/s0378-1135(00)00218-2.
- [161] R. J. Callan et F. B. Garry, « Biosecurity and bovine respiratory disease », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 18, n° 1, p. 57-77, mars 2002, doi: 10.1016/s0749-0720(02)00004-x.
- [162] G. A. Donovan, I. R. Dohoo, D. M. Montgomery, et F. L. Bennett, « Calf and disease factors affecting growth in female Holstein calves in Florida, USA », *Prev. Vet. Med.*, vol. 33, n° 1-4, p. 1-10, janv. 1998, doi: 10.1016/s0167-5877(97)00059-7.
- [163] J. Van Donkersgoed, C. S. Ribble, L. G. Boyer, et H. G. Townsend, « Epidemiological study of enzootic pneumonia in dairy calves in Saskatchewan », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 57, n° 4, p. 247-254, oct. 1993.
- [164] A. M. Virtala, Y. T. Gröhn, G. D. Mechor, et H. N. Erb, « The effect of maternally derived immunoglobulin G on the risk of respiratory disease in heifers during the first 3 months of life », *Prev. Vet. Med.*, vol. 39, n° 1, p. 25-37, mars 1999, doi: 10.1016/s0167-5877(98)00140-8.
- [165] USDA, « Heifer calf health and management practices on U.S. dairy operations », 2007.
- [166] A. Bach, J. Ahedo, et A. Ferrer, « Optimizing weaning strategies of dairy replacement calves », *J. Dairy Sci.*, vol. 93, n° 1, p. 413-419, janv. 2010, doi: 10.3168/jds.2009-2682.
- [167] J. Quigley, « Calf Note #16 - Stress at weaning », p. 2, 2001.
- [168] P. J. Gorden et P. Plummer, « Control, Management, and Prevention of Bovine Respiratory Disease in Dairy Calves and Cows », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 26, n° 2, p. 243-259, juill. 2010, doi: 10.1016/j.cvfa.2010.03.004.
- [169] B. J. Nonnecke, M. R. Foote, B. L. Miller, M. Fowler, T. E. Johnson, et R. L. Horst, « Effects of chronic environmental cold on growth, health, and select metabolic and immunologic responses of preruminant calves », *J. Dairy Sci.*, vol. 92, n° 12, p. 6134-6143, déc. 2009, doi: 10.3168/jds.2009-2517.
- [170] A. L. Stanton *et al.*, « The effect of treatment with long-acting antibiotic at postweaning movement on respiratory disease and on growth in commercial dairy calves », *J. Dairy Sci.*, vol. 93, n° 2, p. 574-581, févr. 2010, doi: 10.3168/jds.2009-2414.
- [171] F. Maunsell et G. A. Donovan, « Biosecurity and risk management for dairy replacements », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 24, n° 1, p. 155-190, mars 2008, doi: 10.1016/j.cvfa.2007.10.007.
- [172] U. Vecht, G. E. Shook, R. D. Politiek, G. Grootenhuis, W. J. Koops, et D. G. Groothuis, « Effect of bull selection for somatic cell count in first lactation on cell counts and pathogens in later lactations », *J. Dairy Sci.*, vol. 68, n° 11, p. 2995-3003, nov. 1985, doi: 10.3168/jds.S0022-0302(85)81195-4.
- [173] M. M. Schutz, « Genetic evaluation of somatic cell scores for United States dairy cattle », *J. Dairy Sci.*, vol. 77, n° 7, p. 2113-2129, juill. 1994, doi: 10.3168/jds.S0022-0302(94)77154-X.
- [174] M. S. Lund, J. Jensen, et P. H. Petersen, « Estimation of genetic and phenotypic parameters for clinical mastitis, somatic cell production deviance, and protein yield in dairy cattle using Gibbs sampling », *J. Dairy Sci.*, vol. 82, n° 5, p. 1045-1051, mai 1999, doi: 10.3168/jds.S0022-0302(99)75325-7.
- [175] B. Heringstad, D. Gianola, Y. M. Chang, J. Odegård, et G. Klemetsdal, « Genetic associations between clinical mastitis and somatic cell score in early first-lactation cows », *J. Dairy Sci.*, vol. 89, n° 6, p. 2236-2244, juin 2006, doi: 10.3168/jds.S0022-0302(06)72295-0.
- [176] K. L. Parker Gaddis, F. Tiezzi, J. B. Cole, J. S. Clay, et C. Maltecca, « Genomic prediction of disease occurrence using producer-recorded health data: a comparison of

- methods », *Genet. Sel. Evol. GSE*, vol. 47, p. 41, mai 2015, doi: 10.1186/s12711-015-0093-9.
- [177] L. M. Berning et G. E. Shook, « Prediction of mastitis using milk somatic cell count, N-acetyl-beta-D-glucosaminidase, and lactose », *J. Dairy Sci.*, vol. 75, n° 7, p. 1840-1848, juill. 1992, doi: 10.3168/jds.S0022-0302(92)77943-0.
- [178] E. Norberg, « Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review », *Livest. Prod. Sci.*, vol. 2-3, n° 96, p. 129-139, 2005, doi: 10.1016/j.livprodsci.2004.12.014.
- [179] « Defining candidate genes for mastitis resistance in cattle: the role of lactoferrin and lysozyme - Seyfert - 1996 - Journal of Animal Breeding and Genetics - Wiley Online Library ». <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1439-0388.1996.tb00616.x> (consulté le mai 07, 2021).
- [180] R. J. Wall *et al.*, « Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection », *Nat. Biotechnol.*, vol. 23, n° 4, p. 445-451, avr. 2005, doi: 10.1038/nbt1078.
- [181] N. E. Muggli-Cockett, L. V. Cundiff, et K. E. Gregory, « Genetic analysis of bovine respiratory disease in beef calves during the first year of life », *J. Anim. Sci.*, vol. 70, n° 7, p. 2013-2019, juill. 1992, doi: 10.2527/1992.7072013x.
- [182] G. D. Snowder, L. D. Van Vleck, L. V. Cundiff, et G. L. Bennett, « Influence of breed, heterozygosity, and disease incidence on estimates of variance components of respiratory disease in preweaned beef calves », *J. Anim. Sci.*, vol. 83, n° 6, p. 1247-1261, juin 2005, doi: 10.2527/2005.8361247x.
- [183] G. D. Snowder, L. D. Van Vleck, L. V. Cundiff, et G. L. Bennett, « Bovine respiratory disease in feedlot cattle: environmental, genetic, and economic factors », *J. Anim. Sci.*, vol. 84, n° 8, p. 1999-2008, août 2006, doi: 10.2527/jas.2006-046.
- [184] R. Rupp *et al.*, « A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model », *PLoS Genet.*, vol. 11, n° 12, p. e1005629, déc. 2015, doi: 10.1371/journal.pgen.1005629.
- [185] K. L. Smith et F. L. Schanbacher, « Lactoferrin as a factor of resistance to infection of the bovine mammary gland », *J. Am. Vet. Med. Assoc.*, vol. 170, n° 10 Pt 2, p. 1224-1227, mai 1977.
- [186] C. J.s, « The Escherichia coli J5 vaccine: investigating a new tool to combat coliform mastitis. », *Vet. Med.*, 1991, Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://agris.fao.org/agris-search/search.do?recordID=US19910102071>
- [187] J. W. Tyler, J. S. Cullor, S. J. Spier, et B. P. Smith, « Immunity targeting common core antigens of gram-negative bacteria », *J. Vet. Intern. Med.*, vol. 4, n° 1, p. 17-25, févr. 1990, doi: 10.1111/j.1939-1676.1990.tb00870.x.
- [188] J. W. Tyler, J. S. Cullor, et J. D. Dellinger, « Cross-reactive affinity purification of immunoglobulin recognizing common gram-negative bacterial core antigens », *J. Immunol. Methods*, vol. 129, n° 2, p. 221-226, mai 1990, doi: 10.1016/0022-1759(90)90442-x.
- [189] G. M. Tomita, D. A. Todhunter, J. S. Hogan, et K. L. Smith, « Antigenic crossreactivity and lipopolysaccharide neutralization properties of bovine immunoglobulin G », *J. Dairy Sci.*, vol. 78, n° 12, p. 2745-2752, déc. 1995, doi: 10.3168/jds.S0022-0302(95)76905-3.
- [190] A. Chaiyotwittayakun, J. L. Burton, P. S. D. Weber, K. Kizilkaya, F. F. Cardoso, et R. J. Erskine, « Hyperimmunization of steers with J5 *Escherichia coli* bacterin: effects on isotype-specific serum antibody responses and cross reactivity with heterogeneous gram-negative bacteria », *J. Dairy Sci.*, vol. 87, n° 10, p. 3375-3385, oct. 2004, doi: 10.3168/jds.S0022-0302(04)73473-6.

- [191] J. W. Tyler, J. S. Cullor, B. I. Osburn, R. B. Bushnell, et B. W. Fenwick, « Relationship between serologic recognition of Escherichia coli 0111:B4 (J5) and clinical coliform mastitis in cattle », *Am. J. Vet. Res.*, vol. 49, n° 11, p. 1950-1954, nov. 1988.
- [192] R. N. González, J. S. Cullor, D. E. Jasper, T. B. Farver, R. B. Bushnell, et M. N. Oliver, « Prevention of clinical coliform mastitis in dairy cows by a mutant Escherichia coli vaccine », *Can. J. Vet. Res. Rev. Can. Rech. Veterinaire*, vol. 53, n° 3, p. 301-305, juill. 1989.
- [193] J. S. Hogan, W. P. Weiss, D. A. Todhunter, K. L. Smith, et P. S. Schoenberger, « Efficacy of an Escherichia coli J5 mastitis vaccine in an experimental challenge trial », *J. Dairy Sci.*, vol. 75, n° 2, p. 415-422, févr. 1992, doi: 10.3168/jds.S0022-0302(92)77777-7.
- [194] G. M. Tomita, C. H. Ray, S. C. Nickerson, W. E. Owens, et G. F. Gallo, « A comparison of two commercially available Escherichia coli J5 vaccines against E. coli intramammary challenge », *J. Dairy Sci.*, vol. 83, n° 10, p. 2276-2281, oct. 2000, doi: 10.3168/jds.S0022-0302(00)75112-5.
- [195] J. S. Hogan, K. L. Smith, D. A. Todhunter, et P. S. Schoenberger, « Field trial to determine efficacy of an Escherichia coli J5 mastitis vaccine », *J. Dairy Sci.*, vol. 75, n° 1, p. 78-84, janv. 1992, doi: 10.3168/jds.S0022-0302(92)77741-8.
- [196] D. J. Wilson, Y. T. Grohn, G. J. Bennett, R. N. González, Y. H. Schukken, et J. Spatz, « Comparison of J5 vaccinates and controls for incidence, etiologic agent, clinical severity, and survival in the herd following naturally occurring cases of clinical mastitis », *J. Dairy Sci.*, vol. 90, n° 9, p. 4282-4288, sept. 2007, doi: 10.3168/jds.2007-0160.
- [197] D. J. Wilson, B. A. Mallard, J. L. Burton, Y. H. Schukken, et Y. T. Grohn, « Association of Escherichia coli J5-specific serum antibody responses with clinical mastitis outcome for J5 vaccinate and control dairy cattle », *Clin. Vaccine Immunol. CVI*, vol. 16, n° 2, p. 209-217, févr. 2009, doi: 10.1128/CVI.00324-08.
- [198] R. J. Erskine, E. J. VanDyk, P. C. Bartlett, J. L. Burton, et M. C. Boyle, « Effect of hyperimmunization with an Escherichia coli J5 bacterin in adult lactating dairy cows », *J. Am. Vet. Med. Assoc.*, vol. 231, n° 7, p. 1092-1097, oct. 2007, doi: 10.2460/javma.231.7.1092.
- [199] R. J. Erskine, A. R. Brockett, N. D. Beeching, R. W. Hull, et P. C. Bartlett, « Effect of changes in number of doses and anatomic location for administration of an Escherichia coli bacterin on serum IgG1 and IgG2 concentrations in dairy cows », *Am. J. Vet. Res.*, vol. 71, n° 1, p. 120-124, janv. 2010, doi: 10.2460/ajvr.71.1.120.
- [200] A. M. McClure, E. E. Christopher, W. A. Wolff, W. H. Fales, G. F. Krause, et J. Miramonti, « Effect of Re-17 mutant Salmonella typhimurium bacterin toxoid on clinical coliform mastitis », *J. Dairy Sci.*, vol. 77, n° 8, p. 2272-2280, août 1994, doi: 10.3168/jds.S0022-0302(94)77170-8.
- [201] S. C. Nickerson, W. E. Owens, G. M. Tomita, et P. W. Widel, « Vaccinating dairy heifers with a Staphylococcus aureus bacterin reduces mastitis at calving », *Large Anim. Pract.*, 1999, Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://agris.fao.org/agris-search/search.do?recordID=US201302917449>
- [202] J. R. Middleton, J. Ma, C. L. Rinehart, V. N. Taylor, C. D. Luby, et B. J. Steevens, « Efficacy of different Lysigin formulations in the prevention of Staphylococcus aureus intramammary infection in dairy heifers », *J. Dairy Res.*, vol. 73, n° 1, p. 10-19, févr. 2006, doi: 10.1017/S0022029905001354.
- [203] J. R. Middleton, C. D. Luby, et D. S. Adams, « Efficacy of vaccination against staphylococcal mastitis: a review and new data », *Vet. Microbiol.*, vol. 134, n° 1-2, p. 192-198, févr. 2009, doi: 10.1016/j.vetmic.2008.09.053.

- [204] « Development of a Staphylococcus aureus vaccine against mastitis in dairy cows. I. Challenge trials - PubMed ». <https://pubmed.ncbi.nlm.nih.gov/12753773/> (consulté le mai 07, 2021).
- [205] A. Prenafeta, R. March, A. Foix, I. Casals, et L. Costa, « Study of the humoral immunological response after vaccination with a Staphylococcus aureus biofilm-embedded bacterin in dairy cows: possible role of the exopolysaccharide specific antibody production in the protection from Staphylococcus aureus induced mastitis », *Vet. Immunol. Immunopathol.*, vol. 134, n° 3-4, p. 208-217, avr. 2010, doi: 10.1016/j.vetimm.2009.09.020.
- [206] Jiminez LM, Romero C, « Efficacy of vaccination on mastitis epidemiology: field study », Proceedings of the 50th Annual National Mastitis Council Meeting. Verona (WI): National Mastitis Council; 2011.
- [207] Noguera M, March R, Guix R, et al, « Evaluation of the efficacy of a new vaccine against bovine mastitis caused by CNS field trial results », Proceedings of the 50th Annual National Mastitis Council Meeting. Verona (WI): National Mastitis Council, 2011.
- [208] S. Piepers, A. Prenafeta, J. Verbeke, A. De Visscher, R. March, et S. De Vlieghe, « Immune response after an experimental intramammary challenge with killed Staphylococcus aureus in cows and heifers vaccinated and not vaccinated with Startvac, a polyvalent mastitis vaccine », *J. Dairy Sci.*, vol. 100, n° 1, p. 769-782, janv. 2017, doi: 10.3168/jds.2016-11269.
- [209] B. M. Jayarao, B. E. Gillespie, M. J. Lewis, H. H. Dowlen, et S. P. Oliver, « Epidemiology of Streptococcus uberis intramammary infections in a dairy herd », *Zentralblatt Vet. Reihe B J. Vet. Med. Ser. B*, vol. 46, n° 7, p. 433-442, sept. 1999, doi: 10.1046/j.1439-0450.1999.00254.x.
- [210] B. Tamilselvam, R. A. Almeida, J. R. Dunlap, et S. P. Oliver, « Streptococcus uberis internalizes and persists in bovine mammary epithelial cells », *Microb. Pathog.*, vol. 40, n° 6, p. 279-285, juin 2006, doi: 10.1016/j.micpath.2006.02.006.
- [211] J. M. Finch, A. Winter, A. W. Walton, et J. A. Leigh, « Further studies on the efficacy of a live vaccine against mastitis caused by Streptococcus uberis », *Vaccine*, vol. 15, n° 10, p. 1138-1143, juill. 1997, doi: 10.1016/s0264-410x(96)00307-6.
- [212] D. A. Luther, R. A. Almeida, et S. P. Oliver, « Elucidation of the DNA sequence of Streptococcus uberis adhesion molecule gene (sua) and detection of sua in strains of Streptococcus uberis isolated from geographically diverse locations », *Vet. Microbiol.*, vol. 128, n° 3-4, p. 304-312, avr. 2008, doi: 10.1016/j.vetmic.2007.10.015.
- [213] M. E. Prado *et al.*, « Vaccination of dairy cows with recombinant Streptococcus uberis adhesion molecule induces antibodies that reduce adherence to and internalization of S. uberis into bovine mammary epithelial cells », *Vet. Immunol. Immunopathol.*, vol. 141, n° 3-4, p. 201-208, juin 2011, doi: 10.1016/j.vetimm.2011.02.023.
- [214] W. D. Yates, B. F. Kingscote, J. A. Bradley, et D. Mitchell, « The relationship of serology and nasal microbiology to pulmonary lesions in feedlot cattle », *Can. J. Comp. Med. Rev. Can. Med. Comp.*, vol. 47, n° 3, p. 375-378, juill. 1983.
- [215] D. Griffin, « Feedlot diseases », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 14, n° 2, p. 199-231, juill. 1998, doi: 10.1016/s0749-0720(15)30251-6.
- [216] J. A. Roth et L. J. Perino, « Immunology and prevention of infection in feedlot cattle », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 14, n° 2, p. 233-256, juill. 1998, doi: 10.1016/s0749-0720(15)30252-8.
- [217] D. E. (Oregon S. U. Hansen, R. Syvrud, et D. Armstrong, « Effectiveness of a bovine respiratory syncytial virus vaccine in reducing the risk of respiratory disease », *Agri-Pract. USA*, 1992, Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://agris.fao.org/agris-search/search.do?recordID=US9407538>

- [218] « The Effect of bovine respiratory syncytial virus vaccination on health, feedlot performance and carcass characteristics of feeder cattle | The Bovine Practitioner », Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://journals.tdl.org/bovine/index.php/bovine/article/view/1717>
- [219] P. Kerkhofs, M. Tignon, H. Petry, I. Mawhinney, et B. Sustronck, « Immune responses to bovine respiratory syncytial virus (BRSV) following use of an inactivated BRSV-PI3-Mannheimia haemolytica vaccine and a modified live BRSV-BVDV vaccine », *Vet. J. Lond. Engl.* 1997, vol. 167, n° 2, p. 208-210, mars 2004, doi: 10.1016/S1090-0233(03)00078-9.
- [220] K. F. Fairbanks, J. Campbell, et C. C. L. Chase, « Rapid onset of protection against infectious bovine rhinotracheitis with a modified-live virus multivalent vaccine », *Vet. Ther. Res. Appl. Vet. Med.*, vol. 5, n° 1, p. 17-25, 2004.
- [221] B. De Groot, C. E. Dewey, D. D. Griffin, L. J. Perino, R. A. Moxley, et G. L. Hahn, « Effect of booster vaccination with a multivalent clostridial bacterin-toxoid on sudden death syndrome mortality rate among feedlot cattle », *J. Am. Vet. Med. Assoc.*, vol. 211, n° 6, p. 749-753, sept. 1997.
- [222] Veterinary Learning Systems, *Bovine Respiratory Disease: Sourcebook for the Veterinary Professional*. 1995.
- [223] D. A. Mosier, « Bacterial pneumonia », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 13, n° 3, p. 483-493, nov. 1997, doi: 10.1016/s0749-0720(15)30310-8.
- [224] C. W. Purdy, R. H. Raleigh, J. K. Collins, J. L. Watts, et D. C. Straus, « Serotyping and enzyme characterization of Pasteurella haemolytica and Pasteurella multocida isolates recovered from pneumonic lungs of stressed feeder calves », *Curr. Microbiol.*, vol. 34, n° 4, p. 244-249, avr. 1997, doi: 10.1007/s002849900177.
- [225] J. W. McBride, E. J. Wozniak, A. W. Brewer, D. K. Naydan, et B. I. Osburn, « Evidence of Pasteurella haemolytica linked immune complex disease in natural and experimental models », *Microb. Pathog.*, vol. 26, n° 4, p. 183-193, avr. 1999, doi: 10.1006/mpat.1998.0266.
- [226] « Comparative efficacy and duration of immunity of commercial Pasteurella haemolytica vaccines | The Bovine Practitioner », nov. 2019, Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: <https://journals.tdl.org/bovine/index.php/bovine/article/view/1548>
- [227] A. W. Confer, S. Ayalew, R. J. Panciera, M. Montelongo, L. C. Whitworth, et J. D. Hammer, « Immunogenicity of recombinant Mannheimia haemolytica serotype 1 outer membrane protein PlpE and augmentation of a commercial vaccine », *Vaccine*, vol. 21, n° 21-22, p. 2821-2829, juin 2003, doi: 10.1016/s0264-410x(03)00213-5.
- [228] S. R.a, G. D.r, et H. R.b, « Improving the performance of stocker and feedlot calves with a live Pasteurella haemolytica vaccine. », *Vet. Med.*, 1986, Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://agris.fao.org/agris-search/search.do?recordID=US19870077232>
- [229] K. J. Hill WJ, « The effects of Septimmune on health and performance of stressed stocker cattle », 1993.
- [230] K. F. Lechtenberg, R. A. Smith, et G. L. Stokka, « Feedlot health and management », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 14, n° 2, p. 177-197, juill. 1998, doi: 10.1016/s0749-0720(15)30250-4.
- [231] « Evaluating the Efficacy of a Haemophilus Somnus Bacterin in a Controlled Field Trial | The Bovine Practitioner », Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://journals.tdl.org/bovine/index.php/bovine/article/view/3663>
- [232] C. A. Hjerpe, « The role of mycoplasma in bovine respiratory disease », *Vet. Med. Small Anim. Clin. VM SAC*, vol. 75, n° 2, p. 297-298, févr. 1980.

- [233] F. M. Shahriar, E. G. Clark, E. Janzen, K. West, et G. Wobeser, « Coinfection with bovine viral diarrhoea virus and *Mycoplasma bovis* in feedlot cattle with chronic pneumonia », *Can. Vet. J. Rev. Veterinaire Can.*, vol. 43, n° 11, p. 863-868, nov. 2002.
- [234] R. A. J. Nicholas, R. D. Ayling, et L. P. Stipkovits, « An experimental vaccine for calf pneumonia caused by *Mycoplasma bovis*: clinical, cultural, serological and pathological findings », *Vaccine*, vol. 20, n° 29-30, p. 3569-3575, oct. 2002, doi: 10.1016/s0264-410x(02)00340-7.
- [235] C. J. Howard, E. J. Stott, L. H. Thomas, R. N. Gourlay, et G. Taylor, « Protection against respiratory disease in calves induced by vaccines containing respiratory syncytial virus, parainfluenza type 3 virus, *Mycoplasma bovis* and M dispar », *Vet. Rec.*, vol. 121, n° 16, p. 372-376, oct. 1987, doi: 10.1136/vr.121.16.372.
- [236] R. W. Fulton et L. J. Burge, « Bovine viral diarrhoea virus types 1 and 2 antibody response in calves receiving modified live virus or inactivated vaccines », *Vaccine*, vol. 19, n° 2-3, p. 264-274, sept. 2000, doi: 10.1016/s0264-410x(00)00168-7.
- [237] J. J. Endsley, M. J. Quade, B. Terhaar, et J. A. Roth, « Bovine viral diarrhoea virus type 1- and type 2-specific bovine T lymphocyte-subset responses following modified-live virus vaccination », *Vet. Ther. Res. Appl. Vet. Med.*, vol. 3, n° 4, p. 364-372, 2002.
- [238] « Comparison of neutralizing antibodies to type 1a, 1b and 2 bovine viral diarrhoea virus from experimentally infected and vaccinated cattle | The Bovine Practitioner », Consulté le: mai 07, 2021. [En ligne]. Disponible sur: <https://journals.tdl.org/bovine/index.php/bovine/article/view/1644>
- [239] C. L. Kelling, « Evolution of bovine viral diarrhoea virus vaccines », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 20, n° 1, p. 115-129, mars 2004, doi: 10.1016/j.cvfa.2003.11.001.
- [240] R. W. Fulton, J. T. Saliki, L. J. Burge, et M. E. Payton, « Humoral immune response and assessment of vaccine virus shedding in calves receiving modified live virus vaccines containing bovine herpesvirus-1 and bovine viral diarrhoea virus 1a », *J. Vet. Med. B Infect. Dis. Vet. Public Health*, vol. 50, n° 1, p. 31-37, févr. 2003, doi: 10.1046/j.1439-0450.2003.00608.x.
- [241] P. Ruegg, « Relationship between bulk tank milk somatic cell count and antibiotic residues », *Proc 44th Natl. Mastit. Counc. Mtg*, janv. 2005.
- [242] C. Boireau *et al.*, « Antimicrobial resistance in bacteria isolated from mastitis in dairy cattle in France, 2006-2016 », *J. Dairy Sci.*, vol. 101, n° 10, p. 9451-9462, oct. 2018, doi: 10.3168/jds.2018-14835.
- [243] B. Faye et L. Pérochon, « [Mortality of dairy cows in an ecopathologic survey in Brittany] », *Vet. Res.*, vol. 26, n° 2, p. 124-131, 1995.
- [244] « Evaluation de l'impact économique de la santé en élevage laitier - Enquêtes (...) - 3R - Rencontres autour des Recherches sur les Ruminants ». <http://www.journees3r.fr/spip.php?article1566> (consulté le mai 07, 2021).
- [245] « Respiratory diseases of feedlot cattle in central USA | The Bovine Practitioner », nov. 2019, Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: <https://journals.tdl.org/bovine/index.php/bovine/article/view/2441>
- [246] E. Gay et J. Barnouin, « A nation-wide epidemiological study of acute bovine respiratory disease in France », *Prev. Vet. Med.*, vol. 89, n° 3-4, p. 265-271, juin 2009, doi: 10.1016/j.prevetmed.2009.02.013.
- [247] B. P. Holland *et al.*, « Effect of bovine respiratory disease during preconditioning on subsequent feedlot performance, carcass characteristics, and beef attributes », *J. Anim. Sci.*, vol. 88, n° 7, p. 2486-2499, juill. 2010, doi: 10.2527/jas.2009-2428.
- [248] N. Cernicchiaro, B. J. White, D. G. Renter, et A. H. Babcock, « Evaluation of economic and performance outcomes associated with the number of treatments after an initial

- diagnosis of bovine respiratory disease in commercial feeder cattle », *Am. J. Vet. Res.*, vol. 74, n° 2, p. 300-309, févr. 2013, doi: 10.2460/ajvr.74.2.300.
- [249] « Case report | The Bovine Practitioner », nov. 2019, Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: <https://bovine-ojs-tamu.tdl.org/bovine/index.php/bovine/article/view/2798>
- [250] B. J. White et D. G. Renter, « Bayesian estimation of the performance of using clinical observations and harvest lung lesions for diagnosing bovine respiratory disease in post-weaned beef calves », *J. Vet. Diagn. Investig. Off. Publ. Am. Assoc. Vet. Lab. Diagn. Inc.*, vol. 21, n° 4, p. 446-453, juill. 2009, doi: 10.1177/104063870902100405.
- [251] N. K. Chirase et L. W. Greene, « Dietary zinc and manganese sources administered from the fetal stage onwards affect immune response of transit stressed and virus infected offspring steer calves », *Anim. Feed Sci. Technol.*, vol. 93, n° 3, p. 217-228, oct. 2001, doi: 10.1016/S0377-8401(01)00277-2.
- [252] K. R. Brooks, K. C. Raper, C. E. Ward, B. P. Holland, C. R. Krehbiel, et D. L. Step, « Economic effects of bovine respiratory disease on feedlot cattle during backgrounding and finishing phases1 », *Prof. Anim. Sci.*, vol. 27, n° 3, p. 195-203, juin 2011, doi: 10.15232/S1080-7446(15)30474-5.
- [253] B. K. Wilson, D. L. Step, C. L. Maxwell, C. A. Gifford, C. J. Richards, et C. R. Krehbiel, « Effect of bovine respiratory disease during the receiving period on steer finishing performance, efficiency, carcass characteristics, and lung scores », *Prof. Anim. Sci.*, vol. 33, n° 1, p. 24-36, févr. 2017, doi: 10.15232/pas.2016-01554.
- [254] C. Blakebrough-Hall, J. P. McMeniman, et L. A. González, « An evaluation of the economic effects of bovine respiratory disease on animal performance, carcass traits, and economic outcomes in feedlot cattle defined using four BRD diagnosis methods », *J. Anim. Sci.*, vol. 98, n° 2, févr. 2020, doi: 10.1093/jas/skaa005.
- [255] W. H. Chang et G. R. Carter, « Multiple drug resistance in *Pasteurella multocida* and *Pasteurella haemolytica* from cattle and swine », *J. Am. Vet. Med. Assoc.*, vol. 169, n° 7, p. 710-712, oct. 1976.
- [256] A. M. O'Connor *et al.*, « A systematic review and network meta-analysis of injectable antibiotic options for the control of bovine respiratory disease in the first 45 days post arrival at the feedlot », *Anim. Health Res. Rev.*, vol. 20, n° 2, p. 163-181, déc. 2019, doi: 10.1017/S1466252320000031.
- [257] C. Herrenschmidt, « Marcia Ascher, Mathématiques d'ailleurs. Nombres, formes et jeux dans les sociétés traditionnelles », *L'Homme Rev. Fr. D'anthropologie*, n° 153, Art. n° 153, janv. 2000.
- [258] B. Pascal, *Traité du triangle arithmétique, avec quelques autres petits traités sur la même matière. Par monsieur Pascal*. Guil. Desprez, 1665.
- [259] www.bibliopolis.com, « Varia Opera Mathematica...accesserunt selectae quaedam eiusdem Epistolae, vel ad ipsum à plerisque doctissimis viris Gallicè, Latinè, vel Italicè, de rebus ad Mathematicas disciplinas, aut Physicam », *JONATHAN A. HILL, BOOKSELLER, INC.* <https://www.jonathanahill.com/pages/books/3105/pierre-de-fermat/varia-opera-mathematica-accesserunt-selectae-quaedam-eiusdem-epistolae-vel-ad-ipsum-a-plerisque> (consulté le oct. 04, 2021).
- [260] DBNL, « Christiaan Huygens, Oeuvres complètes. Tome XIV. Probabilités. Travaux de mathématiques pures 1655-1666 · dbnl », *DBNL*. https://www.dbnl.org/tekst/huyg003oeuv14_01/huyg003oeuv14_01_0055.php (consulté le oct. 04, 2021).
- [261] M. Mattmüller, « The difficult birth of stochastics: Jacob Bernoulli's *Ars Conjectandi* (1713) », *Hist. Math.*, vol. 41, n° 3, p. 277-290, août 2014, doi: 10.1016/j.hm.2014.04.001.

- [262] C. Duflo, « Antoine Deparcieux : Essai sur les probabilités de la durée de la vie humaine (1746). Addition à l'Essai (1760). Réédition en fac-similé avec une introduction et des notes de C. Behar, accompagnée de contributions de G. Gallais-Hamonno, C. Tietsch, J. Berthon, 2003 », *Dix-Huit. Siècle*, vol. 36, n° 1, p. 601-601, 2004.
- [263] « Probabilité conditionnelle », *Wikipédia*. août 17, 2021. Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: https://fr.wikipedia.org/w/index.php?title=Probabilit%C3%A9_conditionnelle&oldid=185563448
- [264] « Biographie de Thomas Bayes ». <https://www.bibmath.net/bios/index.php?action=affiche&quoi=bayes> (consulté le oct. 04, 2021).
- [265] « Jonathan Sachs, “1786/1801: William Playfair, Statistical Graphics, and the Meaning of an Event” | BRANCH ». http://www.branchcollective.org/?ps_articles=jonathan-sachs-17861801-william-playfair-statistical-graphics-and-the-meaning-of-an-event (consulté le oct. 04, 2021).
- [266] S. Rogers, « Florence Nightingale, datajournalist: information has always been beautiful », *The Guardian*, août 13, 2010. Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: <https://www.theguardian.com/news/datablog/2010/aug/13/florence-nightingale-graphics>
- [267] *Mémoires de la classe des sciences mathématiques et physiques de l'Institut National de France*. Institut National de France, 1809.
- [268] Aimé FUCHS, « PLAIDOYER POUR LA LOI NORMALE », *Pour la science*, 1995.
- [269] « The Concise Encyclopedia of Statistics | Yadolah Dodge | Springer ». <https://www.springer.com/gp/book/9780387328331> (consulté le oct. 04, 2021).
- [270] R. W. Farebrother, « Adrien-Marie Legendre », in *Statisticians of the Centuries*, C. C. Heyde, E. Seneta, P. Crépel, S. E. Fienberg, et J. Gani, Éd. New York, NY: Springer, 2001, p. 101-104. doi: 10.1007/978-1-4613-0179-0_20.
- [271] S. M. Stigler, « Karl Pearson's Theoretical Errors and the Advances They Inspired », *Stat. Sci.*, vol. 23, n° 2, p. 261-271, 2008.
- [272] J. Venn, « I. On the diagrammatic and mechanical representation of propositions and reasonings », *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, vol. 10, n° 59, p. 1-18, juill. 1880, doi: 10.1080/14786448008626877.
- [273] E. Universalis, « GALTON sir FRANCIS », *Encyclopædia Universalis*. <https://www.universalis.fr/encyclopedie/galton-sir-francis/> (consulté le oct. 04, 2021).
- [274] « Spearman Rank Correlation Coefficient », in *The Concise Encyclopedia of Statistics*, New York, NY: Springer, 2008, p. 502-505. doi: 10.1007/978-0-387-32833-1_379.
- [275] Student, « The Probable Error of a Mean », *Biometrika*, vol. 6, n° 1, p. 1-25, 1908, doi: 10.2307/2331554.
- [276] W. G. Cochran, « Fisher and the Analysis of Variance », in *R.A. Fisher: An Appreciation*, New York, NY, 1980, p. 17-34. doi: 10.1007/978-1-4612-6079-0_4.
- [277] « The Impact of R. A. Fisher on Statistics: Journal of the American Statistical Association: Vol 46, No 253 ». <https://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500765> (consulté le oct. 04, 2021).
- [278] J. F. Box, « R.A. Fisher and the Design of Experiments, 1922–1926 », *Am. Stat.*, vol. 34, n° 1, p. 1-7, févr. 1980, doi: 10.1080/00031305.1980.10482701.
- [279] « The testing of statistical hypotheses in relation to probabilities a priori | Mathematical Proceedings of the Cambridge Philosophical Society | Cambridge Core ». <https://www.cambridge.org/core/journals/mathematical-proceedings-of-the-cambridge->

- philosophical-society/article/testing-of-statistical-hypotheses-in-relation-to-probabilities-a-priori/65C6E3D534996282114D4E16FCA3E73C (consulté le oct. 04, 2021).
- [280] J. Neyman, « On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection », *J. R. Stat. Soc.*, vol. 97, n° 4, p. 558-625, 1934, doi: 10.2307/2342192.
- [281] I. J. GOOD, « Studies in the History of Probability and Statistics. XXXVII A. M. Turing's statistical work in World War II », *Biometrika*, vol. 66, n° 2, p. 393-396, août 1979, doi: 10.1093/biomet/66.2.393.
- [282] R. Doll et A. B. Hill, « Smoking and carcinoma of the lung; preliminary report », *Br. Med. J.*, vol. 2, n° 4682, p. 739-748, sept. 1950, doi: 10.1136/bmj.2.4682.739.
- [283] E. L. Kaplan et P. Meier, « Nonparametric Estimation from Incomplete Observations », *J. Am. Stat. Assoc.*, vol. 53, n° 282, p. 457-481, 1958, doi: 10.2307/2281868.
- [284] H. B. Mann et D. R. Whitney, « On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other », *Ann. Math. Stat.*, vol. 18, n° 1, p. 50-60, mars 1947, doi: 10.1214/aoms/1177730491.
- [285] F. Wilcoxon, « Individual Comparisons by Ranking Methods », *Biom. Bull.*, vol. 1, n° 6, p. 80-83, 1945, doi: 10.2307/3001968.
- [286] J. W. Tukey, *Exploratory data analysis*. Reading, Mass.: Addison-Wesley Pub. Co., 1977.
- [287] « Edward Tufte: Books - The Visual Display of Quantitative Information ». https://www.edwardtufte.com/tufte/books_vdqi (consulté le oct. 04, 2021).
- [288] « Over 16 years of R Project history (Revolutions) ». <https://blog.revolutionanalytics.com/2016/03/16-years-of-r-history.html> (consulté le oct. 04, 2021).
- [289] « I. Mathematical contributions to the theory of evolution. —XI. On the influence of natural selection on the variability and correlation of organs | Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character ». <https://royalsocietypublishing.org/doi/10.1098/rsta.1903.0001> (consulté le oct. 04, 2021).
- [290] K. Pearson, « LIII. On lines and planes of closest fit to systems of points in space », *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, vol. 2, n° 11, p. 559-572, nov. 1901, doi: 10.1080/14786440109462720.
- [291] H. Hotelling, « Analysis of a complex of statistical variables into principal components », *J. Educ. Psychol.*, vol. 24, n° 6, p. 417-441, 1933, doi: 10.1037/h0071325.
- [292] J. P. Benzécri, « Histoire et préhistoire de l'analyse des données. Partie V L'analyse des correspondances », *Cah. Anal. Données*, vol. 2, n° 1, p. 9-40, 1977.
- [293] « Multidimensional quantification: With the applications to analysis of social phenomena ». <https://scienceon.kisti.re.kr/srch/selectPORSrchArticle.do?cn=NART60073215> (consulté le oct. 04, 2021).
- [294] S. NISHISATO, *Analysis of Categorical Data: Dual Scaling and its Applications*. University of Toronto Press, 1980. Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: <https://www.jstor.org/stable/10.3138/j.ctvcj2v1m>
- [295] B. Escofier et J. Pagès, *Analyses factorielles simples et multiples. Objectifs méthodes et interprétation*. Dunod, 2008. Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: <https://hal.archives-ouvertes.fr/hal-00382085>
- [296] B. Escofier, « Traitement simultané de variables qualitatives et quantitatives en analyse factorielle », *Cah. Anal. Données*, vol. 4, n° 2, p. 137-146, 1979.
- [297] « Details - Caroli Linnaei...Systema naturae per regna tria naturae :secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. - Biodiversity

- Heritage Library ». <https://www.biodiversitylibrary.org/bibliography/542> (consulté le oct. 04, 2021).
- [298] A. P. Dempster, N. M. Laird, et D. B. Rubin, « Maximum Likelihood from Incomplete Data via the EM Algorithm », *J. R. Stat. Soc. Ser. B Methodol.*, vol. 39, n° 1, p. 1-38, 1977.
- [299] J. Sander, M. Ester, H.-P. Kriegel, et X. Xu, « Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications », *Data Min. Knowl. Discov.*, vol. 2, n° 2, p. 169-194, juin 1998, doi: 10.1023/A:1009745219419.
- [300] X. Jin et J. Han, « K-Means Clustering », in *Encyclopedia of Machine Learning*, C. Sammut et G. I. Webb, Éd. Boston, MA: Springer US, 2010, p. 563-564. doi: 10.1007/978-0-387-30164-8_425.
- [301] « Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification | BibSonomy ». <https://www.bibsonomy.org/bibtex/c86383cba8cfe00d5e6ef200016aca3f> (consulté le oct. 04, 2021).
- [302] F. Husson, J. Josse, et J. Pages, « Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? », p. 17.
- [303] « ESCOUFIER DATA SCIENCE AND ITS APPLICATIONS by Yves Escoufier, Chikio Hayashi, Bernard Fichet | 9780122417702 | Get Textbooks | New Textbooks | Used Textbooks | College Textbooks - GetTextbooks.com ». <https://www.gettextbooks.com/isbn/9780122417702/> (consulté le oct. 04, 2021).
- [304] « Mechanical Intelligence, Volume 1 - 1st Edition ». <https://www.elsevier.com/books/mechanical-intelligence/ince/978-0-444-88058-1> (consulté le oct. 04, 2021).
- [305] S. J. Russell et P. Norvig, *Artificial intelligence: a modern approach*. Upper Saddle River: Prentice Hall, 1995.
- [306] « Comment le « deep learning » révolutionne l'intelligence artificielle », *Le Monde.fr*, juill. 24, 2015. Consulté le: oct. 04, 2021. [En ligne]. Disponible sur: https://www.lemonde.fr/pixels/article/2015/07/24/comment-le-deep-learning-revolutionne-l-intelligence-artificielle_4695929_4408996.html
- [307] www.bibliopolis.com, « Ueber die chemische Zusammensetzung der Eiterzellen IN: Hoppe-Seyler ed Medicinsich-chemische Untersuchungen, Heft 4, 441-60 by Miescher on Jeremy Norman & Co., Inc », *Jeremy Norman & Co., Inc*. <https://www.jnorman.com/pages/books/42605/miescher/ueber-die-chemische-zusammensetzung-der-eiterzellen-in-hoppe-seyler-ed-medicinsich-chemische> (consulté le mai 07, 2021).
- [308] J. D. Watson et F. H. Crick, « Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid », *Nature*, vol. 171, n° 4356, p. 737-738, avr. 1953, doi: 10.1038/171737a0.
- [309] K. Hall, « William Astbury and the biological significance of nucleic acids, 1938-1951 », *Stud. Hist. Philos. Biol. Biomed. Sci.*, vol. 42, n° 2, p. 119-128, juin 2011, doi: 10.1016/j.shpsc.2010.11.018.
- [310] F. Crick, « Central dogma of molecular biology », *Nature*, vol. 227, n° 5258, p. 561-563, août 1970, doi: 10.1038/227561a0.
- [311] B. Whyte, « Maurice Wilkins: the third man of the double helix », *Drug Discov. Today*, vol. 9, n° 22, p. 956-957, nov. 2004, doi: 10.1016/S1359-6446(04)03286-6.
- [312] E. Noor, S. Cherkaoui, et U. Sauer, « Biological insights through omics data integration », *Curr. Opin. Syst. Biol.*, vol. 15, p. 39-47, juin 2019, doi: 10.1016/j.coisb.2019.03.007.

- [313] R. C. Gentleman *et al.*, « Bioconductor: open software development for computational biology and bioinformatics », *Genome Biol.*, vol. 5, n° 10, p. R80, 2004, doi: 10.1186/gb-2004-5-10-r80.
- [314] R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, et A. Regev, « Spatial reconstruction of single-cell gene expression data », *Nat. Biotechnol.*, vol. 33, n° 5, p. 495-502, mai 2015, doi: 10.1038/nbt.3192.
- [315] S. G. Rodriques *et al.*, « Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution », *Science*, vol. 363, n° 6434, p. 1463-1467, mars 2019, doi: 10.1126/science.aaw1219.
- [316] G. Wang, J. R. Moffitt, et X. Zhuang, « Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy », *Sci. Rep.*, vol. 8, n° 1, p. 4847, mars 2018, doi: 10.1038/s41598-018-22297-7.
- [317] Y. Zhou, E. Jia, M. Pan, X. Zhao, et Q. Ge, « Encoding Method of Single-cell Spatial Transcriptomics Sequencing », *Int. J. Biol. Sci.*, vol. 16, n° 14, p. 2663-2674, 2020, doi: 10.7150/ijbs.43887.
- [318] C. Baccin *et al.*, « Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization », *Nat. Cell Biol.*, vol. 22, n° 1, p. 38-48, janv. 2020, doi: 10.1038/s41556-019-0439-6.
- [319] M. Pavkovic, B. Riefke, K. Gutberlet, M. Raschke, et H. Ellinger-Ziegelbauer, « Comparison of the MesoScale Discovery and Luminex multiplex platforms for measurement of urinary biomarkers in a cisplatin rat kidney injury model », *J. Pharmacol. Toxicol. Methods*, vol. 69, n° 2, p. 196-204, avr. 2014, doi: 10.1016/j.vascn.2013.11.003.
- [320] S. Sourial, M. Marcusson-Ståhl, et K. Cederbrant, « Meso Scale Discovery and Luminex Comparative Analysis of Calbindin D28K », *J. Biomed. Biotechnol.*, vol. 2009, p. 187426, 2009, doi: 10.1155/2009/187426.
- [321] M. Martin, « Cutadapt removes adapter sequences from high-throughput sequencing reads », *EMBnet.journal*, vol. 17, n° 1, Art. n° 1, mai 2011, doi: 10.14806/ej.17.1.200.
- [322] C. Otto, P. F. Stadler, et S. Hoffmann, « Lacking alignments? The next-generation sequencing mapper segemehl revisited », *Bioinforma. Oxf. Engl.*, vol. 30, n° 13, p. 1837-1843, juill. 2014, doi: 10.1093/bioinformatics/btu146.
- [323] C. Trapnell, L. Pachter, et S. L. Salzberg, « TopHat: discovering splice junctions with RNA-Seq », *Bioinforma. Oxf. Engl.*, vol. 25, n° 9, p. 1105-1111, mai 2009, doi: 10.1093/bioinformatics/btp120.
- [324] B. Langmead et S. L. Salzberg, « Fast gapped-read alignment with Bowtie 2 », *Nat. Methods*, vol. 9, n° 4, p. 357-359, mars 2012, doi: 10.1038/nmeth.1923.
- [325] A. Dobin *et al.*, « STAR: ultrafast universal RNA-seq aligner », *Bioinforma. Oxf. Engl.*, vol. 29, n° 1, p. 15-21, janv. 2013, doi: 10.1093/bioinformatics/bts635.
- [326] D. Kim, B. Langmead, et S. L. Salzberg, « HISAT: a fast spliced aligner with low memory requirements », *Nat. Methods*, vol. 12, n° 4, Art. n° 4, avr. 2015, doi: 10.1038/nmeth.3317.
- [327] N. L. Bray, H. Pimentel, P. Melsted, et L. Pachter, « Near-optimal probabilistic RNA-seq quantification », *Nat. Biotechnol.*, vol. 34, n° 5, Art. n° 5, mai 2016, doi: 10.1038/nbt.3519.
- [328] R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, et C. Kingsford, « Salmon provides fast and bias-aware quantification of transcript expression », *Nat. Methods*, vol. 14, n° 4, p. 417-419, avr. 2017, doi: 10.1038/nmeth.4197.
- [329] H. Li *et al.*, « The Sequence Alignment/Map format and SAMtools », *Bioinforma. Oxf. Engl.*, vol. 25, n° 16, p. 2078-2079, août 2009, doi: 10.1093/bioinformatics/btp352.

- [330] D. Kim, J. M. Paggi, C. Park, C. Bennett, et S. L. Salzberg, « Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype », *Nat. Biotechnol.*, vol. 37, n° 8, p. 907-915, août 2019, doi: 10.1038/s41587-019-0201-4.
- [331] D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, et S. L. Salzberg, « TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions », *Genome Biol.*, vol. 14, n° 4, p. R36, avr. 2013, doi: 10.1186/gb-2013-14-4-r36.
- [332] B. Langmead, C. Trapnell, M. Pop, et S. L. Salzberg, « Ultrafast and memory-efficient alignment of short DNA sequences to the human genome », *Genome Biol.*, vol. 10, n° 3, p. R25, 2009, doi: 10.1186/gb-2009-10-3-r25.
- [333] H. Li et R. Durbin, « Fast and accurate short read alignment with Burrows-Wheeler transform », *Bioinforma. Oxf. Engl.*, vol. 25, n° 14, p. 1754-1760, juill. 2009, doi: 10.1093/bioinformatics/btp324.
- [334] C. Trapnell *et al.*, « Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation », *Nat. Biotechnol.*, vol. 28, n° 5, p. 511-515, mai 2010, doi: 10.1038/nbt.1621.
- [335] M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, et S. L. Salzberg, « StringTie enables improved reconstruction of a transcriptome from RNA-seq reads », *Nat. Biotechnol.*, vol. 33, n° 3, p. 290-295, mars 2015, doi: 10.1038/nbt.3122.
- [336] S. Anders, P. T. Pyl, et W. Huber, « HTSeq--a Python framework to work with high-throughput sequencing data », *Bioinforma. Oxf. Engl.*, vol. 31, n° 2, p. 166-169, janv. 2015, doi: 10.1093/bioinformatics/btu638.
- [337] B. Li et C. N. Dewey, « RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome », *BMC Bioinformatics*, vol. 12, n° 1, p. 323, août 2011, doi: 10.1186/1471-2105-12-323.
- [338] Y. Liao, G. K. Smyth, et W. Shi, « featureCounts: an efficient general purpose program for assigning sequence reads to genomic features », *Bioinforma. Oxf. Engl.*, vol. 30, n° 7, p. 923-930, avr. 2014, doi: 10.1093/bioinformatics/btt656.
- [339] M. Teng *et al.*, « A benchmark for RNA-seq quantification pipelines », *Genome Biol.*, vol. 17, n° 1, p. 74, avr. 2016, doi: 10.1186/s13059-016-0940-1.
- [340] C. Zhang, B. Zhang, L.-L. Lin, et S. Zhao, « Evaluation and comparison of computational tools for RNA-seq isoform quantification », *BMC Genomics*, vol. 18, n° 1, p. 583, août 2017, doi: 10.1186/s12864-017-4002-1.
- [341] Raivo Kolde, *Pheatmap : Pretty Heatmaps*. 2019.
- [342] S. Anders et W. Huber, « Differential expression analysis for sequence count data », *Genome Biol.*, vol. 11, n° 10, p. R106, 2010, doi: 10.1186/gb-2010-11-10-r106.
- [343] M. I. Love, W. Huber, et S. Anders, « Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 », *Genome Biol.*, vol. 15, n° 12, déc. 2014, doi: 10.1186/s13059-014-0550-8.
- [344] M. D. Robinson, D. J. McCarthy, et G. K. Smyth, « edgeR: a Bioconductor package for differential expression analysis of digital gene expression data », *Bioinforma. Oxf. Engl.*, vol. 26, n° 1, p. 139-140, janv. 2010, doi: 10.1093/bioinformatics/btp616.
- [345] M. E. Ritchie *et al.*, « limma powers differential expression analyses for RNA-sequencing and microarray studies », *Nucleic Acids Res.*, vol. 43, n° 7, p. e47, avr. 2015, doi: 10.1093/nar/gkv007.
- [346] N. J. Schurch *et al.*, « How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? », *RNA N. Y. N.*, vol. 22, n° 6, p. 839-851, juin 2016, doi: 10.1261/rna.053959.115.
- [347] M.-A. Dillies *et al.*, « A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis », *Brief. Bioinform.*, vol. 14, n° 6, p. 671-683, nov. 2013, doi: 10.1093/bib/bbs046.

- [348] H. Pimentel, N. L. Bray, S. Puente, P. Melsted, et L. Pachter, « Differential analysis of RNA-seq incorporating quantification uncertainty », *Nat. Methods*, vol. 14, n° 7, p. 687-690, juill. 2017, doi: 10.1038/nmeth.4324.
- [349] J. Costa-Silva, D. Domingues, et F. M. Lopes, « RNA-Seq differential expression analysis: An extended review and a software tool », *PloS One*, vol. 12, n° 12, p. e0190152, 2017, doi: 10.1371/journal.pone.0190152.
- [350] M. I. Love, S. Anders, V. Kim, et W. Huber, « RNA-Seq workflow: gene-level exploratory analysis and differential expression », *F1000Research*, vol. 4, p. 1070, 2015, doi: 10.12688/f1000research.7035.1.
- [351] H. Sanz *et al.*, « drLumi: An open-source package to manage data, calibrate, and conduct quality control of multiplex bead-based immunoassays data analysis », *PLOS ONE*, vol. 12, n° 11, p. e0187901, nov. 2017, doi: 10.1371/journal.pone.0187901.
- [352] E. J. Breen, W. Tan, et A. Khan, « The Statistical Value of Raw Fluorescence Signal in Luminex xMAP Based Multiplex Immunoassays », *Sci. Rep.*, vol. 6, n° 1, juill. 2016, doi: 10.1038/srep26996.
- [353] L. Cohen *et al.*, « A Modular Cytokine Analysis Method Reveals Novel Associations With Clinical Phenotypes and Identifies Sets of Co-signaling Cytokines Across Influenza Natural Infection Cohorts and Healthy Controls », *Front. Immunol.*, vol. 10, p. 1338, 2019, doi: 10.3389/fimmu.2019.01338.
- [354] N. Tierney, « visdat: Visualising Whole Data Frames », *J. Open Source Softw.*, vol. 2, n° 16, p. 355, août 2017, doi: 10.21105/joss.00355.
- [355] P. Qiu, « Embracing the dropouts in single-cell RNA-seq analysis », *Nat. Commun.*, vol. 11, p. 1169, mars 2020, doi: 10.1038/s41467-020-14976-9.
- [356] J. Josse et F. Husson, « missMDA: A Package for Handling Missing Values in Multivariate Data Analysis », *J. Stat. Softw.*, vol. 70, n° 1, Art. n° 1, avr. 2016, doi: 10.18637/jss.v070.i01.
- [357] Z. Zhang, « Multiple imputation with multivariate imputation by chained equation (MICE) package », *Ann. Transl. Med.*, vol. 4, n° 2, p. 30, janv. 2016, doi: 10.3978/j.issn.2305-5839.2015.12.63.
- [358] A. Kassambara, « Practical Guide to Principal Component Methods in R », p. 29.
- [359] G. Sturm *et al.*, « Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology », *Bioinformatics*, vol. 35, n° 14, p. i436-i445, juill. 2019, doi: 10.1093/bioinformatics/btz363.
- [360] F. Avila Cobos, J. Alquicira-Hernandez, J. E. Powell, P. Mestdagh, et K. De Preter, « Benchmarking of cell type deconvolution pipelines for transcriptomics data », *Nat. Commun.*, vol. 11, n° 1, p. 5650, nov. 2020, doi: 10.1038/s41467-020-19015-1.
- [361] H. Jin et Z. Liu, « A benchmark for RNA-seq deconvolution analysis under dynamic testing environments », *Genome Biol.*, vol. 22, n° 1, p. 102, avr. 2021, doi: 10.1186/s13059-021-02290-6.
- [362] B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, et A. A. Alizadeh, « Profiling tumor infiltrating immune cells with CIBERSORT », *Methods Mol. Biol. Clifton NJ*, vol. 1711, p. 243-259, 2018, doi: 10.1007/978-1-4939-7493-1_12.
- [363] J. Racle et D. Gfeller, « EPIC: A Tool to Estimate the Proportions of Different Cell Types from Bulk Gene Expression Data », *Methods Mol. Biol. Clifton NJ*, vol. 2120, p. 233-248, 2020, doi: 10.1007/978-1-0716-0327-7_17.
- [364] D. Aran, Z. Hu, et A. J. Butte, « xCell: digitally portraying the tissue cellular heterogeneity landscape », *Genome Biol.*, vol. 18, n° 1, déc. 2017, doi: 10.1186/s13059-017-1349-1.

- [365] E. Becht *et al.*, « Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression », *Genome Biol.*, vol. 17, n° 1, p. 218, oct. 2016, doi: 10.1186/s13059-016-1070-5.
- [366] F. Finotello *et al.*, « Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data », *Genome Med.*, vol. 11, n° 1, p. 34, mai 2019, doi: 10.1186/s13073-019-0638-6.
- [367] R. Du, V. Carey, et S. T. Weiss, « deconvSeq: deconvolution of cell mixture distribution in sequencing data », *Bioinforma. Oxf. Engl.*, vol. 35, n° 24, p. 5095-5102, déc. 2019, doi: 10.1093/bioinformatics/btz444.
- [368] K. Kang *et al.*, « CDSeq: A novel complete deconvolution method for dissecting heterogeneous samples using gene expression data », *PLOS Comput. Biol.*, vol. 15, n° 12, p. e1007510, déc. 2019, doi: 10.1371/journal.pcbi.1007510.
- [369] E. Birney *et al.*, « An Overview of Ensembl », *Genome Res.*, vol. 14, n° 5, p. 925-928, mai 2004, doi: 10.1101/gr.1860604.
- [370] G. R. Brown *et al.*, « Gene: a gene-centered information resource at NCBI », *Nucleic Acids Res.*, vol. 43, n° Database issue, p. D36-42, janv. 2015, doi: 10.1093/nar/gku1055.
- [371] D. Smedley *et al.*, « BioMart--biological queries made easy », *BMC Genomics*, vol. 10, p. 22, janv. 2009, doi: 10.1186/1471-2164-10-22.
- [372] R. J. Alexa A, *TopGO : Enrichment Analysis for Gene Ontology*.
- [373] G. Yu, L.-G. Wang, Y. Han, et Q.-Y. He, « clusterProfiler: an R package for comparing biological themes among gene clusters », *Omics J. Integr. Biol.*, vol. 16, n° 5, p. 284-287, mai 2012, doi: 10.1089/omi.2011.0118.
- [374] P. Fonseka, M. Pathan, S. V. Chitti, T. Kang, et S. Mathivanan, « FunRich enables enrichment analysis of OMICs datasets », *J. Mol. Biol.*, p. 166747, déc. 2020, doi: 10.1016/j.jmb.2020.166747.
- [375] « Gene Ontology Resource », *Gene Ontology Resource*. <http://geneontology.org/stats.html> (consulté le oct. 04, 2021).
- [376] A. Subramanian *et al.*, « Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles », *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, n° 43, p. 15545-15550, oct. 2005, doi: 10.1073/pnas.0506580102.
- [377] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, et J. P. Mesirov, « Molecular signatures database (MSigDB) 3.0 », *Bioinforma. Oxf. Engl.*, vol. 27, n° 12, p. 1739-1740, juin 2011, doi: 10.1093/bioinformatics/btr260.
- [378] B. Zhang, S. Kirov, et J. Snoddy, « WebGestalt: an integrated system for exploring gene sets in various biological contexts », *Nucleic Acids Res.*, vol. 33, n° Web Server issue, p. W741-748, juill. 2005, doi: 10.1093/nar/gki475.
- [379] D. W. Huang, B. T. Sherman, et R. A. Lempicki, « Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources », *Nat. Protoc.*, vol. 4, n° 1, p. 44-57, 2009, doi: 10.1038/nprot.2008.211.
- [380] J. Chen, E. E. Bardes, B. J. Aronow, et A. G. Jegga, « ToppGene Suite for gene list enrichment analysis and candidate gene prioritization », *Nucleic Acids Res.*, vol. 37, n° suppl_2, p. W305-W311, juill. 2009, doi: 10.1093/nar/gkp427.
- [381] L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, et H. Peterson, « gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler », *F1000Research*, vol. 9, p. ELIXIR-709, nov. 2020, doi: 10.12688/f1000research.24956.2.
- [382] F. Tian, D.-C. Yang, Y.-Q. Meng, J. Jin, et G. Gao, « PlantRegMap: charting functional regulatory maps in plants », *Nucleic Acids Res.*, vol. 48, n° D1, p. D1104-D1113, janv. 2020, doi: 10.1093/nar/gkz1020.

- [383] D. A. Barbie *et al.*, « Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1 », *Nature*, vol. 462, n° 7269, p. 108-112, nov. 2009, doi: 10.1038/nature08460.
- [384] D. Chaussabel *et al.*, « A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus », *Immunity*, vol. 29, n° 1, p. 150-164, juill. 2008, doi: 10.1016/j.immuni.2008.05.012.
- [385] S. Li *et al.*, « Molecular signatures of antibody responses derived from a systems biology study of five human vaccines », *Nat. Immunol.*, vol. 15, n° 2, p. 195-204, févr. 2014, doi: 10.1038/ni.2789.
- [386] January Weiner, *tmod: Feature Set Enrichment Analysis for Metabolomics and Transcriptomics*. 2020.
- [387] D. Chaussabel et N. Baldwin, « Democratizing systems immunology with modular transcriptional repertoire analyses », *Nat. Rev. Immunol.*, vol. 14, n° 4, p. 271-280, avr. 2014, doi: 10.1038/nri3642.
- [388] D. Rinchai *et al.*, « BloodGen3Module: Blood transcriptional module repertoire analysis and visualization using R », *Bioinforma. Oxf. Engl.*, févr. 2021, doi: 10.1093/bioinformatics/btab121.
- [389] A. Rawat *et al.*, « A Neutrophil-Driven Inflammatory Signature Characterizes the Blood Transcriptome Fingerprint of Psoriasis », *Front. Immunol.*, vol. 11, p. 587946, 2020, doi: 10.3389/fimmu.2020.587946.
- [390] « Pathguide: the pathway resource list ». <http://www.pathguide.org/> (consulté le oct. 04, 2021).
- [391] M. Kanehisa, « The KEGG database », *Novartis Found. Symp.*, vol. 247, p. 91-101; discussion 101-103, 119-128, 244-252, 2002.
- [392] D. N. Slenter *et al.*, « WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research », *Nucleic Acids Res.*, vol. 46, n° D1, p. D661-D667, janv. 2018, doi: 10.1093/nar/gkx1064.
- [393] H. Mi et P. Thomas, « PANTHER pathway: an ontology-based pathway database coupled with data analysis tools », *Methods Mol. Biol. Clifton NJ*, vol. 563, p. 123-140, 2009, doi: 10.1007/978-1-60761-175-2_7.
- [394] B. Jassal *et al.*, « The reactome pathway knowledgebase », *Nucleic Acids Res.*, vol. 48, n° D1, p. D498-D503, janv. 2020, doi: 10.1093/nar/gkz1031.
- [395] K. Breuer *et al.*, « InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation », *Nucleic Acids Res.*, vol. 41, n° Database issue, p. D1228-1233, janv. 2013, doi: 10.1093/nar/gks1147.
- [396] B. Zhang et S. Horvath, « A general framework for weighted gene co-expression network analysis », *Stat. Appl. Genet. Mol. Biol.*, vol. 4, p. Article17, 2005, doi: 10.2202/1544-6115.1128.
- [397] P. S. T. Russo *et al.*, « CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses », *BMC Bioinformatics*, vol. 19, n° 1, p. 56, févr. 2018, doi: 10.1186/s12859-018-2053-1.
- [398] P. Shannon *et al.*, « Cytoscape: a software environment for integrated models of biomolecular interaction networks », *Genome Res.*, vol. 13, n° 11, p. 2498-2504, nov. 2003, doi: 10.1101/gr.1239303.
- [399] A. Krämer, J. Green, J. Pollard, et S. Tugendreich, « Causal analysis approaches in Ingenuity Pathway Analysis », *Bioinforma. Oxf. Engl.*, vol. 30, n° 4, p. 523-530, févr. 2014, doi: 10.1093/bioinformatics/btt703.
- [400] F. Rohart, B. Gautier, A. Singh, et K.-A. Lê Cao, « mixOmics: An R package for 'omics feature selection and multiple data integration », *PLOS Comput. Biol.*, vol. 13, n° 11, p. e1005752, nov. 2017, doi: 10.1371/journal.pcbi.1005752.

- [401] P. Rainard et G. Foucras, « A Critical Appraisal of Probiotics for Mastitis Control », *Front. Vet. Sci.*, vol. 5, p. 251, 2018, doi: 10.3389/fvets.2018.00251.
- [402] H. Shin et A. Iwasaki, « A vaccine strategy that protects against genital herpes by establishing local memory T cells », *Nature*, vol. 491, n° 7424, Art. n° 7424, nov. 2012, doi: 10.1038/nature11522.
- [403] P. Roussel *et al.*, « Investigating the contribution of IL-17A and IL-17F to the host response during *Escherichia coli* mastitis », *Vet. Res.*, vol. 46, n° 1, p. 56, juin 2015, doi: 10.1186/s13567-015-0201-4.
- [404] P. Rainard *et al.*, « T Helper 17-Associated Cytokines Are Produced during Antigen-Specific Inflammation in the Mammary Gland », *PLoS ONE*, vol. 8, n° 5, p. e63471, mai 2013, doi: 10.1371/journal.pone.0063471.
- [405] A. El-Sayed et M. Kamel, « Bovine mastitis prevention and control in the post-antibiotic era », *Trop. Anim. Health Prod.*, vol. 53, n° 2, p. 236, mars 2021, doi: 10.1007/s11250-021-02680-9.
- [406] R. J. Erskine, « Vaccination strategies for mastitis », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 28, n° 2, p. 257-270, juill. 2012, doi: 10.1016/j.cvfa.2012.03.002.
- [407] « Un vaccin 100% français administrable par voie nasale, contre la COVID-19 : résultats pré-cliniques positifs », *INRAE Institutionnel*. <https://www.inrae.fr/actualites/vaccin-100-francais-administrable-voie-nasale-contre-covid-19-resultats-pre-cliniques-positifs> (consulté le oct. 07, 2021).
- [408] J. T. Richeson et T. R. Falkner, « Bovine Respiratory Disease Vaccination: What Is the Effect of Timing? », *Vet. Clin. North Am. Food Anim. Pract.*, vol. 36, n° 2, p. 473-485, juill. 2020, doi: 10.1016/j.cvfa.2020.03.013.
- [409] T.-M. Nguyen, A. Shafi, T. Nguyen, et S. Draghici, « Identifying significantly impacted pathways: a comprehensive review and assessment », *Genome Biol.*, vol. 20, n° 1, p. 203, oct. 2019, doi: 10.1186/s13059-019-1790-4.
- [410] R. Henkel, R. Hoehndorf, T. Kacprowski, C. Knüpfer, W. Liebermeister, et D. Waltemath, « Notions of similarity for systems biology models », *Brief. Bioinform.*, vol. 19, n° 1, p. 77-88, janv. 2018, doi: 10.1093/bib/bbw090.
- [411] B. A. Kidd, L. A. Peters, E. E. Schadt, et J. T. Dudley, « Unifying immunology with informatics and multiscale biology », *Nat. Immunol.*, vol. 15, n° 2, p. 118-127, févr. 2014, doi: 10.1038/ni.2787.
- [412] W. Zhou *et al.*, « Longitudinal multi-omics of host-microbe dynamics in prediabetes », *Nature*, vol. 569, n° 7758, p. 663-671, mai 2019, doi: 10.1038/s41586-019-1236-x.
- [413] S. Li *et al.*, « Metabolic Phenotypes of Response to Vaccination in Humans », *Cell*, vol. 169, n° 5, p. 862-877.e17, mai 2017, doi: 10.1016/j.cell.2017.04.026.
- [414] C. Meng, B. Kuster, A. C. Culhane, et A. M. Gholami, « A multivariate approach to the integration of multi-omics datasets », *BMC Bioinformatics*, vol. 15, p. 162, mai 2014, doi: 10.1186/1471-2105-15-162.
- [415] B. Pulendran et M. M. Davis, « The science and medicine of human immunology », *Science*, vol. 369, n° 6511, p. eaay4014, sept. 2020, doi: 10.1126/science.aay4014.
- [416] F. R. Pinu *et al.*, « Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community », *Metabolites*, vol. 9, n° 4, p. E76, avr. 2019, doi: 10.3390/metabo9040076.
- [417] F. Enjalbert, P. Lebreton, et O. Salat, « Effects of copper, zinc and selenium status on performance and health in commercial dairy and beef herds: Retrospective study », *J. Anim. Physiol. Anim. Nutr.*, vol. 90, n° 11-12, p. 459-466, déc. 2006, doi: 10.1111/j.1439-0396.2006.00627.x.

5. Productions scientifiques

5.1. Publications

Nathan Cebron, Sarah Maman, Sarah Walachowski, Blandine Gausserès, Patricia Cunha *et al.* **Th17-related mammary immunity, but not a high systemic Th1 immune response is associated with protection against E. coli mastitis** *NPJ vaccines*, Nature Research 2020, 5 (1), pp.108. (10.1038/s41541-020-00258-4)

Pierre Bessièrre, Marine Wasniewski, Evelyne Picard-Meyer, Alexandre Servat, Thomas Figueroa *et al.* **Intranasal type I interferon treatment is beneficial only when administered before clinical signs onset in the SARS-CoV-2 hamster model**

PLoS Pathogens, Public Library of Science, 2021, 17 (8), pp.e1009427. (10.1371/journal.ppat.1009427)

J. Lesueur, N. Cebron, S. Walachowski, F. Corbiere, S. Barbey, F. Launay, R. Lefebvre, D. Boichard, P. Germon, G. Foucras **Cytokine profiling upon whole blood stimulation with bacteria and TLR ligands reveals commonalities and diversity of the response (En cours de rédaction, Frontiers in immunology)**

Elise Vanbergue, Sébastien Assié, Béatrice Mounaix, Marlène Guiadeur, Anne Aupiais, Fabrice Robert, Didier Andrieu, Olivier Devloo, James Quentin, Nathan Cebron, Gilles Meyer, Aurore Philibert, Renaud Maillard, Gilles Foucras **Preconditioning programs: a solution to reduce bovine respiratory diseases in fattening operations in a French context? (En cours de rédaction Animal)**

Nathan Cebron, Blandine Gausserès, Julie Cournet, Christian Tasca, Clément Boudenne, Lucie Dutheil, Elise Vanbergue, and Gilles Foucras **Integrative analysis of vaccination against bovine respiratory disease in a preconditioning program for beef cattle (En cours de rédaction, Vaccine)**

5.2. Communications orales

Nathan Cebron, Sarah Maman, Céline Chantry-Darmon, Pascal Rainard, Gilles Foucras. **Functional and RNA-seq analyses provide insights into the protective response mechanisms upon immunization against bovine E. coli mastitis**

11. *Symposium of the French Domestic Animal Immunology Network (IAD 2018)*, Mar 2018, Tours, France. 56 p

Nathan Cebron, Sarah Maman Haddad, Pascal Rainard, Gilles Foucras. **Local Th17 immunity upon mammary immunization is protective against E. coli mastitis**

12. *International Veterinary Immunology Symposium (IVIS)*, Aug 2019, Seattle, United States. 176 p

Jérémy Lesueur, Sarah Barbey, Nathan Cebron, Sarah Walachowski, Rachel Lefebvre *et al.* **Development of a bovine/ovine cytokine 15-plex assay for immunoprofiling of the cellular response in ruminants**

12. *International Veterinary Immunology Symposium*, Aug 2019, Seattle, United States. 176 p

Nathan Cebron, Sarah Maman, Céline Chantry-Darmon, Pascal Rainard, Gilles Foucras. **Analyses transcriptomiques de la protection permise par la vaccination contre les mammites à E. coli** *Glande mammaire, Lait*

5.3. Posters

Gilles Foucras, Nathan Cebron, Sarah Maman, Céline Chantry-Darmon, Pascal Rainard. **Looking for correlates of protection against bovine *Escherichia coli* mastitis: the local recruitment of Th17 cells upon mammary immunization**

6. *European Veterinary Immunology Workshop (EVIW)*, Sep 2018, Utrecht, Netherlands. 2018