



**HAL**  
open science

# **Apport des outils bio-informatiques à l'étude de pathologies humaines par l'analyse de réseaux moléculaires et à l'amélioration des pratiques en génétique médicale**

Jean-Marie Ravel

## **► To cite this version:**

Jean-Marie Ravel. Apport des outils bio-informatiques à l'étude de pathologies humaines par l'analyse de réseaux moléculaires et à l'amélioration des pratiques en génétique médicale. Médecine humaine et pathologie. Université de Lorraine, 2023. Français. ⟨NNT : 2023LORR0171⟩. ⟨tel-04504734⟩

**HAL Id: tel-04504734**

**<https://hal.univ-lorraine.fr/tel-04504734v1>**

Submitted on 17 Dec 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization



**UNIVERSITÉ  
DE LORRAINE**

**BIBLIOTHÈQUES  
UNIVERSITAIRES**

## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)  
*(Cette adresse ne permet pas de contacter les auteurs)*

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



UNIVERSITÉ  
DE LORRAINE

BioSE



ECOLE DOCTORALE BioSE  
BIOLOGIE - SANTÉ - ENVIRONNEMENT  
ED 266

Ecole Doctorale BioSE (Biologie – Santé – Environnement)

## Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE de LORRAINE

Mention : « Sciences de la Vie et de la Santé »

par Jean-Marie Ravel

Apport des outils bio-informatiques à l'étude de pathologies  
humaines par l'analyse de réseaux moléculaires et à  
l'amélioration des pratiques en génétique médicale.

Vendredi 13 octobre 2023

### Membres du Jury

Directrices de thèse : Dr Inna Kuperstein, MCU-PH, CR, Inserm, U900, Institut Curie  
Dr Céline Bonnet, UMR Inserm 1256 – NGERE Université de Lorraine

Rapporteurs : Dr Anna Niaraki, PU, Université de Toulouse III Paul Sabatier  
Pr. Jean-Christophe Thalabard, PU-PH émérite, Univ. Paris Centre

Examineurs : Dr Laurence Calzone, CR, Inserm, U900, Institut Curie  
Pr. David Meyre, PU-PH, UMR Inserm 1256 NGERE  
Université de Lorraine  
Dr Carine Bossenmeyer-Pourié, PU, UMR Inserm 1256 NGERE  
Université de Lorraine  
Dr Kévin Yauy, CCA, Génétique clinique, Université de Montpellier

Membre invité : Dr Inna Kuperstein, CR, Inserm, U900, Institut Curie

---

Institut Curie, PSL Research University, Mines Paris Tech, Inserm, U900, 75005  
Paris, France

UMR Inserm 1256 – NGERE – Faculté de médecine, 9 Avenue de la Forêt de  
Haye – CD 50184 – 54505 Vandœuvre-lès-Nancy Cedex, France

Laboratoire de génétique médicale, CHRU-Nancy, F-54000 Nancy, France

*La révolution informatique fait gagner un temps fou aux  
hommes, mais ils le passent avec leur ordinateur !*

---

Khalil Assala

## Remerciements

Ce travail a été effectué conjointement au sein de l'équipe de biologie des systèmes de l'U900 à l'Institut Curie et au sein des équipes de génétique du CHRU de Nancy et de l'unité NGERE à l'université de Lorraine. Je tiens donc à remercier en premier lieu Monsieur le Directeur de Recherche Emmanuel Barillot, Monsieur le Professeur Guéant, Mesdames les Docteurs Laëtitia Lambert et Céline Bonnet pour leur accueil et leur soutien.

Je tiens à remercier très vivement Monsieur le Professeur Jean-Christophe Thalabard et Madame Anna Niaraki d'avoir accepté d'être rapporteur de ce travail ainsi que pour l'analyse et la critique de ce mémoire. J'exprime également ma gratitude à Madame le Professeur Carine Pourrié Bonsemeyer et Monsieur le Professeur David Meyre d'avoir accepté d'être examinateurs de cette thèse. Je remercie également Monsieur le Docteur Kevin Yaou, avec lequel je partage une partie de mon cursus et de ma passion, qui a accepté d'évaluer mon travail aujourd'hui.

Merci également à Monsieur le Professeur Stéphane Zuily et Madame le Docteur Juliette Albuissou, membres du comité de suivi de cette thèse, qui m'ont soutenu et m'ont partagé leurs connaissances pratiques et théoriques. Leur aide, leurs remarques et leurs critiques m'ont beaucoup aidé.

A mes directrices de thèse, Madame le Docteur Inna Kuperstein et Madame le Docteur Céline Bonnet. Merci, Inna, d'avoir accepté de poursuivre les travaux que nous avons initiés il y a plus de sept ans et de m'avoir accordé la possibilité de débiter cette thèse. Merci, Céline, de m'avoir accordé ta confiance, de m'avoir permis de libérer le temps nécessaire à ce travail dans le cadre de mon cursus hospitalo-universitaire. Je te remercie également pour tes commentaires toujours précis et pertinents.

A Madame le Docteur Laurence Calzone : chère Laurence, un simple mot de remerciement ne suffira pas à exprimer ma gratitude pour cette dernière année. Les discussions qui ont émergées avec Monsieur Vincent Noël et Madame Loredana

Martignetti après ce séjour de cohésion à St Malo ont été revitalisant. J'associe tout naturellement les membres de l'équipe U900, présents et passés, et en particulier Vincent, Loredana, Christobal, Andreï, Maria pour leur aide mais également Kati et Caroline pour avoir résolu tous mes déboires logistiques.

Il me tient à cœur de remercier les membres du laboratoire de génétique médicale et du service de génétique clinique du CHRU de Nancy. Le « projet GeneTree » a été initié sous la direction de Monsieur le Professeur Leheup, qui m'a en particulier initié à l'oncogénétique, puis a été soutenu par Madame le Docteur Laëtitia Lambert. Il a très largement bénéficié du retour de plusieurs personnes et en particulier Madame Axelle Rivière, conseillère en génétique, et Madame le Docteur Marion Imbert Bouteille du CLCC de Montpellier. Je remercie également les membres du laboratoire de génétique médicale de leur bienveillance qui m'a permis de me libérer du temps de recherche et de leurs remarques lors des diverses réunions de présentation.

Cette thèse conclut la formation initiée en 2010 avec mon admission au double cursus Médecine Sciences au sein de l'université Paris Descartes. Ce travail se situe dans la continuité de l'élan translationnel donné à ce double cursus initié par Messieurs les Professeurs Jean-Claude Chottard, Philippe Ascher (décédé le 4 octobre 2022), François Amblard et Jean-Christophe Thalabard que je remercie très vivement. C'est donc un réel honneur pour moi d'être examiné ce jour par Monsieur le Professeur Thalabard.

J'associe à ce travail le Docteur Mathilde Renaud, qui a bien voulu me guider dans mon parcours hospitalo-universitaire dès son arrivée au CHRU de Nancy.

Je remercie également mon père, le Docteur Aimé Ravel pour sa relecture attentive de ce manuscrit. Cher Papa, je tiens à t'associer particulièrement à ce travail puisque j'ai probablement « hérité » de toi de cette passion pour la génétique et la recherche. Merci également à toi, chère Maman, car vous m'avez tous deux offert un cadre propice à poursuivre ces longues études.

Merci, chère Manon, pour ton aide précieuse au quotidien ! Merci pour ces 7 premières années de bonheur et les 63 (contractuellement au minimum) à venir ! Tu as toujours été un soutien indéfectible et une aide précieuse aussi bien sur le plan scientifique que logistique, relationnel ou moral 😊.

A mes fils, Philippe et Xavier, qui n'ont pas forcément compris pourquoi, parfois, j'étais grognon !

# Table des matières

Membres du Jury.....	1
Remerciements .....	3
Liste des Figures.....	9
Liste des tableaux.....	11
Liste des abréviations .....	12
Préambule .....	14
1 Introduction.....	15
1. Les grandes lignes de l'évolution de la génétique médicale .....	15
a. La génétique médicale : des maladies pédiatriques rares aux maladies communes de l'adulte .....	15
b. L'émergence du séquençage haut-débit.....	15
c. Omiques : application des données de séquençage haut-débit .....	16
2. Biologie des systèmes et compréhension des maladies complexes .....	18
a. Maladies complexes.....	18
b. Biologie des systèmes .....	20
c. Cartes moléculaires maladie-spécifiques.....	22
d. L'analyse de voies de signalisation .....	25
e. ROMA .....	27
Objectifs de thèses.....	29
2 GeneTree.....	30
1. Introduction .....	30
a. La génétique oncologique .....	30
b. Déroulement de la consultation d'oncogénétique .....	31
c. Problématique.....	32
2. Résultats .....	34
a. Présentation générale .....	34
b. Impacts.....	36
2. Discussion et perspectives.....	37
3 La mort cellulaire régulée .....	40
1. Contexte.....	40
2. Introduction .....	42
3. Matériels et méthodes.....	45
a. Construction de la carte.....	45
b. Règles de sélection de la littérature.....	46
c. Format d'annotation de la carte.....	47

d.	Génération de la carte NaviCell avec NaviCell Factory.....	48
e.	Navigation dans la carte .....	49
4.	Présentation de la carte.....	51
a.	Caractéristiques générales de la carte.....	51
b.	Structure de la carte de la mort cellulaire régulée .....	51
c.	Contenu de la carte du de mort cellulaire régulée .....	53
d.	L'interaction entre les modes de mort cellulaire et les points de commutation .....	58
e.	Comparaison de la carte RCD avec des ressources de voies similaires .....	61
5.	Première application : Alzheimer et cancer du poumon .....	64
6.	Deuxième application : sous types de cancer de l'ovaire .....	71
7.	Discussion.....	73
3.	Analyse de réseaux moléculaires impliqués dans la covid-19.....	76
1.	Introduction .....	76
a.	Mécanismes généraux impliqués dans le covid-19 .....	76
b.	Contexte de ce projet .....	77
2.	Ressources et méthodes .....	82
a.	Les cartes issues de ACSN .....	82
b.	Données utilisées .....	83
c.	Analyse de l'expression différentielle d'expression de gènes (DGE) .....	85
d.	Analyse scRNAseq .....	90
e.	Pseudobulk .....	91
3.	Résultats préliminaires .....	92
a.	Les cellules immunitaires innées des voies respiratoires semblent avoir un profil d'expression génique différent dans les cellules covid-19.....	92
b.	Une méthode plus robuste.....	94
4.	Des disparités d'expression entre les différents types cellulaires .....	95
a.	Les macrophages .....	95
b.	Neutrophiles.....	98
c.	Cellules épithéliales squameuses .....	100
d.	Des mécanismes différents de la grippe ?.....	102
5.	Discussion.....	103
	Perspective : création d'un réseau d'inférence avec Pypath.....	104
4.	Conclusions et perspectives .....	105
	Le premier axe .....	105
	Le second axe .....	106
	Références.....	108

Annexes.....	126
1. Mode d'emploi de l'outils GeneTree.....	126
a. Présentation générale .....	126
b. Création d'une famille.....	127
c. Modification d'une famille au niveau du tableau .....	128
d. Prise en main de l'interface graphique de l'arbre généalogique .....	129
e. Export des différents fichiers.....	130
f. Génération du texte .....	131
g. Interface patient.....	131
Résumé.....	132
Abstract .....	134

## Liste des Figures

Figure 1. Différentes méthodes d'utilisation du séquençage haut-débit. ....	17
Figure 2. Génétique des maladies complexes.....	19
Figure 3. Illustrations de la notion de système biologique.....	21
Figure 4. Réseau de réseaux.....	22
Figure 5. Représentation d'un modèle simplifié en utilisant la notation SBGN. ....	24
Figure 6. Représentation des ensembles de gènes dans ROMA dans le cas de deux échantillons. ....	28
Figure 7. Représentation du temps consacré à la préparation et réalisation d'une consultation.....	31
Figure 8. Comparaison de GeneTree avec les principales solutions existantes. ....	33
Figure 9. Captures d'écrans de l'interface actuelle de l'application. ....	34
Figure 10. Interface dédiée au patient.....	36
Figure 11. Représentation du projet GeneTree. ....	38
Figure 12. Les caractéristiques du cancer, version 2022.....	41
Figure 13. Image du portail ACSN. ....	42
Figure 14. Programmes majeurs de mort cellulaire .....	43
Figure 15. Processus de construction de la carte de la mort cellulaire régulée.....	47
Figure 16. Page d'annotation d'une entité de la carte RCD au format NaviCell.....	48
Figure 17. Navigation dans la carte de la mort cellulaire régulée, zooms et annotations des entités. ....	50
Figure 18. Représentation simplifiée de l'apoptose.....	54
Figure 19. Nécroptose.....	55
Figure 20. Autophagie et mort cellulaire liée au lysosome.....	56
Figure 21. Parthanatose.....	57
Figure 22. Pyroptose (à gauche), ferroptose (au centre) et nécrose liée à la perméabilité mitochondriale (à droite).....	58
Figure 23. Interaction entre les modes de mort cellulaire régulée.....	59
Figure 24. Comparaison de la carte RCD avec d'autres ressources sur les voies.....	61
Figure 25. Visualisation des scores d'activité moyens des modules ROMA à l'aide de données d'expression provenant d'échantillons d'hippocampe de patient atteint de la maladie	

d'Alzheimer (AD) et d'échantillons de cancer du poumon (LC) non à petites cellules (CPNPC) dans la carte RCD.....	66
Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire .....	72
Figure 27. Covid-19 <i>Disease Map</i> . .....	78
Figure 28. Cartes spécifiques à un type de cellule pour les macrophages, cellules dendritiques et <i>Natural Killer</i> .....	79
Figure 29. Fonctionnement de Ailani.....	80
Figure 30. Analyse de la composition de la carte NK.....	81
Figure 31. Représentation schématique du flux de travail expérimental extraite de (Chua et al., 2020) .....	84
Figure 32. Analyse de l'expression différentielle des gènes (DGE).....	86
Figure 33. Impact de la profondeur de séquençage sur l'analyse DGE. ....	87
Figure 34. Impact de la longueur du gène sur l'analyse de DGE. ....	87
Figure 35. Impact de la composition de l'ARN sur l'analyse de la DGE. ....	88
Figure 36. Schéma d'un flux de travail typique d'une analyse RNA-seq de cellule unique extraite de (Luecken & Theis, 2019).....	90
Figure 37. Représentation de l'expression génique moyenne des cellules dendritiques, macrophages et natural killer. ....	92
Figure 38. Analyse ROMA au niveau des cellules macrophages non-résidents.....	93
Figure 39. Boîte à moustache des données d'expression après le processus initial de traitement des données pour les cellules Macrophages non-résidentes.....	94
Figure 40. Boîte à moustache des données d'expression après le processus définitif de traitement des données.....	94
Figure 41. Modules <i>overdispersed not shifted</i> dans les cellules macrophages non-résidentes. ....	97
Figure 42. Carte de chaleur de l'analyse ROMA des cellules Neutrophiles. ....	99
Figure 43. Carte de chaleur de l'analyse ROMA des cellules squameuses montrant les modules <i>Overdispersed not shifted</i> .....	101
Figure 44. Carte de chaleur de l'analyse ROMA des cellules squameuses montrant les modules <i>Overdispersed not shifted</i> en analysant les modules de la mort cellulaire régulée (Ravel et al., 2020).....	102
Figure 45. Analyse ROMA des macrophages M1, modules <i>shifted</i> . ....	103

Figure 46. Barre de navigation et section « tableau » de l'interface médecin.....	127
Figure 47. Menu déroulant 'Nouvelle Famille' .....	128
Figure 48. Interface Arbre généalogique.....	129
Figure 49. « Widgets » autour d'un individu.....	130

## Liste des tableaux

Tableau 1. Structure et contenu de la carte RCD.....	52
Tableau 2. Gènes contribuant le plus à chaque module pour les ensembles de données sur la maladie d'Alzheimer et le cancer du poumon. ....	68

## Liste des abréviations

ACSN : *Atlas of Cancer Signalling Network*  
ADN : acide désoxyribonucléique  
ADCD : mort cellulaire dépendante de l'autophagie  
Ailani : *Artificial Intelligence LANguage Interface*  
AMP : 5' adénosine monophosphate  
AMPK : kinase dépendante de l'AMP  
ARN : acide ribonucléique  
ATAC-seq : *Assay for Transposase-Accessible Chromatin with highthroughput sequencing*  
ATP : 5' adénosine triphosphate  
BOADICEA : *Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm*  
BRCA1 : *Breast cancer 1*  
CHRU : Centre hospitalier régional universitaire  
ChIP : Chromatin Immunoprecipitation Sequencing (ChIP-Seq)  
CLCC : Centre de Lutte Contre le Cancer  
C19DMap carte de la COVID-19 du *Disease Map*  
CPM : comptage par million  
DGE : expression différentielle d'expression de gènes  
GWAS : genome wide association study  
EBV : Virus d'Epstein-Barr  
EMT : *Epithelial-mesenchymal transition*, Transition épithélio-mésenchymateuse  
FCS : *Functional Class Scoring*  
GSEA : *gene set enrichment analysis*, analyse d'enrichissement de l'ensemble des gènes  
HDS, Hébergeur de Données de Santé  
HLA : *Human Leukocyte Antigen*  
HPO : *Human Phenotype Ontology*  
Hi-C-Seq : *high-throughput chromatin conformation capture*  
IAP : protéines inhibitrices de l'apoptose  
ICD : mort cellulaire immunogène  
JSON : *JavaScript Object Notation*  
KEGG : *Kyoto Encyclopedia of Genes and Genomes*  
LDCD : mort cellulaire dépendante du lysosome  
LPS : lipopolysaccharides  
MAMPs : microbe-associated molecular patterns  
MIRIAM *Minimal Information Required In the Annotation of biochemical Models*  
MPT : *mitochondrial permeability transition*, transition de perméabilité mitochondriale  
MSigDB : *Human Molecular Signatures Database*  
miARN : micro ARN  
MOMP : perméabilisation de la membrane externe des mitochondries  
MPT : transition de perméabilité des mitochondrie  
NET : *neutrophil extracellular trap*  
NK : *Natural Killer*  
NSCLC : *Non Small Cell Lung Cancer*, cancers du poumon non à petites cellules  
OMS : Organisation mondiale de la santé

OR : odd ratio  
PAR : poly (ADP-ribose) (PAR)  
PC1 : première composante principale (PC1)  
PCR : réaction en chaîne par polymérase  
PRS : scores de risque polygénique  
PTPC : *permeability transition complex*  
RCD : *regulated cell death*, mort cellulaire régulée  
RE : réticulum endoplasmique  
RGPD : Règlement Général sur la Protection des Données  
ROS : *reactive oxygen species*, espèces oxydantes réactives  
RNA Seq : séquençage de l'ARN  
ROMA : *Representation and quantification Of Module Activities*  
SAM : syndrome d'activation macrophagique  
SARS : syndrome respiratoire aigu sévère  
SARS-CoV-2 : coronavirus 2 du syndrome respiratoire aigu sévère  
SNP : *single nucleotide polymorphism*  
SBML : *Systems Biology Markup Language*  
SBGN : *Systems Biology Graphical Notation*  
scRNA seq : séquençage de l'ARN en de cellule unique  
SDRA : syndrome de détresse respiratoire aiguë  
SSGSEA : Single Sample GSEA  
TCA : *tricarboxylic acid cycle*, cycle de Kreps  
TMM : *trimmed mean of M values*  
WGS : *Whole Genome Sequencing*

## Préambule

Ce travail de thèse d'université repose sur deux axes de recherches qui sont complémentaires : l'un, intimement lié à ma pratique médicale au sein du CHRU de Nancy, université de Lorraine, et l'autre, plus fondamental dans le cadre d'un travail à l'Institut Curie, U900. Les travaux selon chacun de ces axes ont été conduits en parallèle. Ainsi la première partie de ce rapport de thèse va développer l'outil qui résulte de mon premier axe de recherche dénommé GeneTree. Celui-ci permet de résoudre un des problèmes récurrents en pratique clinique, la saisie multiple d'informations d'un patient, et permet également d'impliquer le patient dans sa prise en charge en faisant gagner du temps au clinicien. Le deuxième axe de recherche a abouti à deux travaux qui sont présentés ici dans l'ordre chronologique. Le premier travail de ce second axe (qui est donc le deuxième chapitre de cette thèse) avait pour but de compléter un projet en cours au sein de l'Institut Curie : ACSN (*Atlas of Cancer Signalling Network*) qui vise à représenter sous la forme d'une carte les processus moléculaires impliqués dans le cancer. Dans ce cadre, nous traiterons ici de la mort cellulaire programmée. Enfin, en utilisant la même approche de biologie des systèmes, le second travail qui constitue le dernier chapitre, débuté après la pandémie covid-19, a permis d'explorer certaines différences entre les processus cellulaires impliqués au sein des différents groupes de sévérité.

# 1 Introduction

## 1. Les grandes lignes de l'évolution de la génétique médicale

### a. La génétique médicale : des maladies pédiatriques rares aux maladies communes de l'adulte

L'exercice de la génétique s'est d'abord concentré sur l'analyse des maladies rares polymalformatives avec des phénotypes bien caractérisés (syndrome de trisomie 21, syndrome de Cornelia de Lange, etc.). Dans un deuxième temps, l'étude génétique de maladies, y compris les affections communes, s'est développée, par exemple, des familles présentant une forte prévalence de cancer ont été étudiées. En 1990, le gène *BRCA1* a ainsi été identifié par Mary-Claire King et son équipe et a été associé à un fort risque de cancer, en particulier de cancer du sein et de l'ovaire (Stoppa-Lyonnet & Jeanpierre, 2004). Cette découverte a largement ouvert la voie de la génétique oncologique (encore appelée « oncogénétique »).

### b. L'émergence du séquençage haut-débit

La publication de la séquence du génome humain en 2000 est probablement l'une des grandes avancées de la science contemporaine (Lander et al., 2001). Bien qu'incomplet et contenant de nombreuses erreurs, ce séquençage a significativement marqué l'avancée du Projet Génome Humain (*Human Genome Project*). Ce travail a produit des outils de recherche importants qui permettront aux chercheurs d'identifier les gènes impliqués dans la vie normale ainsi que dans les maladies rares et courantes. Ces outils sont connus sous le nom de clonage positionnel (Collins, 1992, p. 200). Ces techniques ont permis de mettre en évidence des gènes liés à une maladie sans avoir à identifier au préalable le produit protéique ou la fonction du gène. Depuis 1986, cette technique a permis d'isoler bien plus de 40 gènes liés à des maladies (Collins & Fink, 1995).

L'émergence, en pratique courante, du séquençage à haut débit en 2005 a profondément modifié l'organisation et les indications du diagnostic moléculaire (Klau et al., 2022). Il est basé sur la génération et la lecture en parallèle de millions de courtes séquences d'ADN indépendantes (*read*). Ces nouvelles technologies facilitent l'identification de variations et

de gènes (connus ou nouveaux) impliqués dans des pathologies et permettent donc d'augmenter le rendement du diagnostic moléculaire (Marshall et al., 2020).

c. Omiques : application des données de séquençage haut-débit

La révolution introduite par les méthodes de séquençage à très haut débit et à bas coût ne s'est pas limitée aux seules séquences des génomes. Tous les domaines de recherche qui impliquent directement ou indirectement une séquence d'acides nucléiques ont pu profiter des opportunités offertes par le séquençage à très haut débit (Figure 1. Différentes méthodes d'utilisation du séquençage haut-débit.). On peut par exemple citer :

- Le séquençage à haut débit du génome ou *WGS (Whole Genome Sequencing)* permet d'analyser l'ensemble de la séquence d'ADN du génome sans qu'il soit nécessaire d'utiliser des techniques de capture sélective pour isoler des régions spécifiques de l'ADN.
- L'analyse de la structure de la chromatine (*Hi-C-seq, ATAC-seq*) est fondée sur la création de liaison covalente au niveau de complexes ADN/protéine
- L'épigénome et les facteurs de transcription (*ChIP-seq, Methyl-seq*) : le *ChIP-seq* combinent l'immunoprécipitation de la chromatine (*ChIP*) avec le séquençage de l'ADN parallèle pour identifier les sites de liaison des protéines associées à l'ADN. Le *Méthyl-seq* va utiliser le bisulfite pour transformer les cytosines non méthylées en uraciles. La comparaison avec le génome de référence permet l'étude de l'épigénome
- L'analyse de la transcription (*RNA-seq*) permet de séquencer et quantifier l'ensemble des transcrits (transcriptome) d'un échantillon.

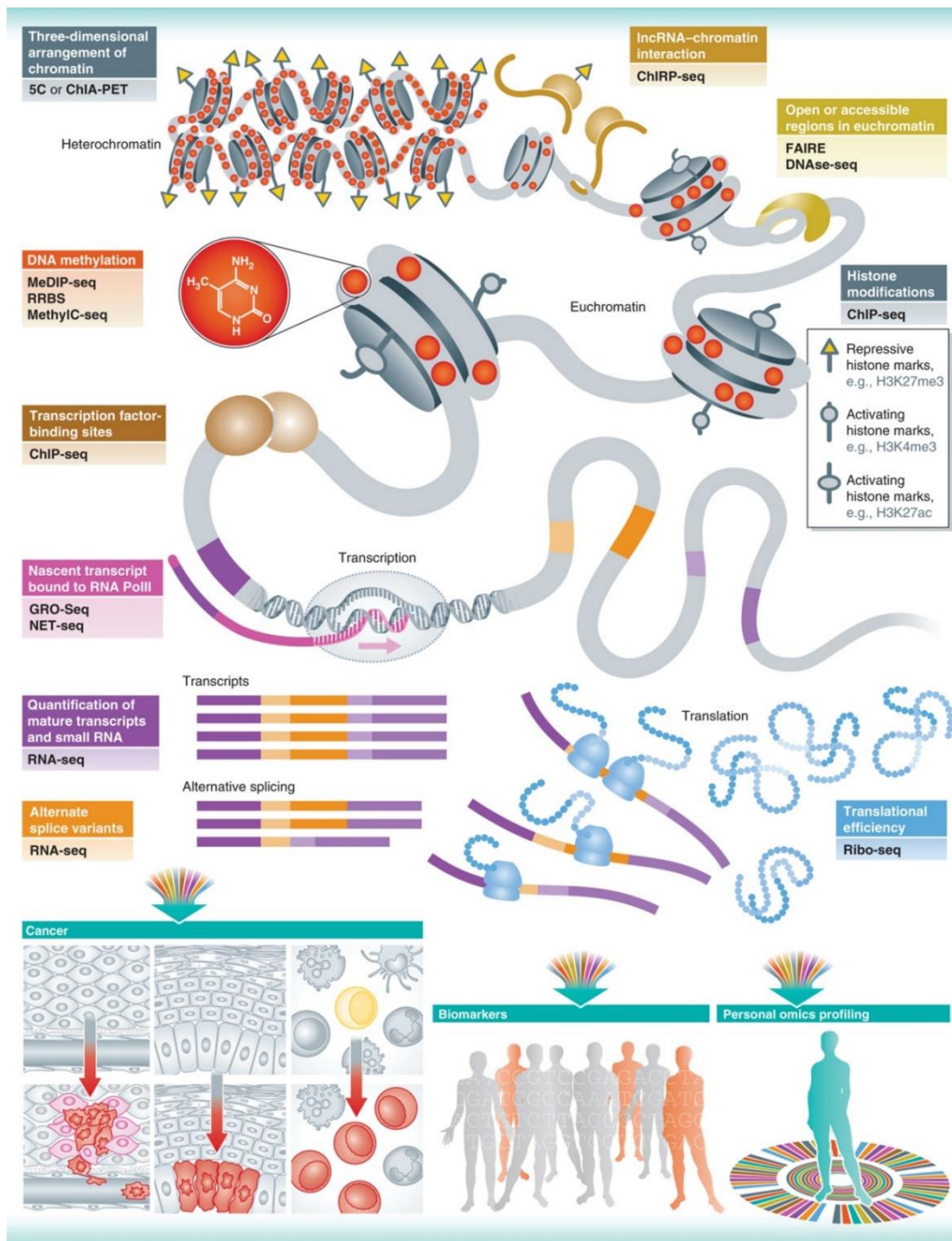


Figure 1. Différentes méthodes d'utilisation du séquençage haut-débit.

Diverses méthodes de séquençage haut-débit peuvent cartographier et quantifier avec précision l'état de la chromatine, les modifications de l'ADN, sa transcription et la traduction. Ces technologies peuvent être appliquées dans divers contextes en médecine, notamment pour mettre à jour les mécanismes de régulation et les profils d'expression qui distinguent les cellules normales des cellules cancéreuses, et pour identifier les biomarqueurs de maladies, en particulier les variations impliquant des sites régulation se situant en dehors de régions codantes. Ensemble, ces méthodes peuvent être utilisées pour établir des profils omiques personnels intégrés afin de cartographier tous les éléments régulateurs et fonctionnels d'un individu. Ces études seront la pierre angulaire de la médecine personnalisée et prédictive. Figure adaptée de (Soon et al., 2013).

## 2. Biologie des systèmes et compréhension des maladies complexes

Avec le développement des technologies de séquençage haut débit, la génétique a pris une toute autre ampleur et n'a plus été restreinte aux pathologies dont l'origine génétique était suspectée. La discipline est également devenue un angle d'analyse de nombreuses pathologies et en particulier les maladies dites complexes. Par exemple, récemment, le séquençage à haut débit a démontré son utilité dans les formes graves de covid-19 chez les personnes jeunes et a permis de mettre en évidence des prédispositions génétiques fréquentes dans cette population (Q. Zhang et al., 2020).

### a. Maladies complexes

Les maladies complexes ou multifactorielles sont définies comme des maladies dont l'étiologie et l'évolution sont déterminées par un certain nombre de facteurs génétiques, environnementaux et épigénétiques (Hunter, 2005). On peut par exemple citer les maladies auto-immunes comme la spondylarthrite axiale (fortement associée à l'allèle *HLA-B27*) (Bowness, 2015), la thrombose veineuse et le facteur V Leiden (Bloemenkamp et al., 1995), les maladies neurologiques comme la maladie d'Alzheimer (Hunter, 2005). Les facteurs génétiques peuvent déterminer la susceptibilité à certaines maladies comme l'hypertension artérielle (Lalouel & Rohrwasser, 2007), ou la sévérité de la présentation clinique comme la covid-19 (Asgari & Pousaz, 2021), ainsi que la réponse à certains traitements (Hunter, 2005). Cependant, ils ne représentent qu'une partie du risque associé aux phénotypes de maladies complexes. Le développement de la maladie dépend fortement de facteurs liés à l'environnement et au mode de vie. Par exemple, l'exposition au tabac est un facteur de risque fort à de nombreuses pathologies communes comme l'asthme et l'hypertension artérielle (Critchley & Unal, 2003). Plusieurs infections ont déjà été associées au déclenchement de pathologies comme Influenza A dans la Narcolepsie (Ravel & Mignot, 2019) ou plus récemment EBV dans la sclérose en plaque (Y. Xu et al., 2021).

Classiquement, l'étude de la genèse d'une pathologie est limitée à un facteur unique (un gène, une protéine, un pathogène). Cependant, les maladies complexes impliquent de nombreuses interactions entre les entités impliquées : leur étude par les approches traditionnelles est ainsi rapidement dépassée.

Les études d'association à l'échelle du génome (*GWAS*), dans lesquelles plusieurs centaines de milliers à plus d'un million de polymorphismes d'un seul nucléotide (*SNP*) sont analysés chez des milliers d'individus, représentent un outil puissant pour étudier l'architecture génétique des maladies complexes (Hardy & Singleton, 2009). Ces études ont permis d'identifier des centaines de variations génétiques associées à ces maladies et ont fourni des informations précieuses sur la complexité de leur architecture génétique (Hindorff et al., 2009). Le *GWAS* représente une avancée importante par rapport aux études conventionnelles sur des gènes candidats, dans lesquelles la taille des échantillons est généralement plus petite et l'explication du mécanisme limitée à quelques individus. L'étude d'association pangénomique repose sur l'hypothèse « maladie commune, variation commune » (Figure 2. Génétique des maladies complexes.), selon laquelle les maladies communes sont attribuables en partie à des variations alléliques présentes dans plus de 1 % de la population (Pritchard, 2001). Les variations les plus courantes, prises individuellement ou en combinaison, confèrent des augmentations relativement faibles du risque (1,1-1,5 fois) et n'expliquent qu'une faible proportion de l'héritabilité. Par exemple, plusieurs centaines de *loci* ont été associés à la taille humaine, un trait complexe classique dont l'héritabilité est estimée à environ 80 %, mais ils n'expliquent qu'environ 5 % de la variance phénotypique malgré des études portant sur des dizaines de milliers de personnes (Visscher, 2008).

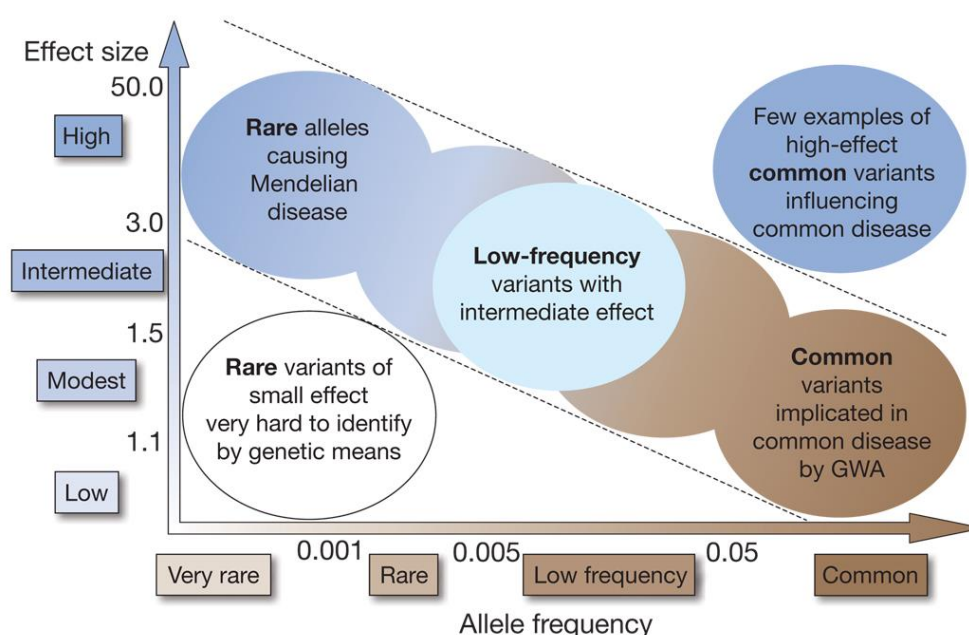


Figure 2. Génétique des maladies complexes.

Représentation de la possibilité de l'identification de variations génétiques en fonction de la fréquence de l'allèle à risque et de l'importance (*OR*, *odd ratio*) de l'effet génétique. Les associations indiquées dans les lignes diagonales en pointillé sont les plus intéressantes.

L'un des aspects de la génomique qui suscite également un intérêt considérable est l'épigénétique, c'est-à-dire les processus de régulation qui contrôlent la transcription des informations codées dans la séquence d'ADN en ARN avant leur traduction en protéines (Relton & Smith, 2010). En effet les gènes sont empaquetés dans la chromatine et des processus dynamiques de remodelage de la chromatine sont nécessaires pour l'étape initiale de l'expression des gènes (transcription), en modifiant l'accessibilité des promoteurs des gènes et des régions régulatrices (Vaissière et al., 2008). Les facteurs épigénétiques sont responsables de ce processus de régulation, dont les principales composantes sont la méthylation de l'ADN, les modifications des histones et l'action de petits ARN non codants. Contrairement à la séquence d'ADN, qui est largement fixée tout au long de la vie, les modèles épigénétiques varient non seulement d'un tissu à l'autre, mais se modifient avec l'âge et sont sensibles aux expositions environnementales (Lister et al., 2009; Relton & Smith, 2010). C'est cette dynamique qui fait que les processus épigénétiques suscitent un intérêt important, car ils se situent à l'interface de l'environnement et du contrôle transcriptionnel. Le rôle des processus épigénétiques aberrants est, par exemple, bien établi dans les troubles rares du développement (Nicholls, 2000).

#### b. Biologie des systèmes

La biologie des systèmes ou biologie systémique, est l'étude des interactions entre les composants d'un système biologique. On peut considérer qu'un « système » est un ensemble d'entités interagissant ou interdépendantes dont l'union forme un tout. Il n'y a pas de limite théorique à la taille d'un système : la définition des limites de l'environnement est assez subjective et dépend de l'angle d'analyse.

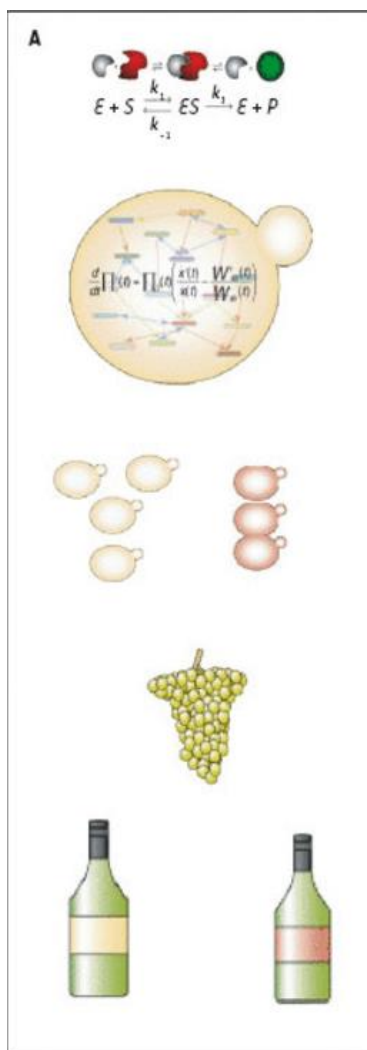


Figure 3. Illustrations de la notion de système biologique.

Schématisation de systèmes biologiques à différentes échelles d'étude. De haut en bas : une enzyme et son substrat, l'ensemble des voies métaboliques dans une levure, la population constituée de plusieurs levures individuelles partageant les mêmes ressources, la grappe de raisin comme écosystème, le vin comme produit d'intégration. (Carvunis et al., 2009)

Par exemple, considérons la Figure 3, si chacun de ces systèmes peut être étudié séparément, ils peuvent aussi être intégrés dans un unique système mixte. Par exemple on peut étudier comment les systèmes métaboliques de différentes levures produisent des produits de dégradation caractéristiques. De ces produits, on peut remonter à la variété des arômes contenus dans un vin puis au goût final du vin.

Un des aspects fondamentaux de la biologie des systèmes est le développement de nouvelles technologies afin d'explorer de nouvelles dimensions du système.

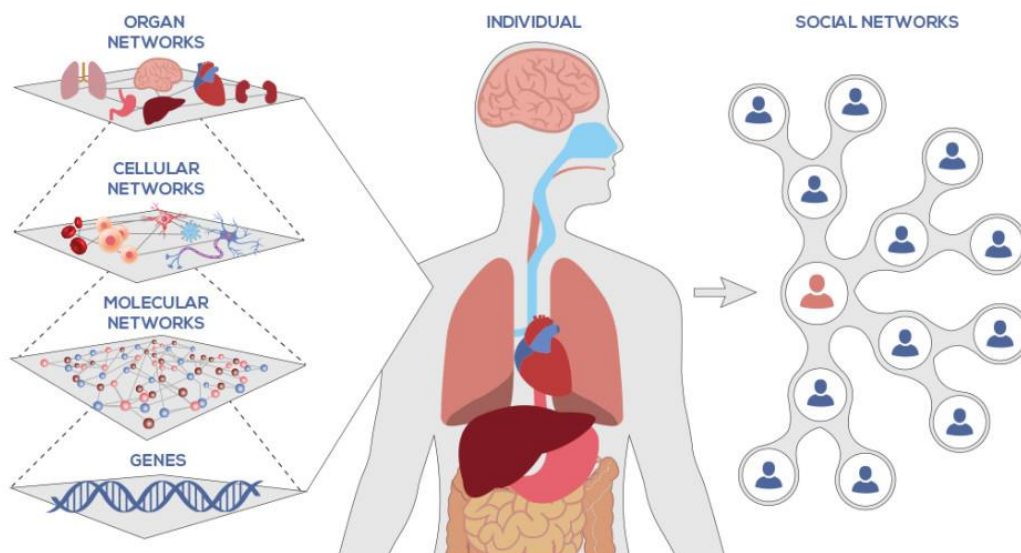


Figure 4. Réseau de réseaux.

Schéma illustrant la notion de « réseau de réseau ». L'individu est constitué d'un réseau d'organes lui-même constitué de réseau de cellules et de réseaux de molécules codés par de nombreux gènes. L'individu lui-même peut s'organiser en réseau social. Source : <https://isbscience.org/about/what-is-systems-biology/>

L'un des principes de la biologie des systèmes est le « réseau des réseaux » illustré par la Figure 4. Réseau de réseaux.. Notre corps est constitué de nombreux réseaux qui sont intégrés et communiquent à de multiples échelles. De notre génome aux molécules et aux cellules qui composent les organes de notre corps, jusqu'à nous-mêmes dans notre monde, nous sommes fondamentalement un réseau de réseaux. La biologie des systèmes étudie ces réseaux à toutes les échelles afin d'intégrer les comportements à différents niveaux, de formuler des hypothèses sur la fonction biologique et de fournir des informations spatiales et temporelles sur les changements biologiques dynamiques. Il ne suffit pas de comprendre une seule partie d'un système pour étudier la complexité de la biologie. C'est pourquoi le cadre du « réseau des réseaux » permet de comprendre pourquoi l'approche de la biologie des systèmes est différente, plus intégrée et plus apte à analyser et à prédire les transitions d'état dans les systèmes biologiques.

### c. Cartes moléculaires maladie-spécifiques

Une carte moléculaire maladie-spécifique en médecine systémique (« systems medicine disease map ») est une représentation des mécanismes d'une maladie, un modèle conceptuel, réalisée par une communauté d'experts (<https://disease-maps.org/>).

Elle est construite autour de groupes d'experts grâce à la lecture extensive de la littérature. Classiquement, une recherche bibliographique en utilisant *PubMed* sera effectuée en utilisant des mots clés spécifiques ou en partant d'articles de références.

La carte est ensuite construite en utilisant un logiciel spécialement dédié tel que CellDesigner (Funahashi et al., 2007) ou Newt (Balci et al., 2021). La couche détaillée de la carte est constituée de voies de signalisation, de métabolismes et de régulations géniques interconnectées, enregistrées dans des formats standard tels que SBGN (Le Novère et al., 2009), SBML (permettant un échange standard des données) et BioPAX (Figure 5. Représentation d'un modèle simplifié en utilisant la notation SBGN.. L'objectif est, comme le rappelle les auteurs de l'article, de résonner avec l'adage de Napoléon Bonaparte « Un bon croquis vaut mieux qu'un long discours ». La représentation standardisée des processus permet d'éviter les erreurs d'interprétation et de donner un sens universel et précis à chaque symbole. Trois langages sont ainsi décrits dans le SBGN : le flux d'activité qui ne s'attarde pas à l'aspect mécanistique, la description de processus qui est la représentation la plus détaillée de la réaction biochimique et la relation d'entité qui ne s'attache pas à la séquence des événements. Un des éléments majeurs de la description de la carte est également l'utilisation d'un système de notation détaillée (référence de l'entité, référence PubMed, etc.).

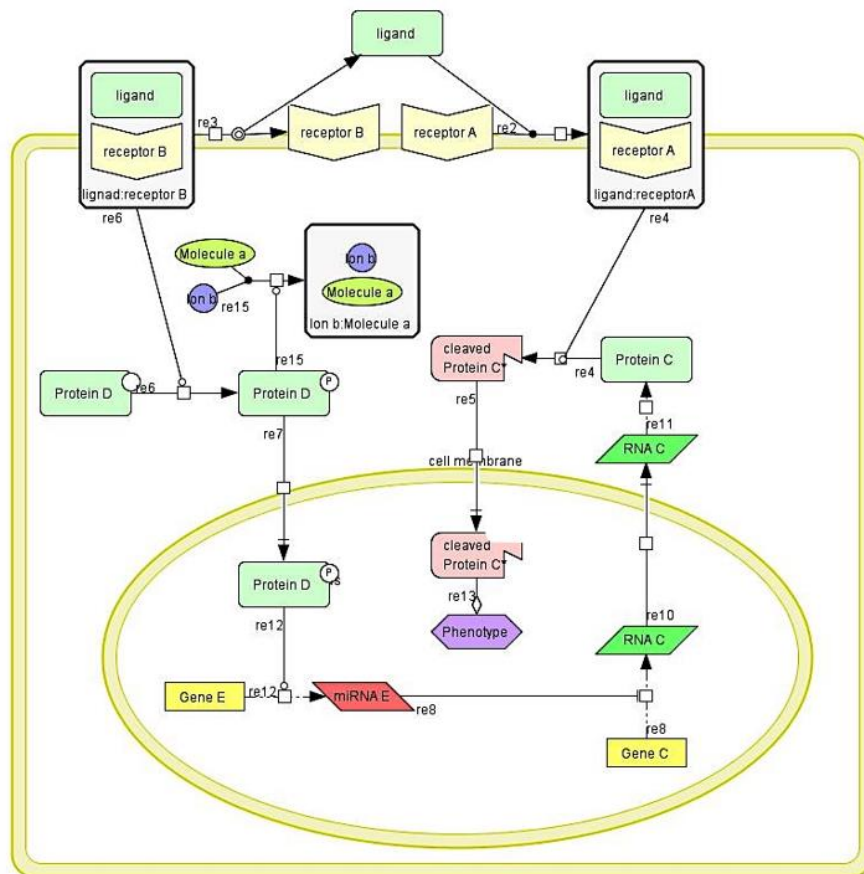


Figure 5. Représentation d'un modèle simplifié en utilisant la notation SBGN.

Cette représentation permet d'illustrer la diversité de la notation SBGN, ici en utilisant le logiciel CellDesigner. La majorité des réactions classiques sont représentées : création d'un complexe, phosphorylation, clivage, traduction, transcription. La représentation permet également la création de phénotype permettant d'explicitier et de terminer la finalité d'une de signalisation. Schéma tiré de (Ravel et al., 2020).

La participation d'experts du domaine (cliniciens et biologistes) garantit que toutes les caractéristiques de la maladie sont couvertes et que les connaissances pertinentes sont représentées de manière adéquate.

Les objectifs de telles cartes étant partagés, plusieurs communautés ont émergé comme *l'Atlas of Cancer Signaling Networks* (<http://acsn.curie.fr>) et le *Disease Map Project* (<http://diseasemaps.org>). Le projet *Disease Maps* est un réseau de groupes qui s'appuient sur des cartes moléculaires pour étudier les mécanismes de maladies (Ostaszewski et al., 2019). Il rassemble des groupes scientifiques et cliniques qui échangent les meilleures pratiques, partagent des informations et développent des outils de biomédecine systémique.

Ainsi, ces communautés ont permis, par l'utilisation de la curation manuelle (étude approfondie de la littérature) ou bien de l'inférence de réseaux à l'aide d'outils bio-informatiques, de produire des cartes moléculaires maladie-spécifique. La maladie de Parkinson (Fujita et al., 2014), l'asthme (Mazein et al., 2021), l'athérosclérose (Parton et al., 2019) ainsi que la polyarthrite rhumatoïde (Zerrouk et al., 2022) ont, par exemple, fait l'objet de cartes moléculaires maladie-spécifique.

Le premier bénéfice évident d'une telle approche est la synthèse graphique permettant la visualisation rapide des nombreux processus interconnectés. Il s'agit littéralement d'une carte d'interaction, organisée et structurée. Elle est généralement disponible en ligne avec un système de navigation permettant une vision globale ou détaillée associée à un outil de requête. Typiquement, MINERVA (Gawron et al., 2016) ou Navicell (Kuperstein et al., 2013) sont deux plateformes mettant à disposition en ligne ces cartes.

Les cartes de maladies individuelles représentent des images contextualisées de diverses pathologies. La comparaison des contenus de cartes maladie-spécifique peut aider à identifier la dérégulation de mécanismes spécifiques à un trouble donné. De plus, les cartes des maladies peuvent fournir un contexte précieux à l'interprétation de nombreuses données comme celles issues du séquençage haut débit. A plus long terme, les cartes des maladies pourraient également devenir une plateforme de prédiction de la réponse aux médicaments à partir de données de réseaux et de bases de données gènes - médicaments.

Un autres des objectifs de ces cartes est l'identification de cibles exploitables en clinique et l'aide à la prise de décision clinique. A cette fin, un des « chaînons manquant » est la transformation relativement automatisée de ces cartes en modèles mathématiques dynamiques exploitables (Büchel et al., 2013; Hemedan et al., 2022).

#### d. L'analyse de voies de signalisation

L'analyse des voies de signalisation ou analyse d'enrichissement fonctionnel, est devenue un des outils classiques d'analyse des données omiques. L'objectif est de déterminer des groupes de gènes impliqués dans un même processus fonctionnel et dont l'expression est différente entre les échantillons de cas par rapport à un contrôle. De cette

manière, ces méthodes cherchent à surmonter le problème de l'interprétation de listes extrêmement vastes de gènes importants, isolés et détachés de leur contexte biologique. Ces listes constituent en effet le principal résultat de nombreuses analyses de données à haut débit, comme l'analyse d'expression différentielle. Les méthodes d'analyse de voies de signalisation donnent un sens aux données biologiques expérimentales à haut débit, facilitant ainsi l'interprétation et la formulation d'hypothèses. En plus des cartes de voies de signalisation, de nombreuses bases de données existent répertoriant des listes de gènes représentant un ensemble fonctionnel sans forcément le réseau moléculaire les liant (García-Campos et al., 2015; Liberzon et al., 2015).

Les applications d'une telle approche sont nombreuses et incluent l'identification de nouvelles cibles thérapeutiques et l'aide à la stratification de risque (Yan et al., 2018). Par exemple, Mack et al. ont pu mettre en évidence la première cible thérapeutique pour l'épendymome, l'un des cancers du cerveau les plus répandus chez les enfants (Mack et al., 2014). En utilisant des données protéomiques, génomiques et transcriptomiques Zhang et al. ont identifié des gènes potentiels *drivers* du cancer du côlon (B. Zhang et al., 2014). D'autres auteurs ont par exemple mis en évidence le rôle clé de *TYRONP* dans la pathophysiologie de la maladie d'Alzheimer à début tardif (B. Zhang et al., 2013).

Près d'une centaine de méthodes d'analyses de voies signalisations ont été développés depuis les années 2000 (Nguyen et al., 2019; Tarca et al., 2013). L'analyse de surreprésentation (ORA) a été la première approche (Tavazoie et al., 1999) utilisée pour identifier les changements d'expression génétique observés en trouvant des groupes de molécules enrichies partageant une fonction similaire. L'inconvénient de telles approches est notamment qu'elles ne peuvent pas être appliquées si aucun gène différentiellement exprimés n'est trouvé dans une étude donnée, par exemple en raison d'un manque de puissance statistique.

Une deuxième génération de méthodes, appelées méthodes de notation des classes fonctionnelles (*Functional Class Scoring* - FCS), permet de ne pas avoir à sélectionner les gènes significatifs dans un premier temps et de fournir un résultat unique pour un ensemble de données. En règle générale, ces méthodes calculent un score à partir de tous

les gènes appartenant à un ensemble de gènes donné, qu'ils soient ou non exprimés de manière différentielle. Ces méthodes comprennent l'analyse d'enrichissement de l'ensemble des gènes (GSEA)(Subramanian et al., 2005).

Un type distinct de méthodes d'analyse des ensembles de gènes calcule un score d'ensemble de gènes dans chaque échantillon individuel à partir des niveaux d'expression des gènes observés. L'association entre le phénotype et les scores de l'ensemble de gènes au niveau de l'échantillon peut être réalisée à l'aide de modèles statistiques classiques. On pourra en particulier citer *single sample GSEA (SSGSEA)* (Barbie et al., 2009) ou le logiciel ROMA décrit ci-après (Martignetti et al., 2016; Najm et al., 2023).

#### e. ROMA

Le progiciel rRoma est une évolution de l'algorithme ROMA, un programme initialement développé en Java (Martignetti et al., 2016; Najm et al., 2023). Il s'agit d'un logiciel libre disponible sur github à l'adresse suivante : [www.github.com/sysbio-curie/rROMA](https://www.github.com/sysbio-curie/rROMA).

rRoma permet de quantifier l'activité d'ensembles de gènes impliqué dans un rôle fonctionnel commun. Ces ensembles de gènes peuvent ainsi correspondre à des gènes ayant les mêmes activités fonctionnelles, des gènes régulés par les mêmes motifs, des gènes appartenant à une même voie de signalisation ou des gènes formant un groupe de gènes fréquemment co-exprimés. L'hypothèse sous-jacente à rRoma est d'évaluer l'activité d'un ensemble de gènes en déterminant la quantité maximale de variance unidimensionnelle, qui est représentée par la première composante principale (PC1) dérivée du gène au sein de l'ensemble. Cette quantité est proportionnelle à l'influence d'un seul facteur latent sur l'expression génétique au sein de l'ensemble de gènes et reflète la variabilité de l'activité de ce facteur à travers les échantillons étudiés.

L'algorithme fait la distinction entre les ensembles de gènes décalés, « shifted », et les ensembles de gènes surdispersés, « overdispersed ». Un ensemble de gènes *shifted* correspond à la situation où l'expression médiane de tous les gènes de l'ensemble de gènes est significativement différente de celle de tous les gènes étudiés, c'est-à-dire que

l'ensemble de gènes présente une expression particulièrement élevée dans au moins un échantillon (Figure 6. Représentation des ensembles de gènes dans ROMA dans le cas de deux échantillons.). Un ensemble de gènes *overdispersed* correspond à la situation dans laquelle la variance expliquée par la PC1 calculée pour les seuls gènes de cet ensemble de gènes est significativement supérieure à la variance d'un ensemble de gènes de même taille sélectionnés au hasard.

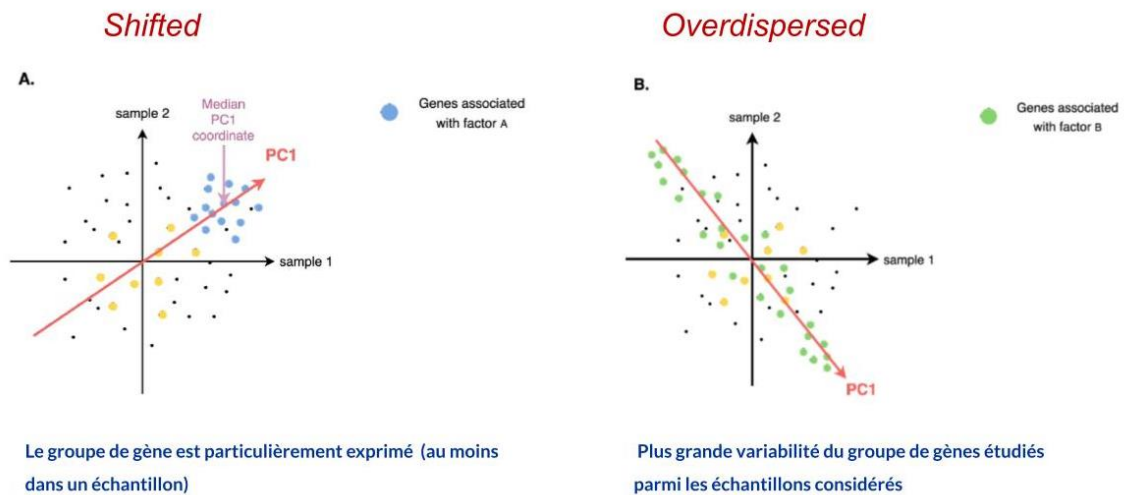


Figure 6. Représentation des ensembles de gènes dans ROMA dans le cas de deux échantillons.

Chaque point représente un gène, sa valeur horizontale correspondant à son expression dans l'échantillon 1 et sa valeur verticale dans l'échantillon 2. Les gènes associés au facteur latent A sont représentés en bleu et la direction PC1 correspondante est représentée en rouge (A). Cet exemple correspond à une voie *shifted*, évaluée par une médiane des projections de gènes sur la direction PC1 éloignée de l'origine de la distribution. Les gènes associés au facteur latent B sont représentés en vert (B) et la direction PC1 correspondante est représentée en rouge. Cet exemple correspond à une voie *overdispersed*, car la PC1 est bien alignée sur la distribution des points. Les gènes en jaune ne sont ni *overdispersed* ni *shifted*, car la PC1 explique une fraction relativement faible de la variance et la médiane des projections sur la PC1 est proche de l'origine pour ce groupe de gènes. Adapté de (Najm et al., 2023)

## Objectifs de thèses.

Les objectifs de cette thèse sont triples :

- (i) Concevoir un outil permettant de créer un arbre généalogique de façon simple mais complète et l'export du texte associé et des scores utiles en génétique oncologique (CanRisk)
- (ii) Construire une carte moléculaire des processus impliqué dans la mort cellulaire programmé et d'en démontrer son intérêt en analysant des données déjà existantes
- (iii) Rechercher des différences dans les processus impliqués dans la covid-19 en utilisant rRoma et en analysant de voies de signalisation classique sans se limiter aux connaissances présentes sur une carte moléculaire.

## 2 GeneTree

Ce premier volet de ma thèse vise à trouver une solution à un problème très pratique rencontré par les cliniciens en génétique médicale, et en particulier en oncogénétique : la redondance des saisies concernant l'histoire familial d'un patient.

GeneTree: using patient-filled electronic medical form to facilitate medical genetics consultation. Jean-Marie Ravel, Axelle Rivière, Marion Imbert-Bouteille, Manon Ravel, Céline Bonnet, Bruno Leheup, Kévin Yaury, Laëtitia Lambert. Article en cours de préparation.

### 1. Introduction

#### a. La génétique oncologique

Entre 5 et 10 % des cancers sont liés à la présence d'une altération génétique constitutionnelle héréditaire dans un gène de prédisposition au cancer (Turnbull & Hodgson, 2005).

Si l'on considère en particulier le cancer du sein, le cancer le plus fréquent chez la femme, il n'existe pas de définition unique du cancer du sein « familial », mais les critères généralement acceptés sont les suivants : (i) au moins trois cas de cancer du sein et/ou de l'ovaire dans une famille ; (ii) deux cas de cancer du sein chez des parents proches, dont au moins un diagnostiqué avant l'âge de 50 ans ; (iii) tout cancer du sein chez l'homme (iv), cancer du sein bilatéral ou cancer du sein et de l'ovaire chez la même patiente (v), cancer du sein avant 35 ans (Balmaña et al., 2011).

Des variations pathogènes dans les gènes *BRCA1* et *BRCA2* sont mises en évidence dans environ 15 à 20 % des familles avec prédisposition au cancer du cancer du sein et de l'ovaire familiale héréditaire (Caputo et al., 2021). Les femmes porteuses de mutations dans les gènes *BRCA1* ou *BRCA2* ont un risque de cancer du sein de 50 % à 85 % au cours de leur vie (King et al., 2003). D'autres gènes rares, mais à forte pénétration, incluent *PTEN* (FitzGerald et al., 1998), *TP53* (Birch et al., 2001), *CDH1* (Pharoah et

al., 2001) et STK11 (Lim et al., 2004), chacun conférant un syndrome clinique distinct. D'autres gènes comme *RAD51C* et *RAD51D* ont ensuite été associés à un risque accru de cancer de l'ovaire (Loveday et al., 2012).

En routine, 32 gènes sont analysés par exemple au CHRU de Nancy dont 13 sont jugés d'intérêt clinique particulier, l'identification d'une variation pathogène ou probablement pathogène chez le patient conduisant à proposer une surveillance spécifique devant un très haut risque.

#### b. Déroulement de la consultation d'oncogénétique

L'étape préalable à la consultation est la validation du bien fondé de celle-ci. Il faut souvent demander des précisions au patient ou au praticien sollicitant la consultation afin de filtrer les situations ne relevant vraisemblablement pas d'une prédisposition génétique au cancer. La consultation initiale dure entre 30 et 45 minutes. La moitié du temps est consacrée à la collecte de l'ensemble des éléments personnels et familiaux et au traçage de l'arbre généalogique (Figure 7. Représentation du temps consacré à la préparation et réalisation d'une consultation.. Plusieurs éléments sont évocateurs de formes héréditaires familiales et vont être recherchés en consultation. On peut citer par exemple l'excès de cas atteints de cancer dans une même famille, un âge précoce de la survenue du cancer, des formes bilatérales ou multifocales (par exemple de cancer de sein).

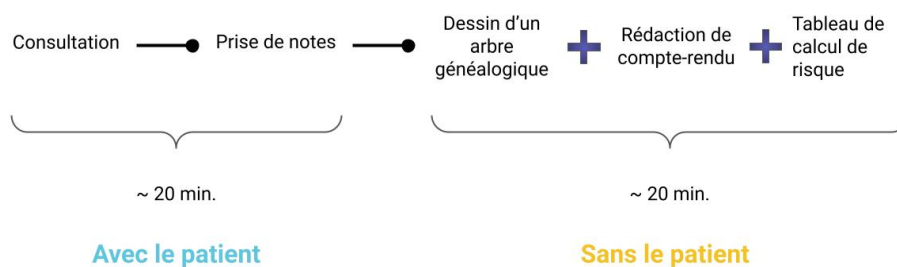


Figure 7. Représentation du temps consacré à la préparation et réalisation d'une consultation.

Ces chiffres sont estimatifs et basés sur notre expérience personnelle et n'incluent pas le temps de préparation en amont de la consultation (recueil des données, examen du dossier numérique, etc.)

A l'issue de cet entretien, le praticien onco-généticien estime à partir des données qu'il a recueillies la probabilité de la présence d'une prédisposition génétique au cancer. Il peut alors proposer une analyse génétique et explique les enjeux et les conséquences d'un tel examen. Il rappelle en particulier au patient les notions de bases de la génétique, le principe de transmission et de prédisposition génétique. Si le patient consent à l'analyse, il signe le formulaire de consentement, un prélèvement est réalisé à l'issue de la consultation et l'analyse effectuée au sein d'un laboratoire de génétique compétent.

Après réception du résultat de l'analyse génétique, deux situations se distinguent. Si une variation pathogène est mise en évidence dans un gène de prédisposition forte au cancer, une surveillance va être proposée en fonction du gène mutée. Si ce n'est pas le cas, une estimation du risque résiduel va être fait et une surveillance particulière, différente de celle de la population générale, va être discutée.

Dans le cadre du cancer du sein, un logiciel de prédiction du risque a été développé par une équipe de Cambridge : Boadicea (Lee et al., 2019) remplacé depuis 2020 par CanRisk (Carver et al., 2021). L'outil CanRisk est une interface web du modèle de prédiction des risques BOADICEA (*Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm*) utilisé pour calculer le risque de cancer du sein et de l'ovaire chez une femme donnée. Il s'agit d'un modèle permettant de prédire le risque de cancer du sein et de l'ovaire chez des femmes non affectées, sur la base d'informations de dépistage de mutations pour des variations génétiques rares (à risque élevé et modéré) de susceptibilité au cancer du sein et de l'ovaire, des variations génétiques communes de susceptibilité au cancer (à l'aide de scores de risque polygénique ou *PRS*), des antécédents familiaux, du mode de vie personnel, des facteurs de risque hormonaux et reproductifs, ainsi que de la densité mammographique dans le cas du cancer du sein.

L'ensemble de ces éléments (résultats et surveillance) est ensuite expliqué au patient lors de la consultation de restitution.

### c. Problématique

La saisie de l'arbre généalogique est centrale dans la consultation de génétique. La saisie des données dans le cadre d'une consultation d'oncogénétique expose au risque de redondance concernant l'histoire familiale d'un patient. Ces informations vont en effet être recueillies au pour la construction de l'arbre généalogique, puis saisies sur l'interface web de CanRisk, et enfin saisies pour la génération du compte rendu de consultation.

Plusieurs propositions permettant la génération d'arbre généalogique existent, aussi bien payantes que gratuites. Néanmoins, à ce jour, aucune solution ne permet la génération d'un arbre généalogique modifiable et exportable rapidement ainsi que le texte associé.

	<i>PedigreeXP</i>	<i>PhenoTips</i>	
Arbre généalogique	✓	✓	✓
Rédaction de CR	✗	✗	✓
Tableau calcul de risque	✓	✓	✓
Stockage sur serveur	✗	✓	✓
Utilisation en UE	✓	✗	✓

Figure 8. Comparaison de GeneTree avec les principales solutions existantes.

En particulier, l'étude des solutions existantes a montré qu'aucun logiciel ne permettait la génération de texte, l'envoi d'un formulaire à un patient (Figure 8. Comparaison de GeneTree avec les principales solutions existantes..

Un des points majeurs compliquant ce projet est la volonté de respecter le RGPD (Règlement Général sur la Protection des Données qui encadre le traitement des données personnelles sur le territoire de l'Union européenne) et les bonnes pratiques en logiciels de santé. En particulier, les données de santé doivent être enregistrées sur un serveur habilité (certifié HDS, Hébergeur de Données de Santé) afin de garantir un haut niveau de sécurité et de confidentialité. Les données de santé sont en effet des données à caractère personnel sont considérées comme sensibles. Elles font à ce titre l'objet d'une protection particulière par les textes (règlement européen sur la protection des données personnelles, Loi Informatique et Libertés, code de la Santé publique, etc.) afin de garantir le respect de la vie privée des personnes et, en l'occurrence, de leurs apparentés.

L'objectif de ce projet était donc de développer un outil qui permettrait (1) une utilisation facile par un clinicien, (2) l'export d'un arbre, d'un score de risque et du texte de compte rendu, (3) une implémentation facile au sein d'un hôpital, si possible sans adaptation importante nécessaire de son système, (4) la sécurité de la saisie des informations des patients.

## 2. Résultats

Une application Web a été créée sur la base de deux modules javascript existants : pedigreeJS (Carver et al., 2018) et handsontable (<https://handsontable.com/>). Elle est hébergée sur Github de façon gratuite. L'ensemble du traitement des données est fait sur le poste client, sans échange avec le serveur.

Cette application est disponible librement et gratuitement en Français et en Anglais et a la particularité d'avoir une interface dédiée au patient : <https://jimouse.github.io/GeneTree/>

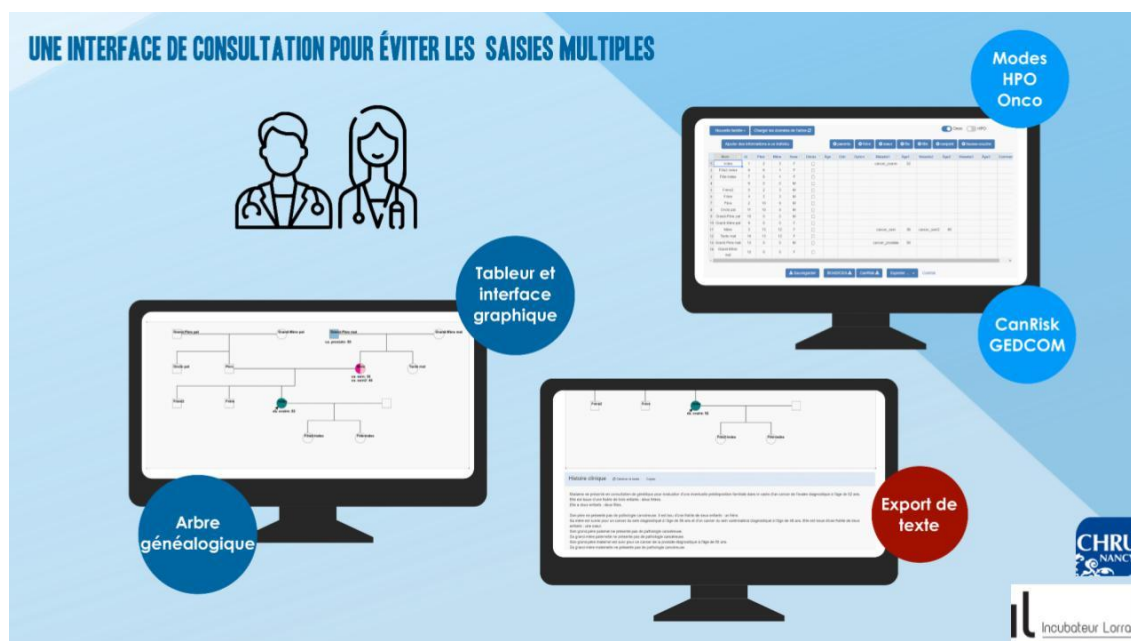


Figure 9. Captures d'écrans de l'interface actuelle de l'application.

A gauche de l'image l'interface du praticien et à droite l'interface spécifiquement conçue pour le patient. Un guide complet est disponible sur l'application.

### a. Présentation générale

Cette application est un outil intégré pour réduire le temps de travail sur l'ordinateur des médecins et des conseillers en génétiques dans le cadre de leur consultation de génétique. Une explication détaillée est disponible en annexe ainsi que sur le site de l'application. Une vidéo YouTube™ expliquant les fonctions de bases a également été publiée (<https://www.youtube.com/watch?v=FneR3XdMjhc>).

Les données portant sur la famille peuvent être chargées (fichiers JSON ou BOADICEA) ou créées à partir d'une structure standard ou personnalisée, puis complétées via un tableau ou une interface graphique, tous deux interconnectés.

L'application propose un mode spécifique pour la consultation d'oncogénétique et un mode utilisant les phénotypes HPO (Human Phenotype Ontology, un vocabulaire standardisé des anomalies phénotypiques rencontrées dans les maladies humaines) contenant la majorité des symptômes et maladies décrites (Köhler et al., 2017). Cette approche permet de reporter les symptômes et maladies de façon reproductible. Chaque terme de l'HPO décrit une anomalie phénotypique.

L'arbre est exportable dans plusieurs formats de fichiers (TSV, JSON, PED, BOADICEA) et modifiable par un éditeur vectoriel (format PDF, SVG) ou imprimable directement.

Enfin, cette application permet la génération automatique à partir du contenu du tableau du texte correspondant à l'histoire personnelle et familiale du patient.

L'interface en ligne optimisée pour le navigateur Google Chrome™ permet une utilisation simple sans installation préalable. Cet outil a été particulièrement pensé pour les consultations d'oncogénétique afin d'éviter la triple saisie (texte – arbre généalogique - score de risque Boadicea ou CanRisk).

Une interface dédiée aux patients a également été créée. Elle permet, en amont de la consultation, au patient de créer l'arbre de façon didactique.



Figure 10. Interface dédiée au patient.

L'application est utilisable sur ordinateur ou sur tablette et, dans une moindre mesure, sur un téléphone portable.

Les données sont ensuite téléchargeables et peuvent être envoyées au médecin correspondant.

## b. Impacts

### *Un gain de temps*

L'utilisation de cet outil permet un gain de temps estimée à 10 à 20 minutes en moyenne par patient en dehors de l'utilisation du questionnaire patient. En particulier, il permet de gagner 5 à 10 minutes sur la génération du courrier (ou de l'arbre) et le même temps sur le remplissage du score canRisk dans le cas de cancer du sein et de l'ovaire. Pour une matinée type moyenne constituée de quatre consultations initiales, le gain de temps est donc d'une heure en moyenne par jour (sur la base d'une demi-journée de consultation par jour, ce qui est classique). Cette approche permet une réutilisation simple des données et en particulier de modifier les données de la personne à étudier (frère, sœur, enfants etc.) en un clic de souris.

### *Des données de qualité*

Si le gain de temps de l'application semble certain, l'intérêt de l'interface vers le patient reste à démontrer. Pour ce faire, une collaboration avec un praticien du Centre de Lutte

Contre le Cancer (CLCC) de Montpellier a été initiée. Les patients pour lesquels un rendez-vous de consultation de génétique oncologique avait été validé se voyait proposer d'utiliser GeneTree en amont de la consultation et d'envoyer les fichiers résultants au service. L'arbre était ensuite complété en consultation par le praticien.

De juin 2022 à janvier 2023, 23/118 (19,5 %) patients ont envoyé leur formulaire au centre de génétique du CLCC de Montpellier. En comparant l'arbre généalogique rempli par le patient et celui établi par le médecin, nous avons constaté que l'histoire familiale remplie par le patient fournissait 42 des 51 (82,4%) antécédents de cancer pour lesquels un examen génétique est recommandé, ce qui représente 71,4% des arbres généalogiques établis par les praticiens avec une description précise.

Ces chiffres soulignent le fait que l'information fournie par le patient est de bonne qualité mais que l'expertise médicale est nécessaire pour la vérifier et la compléter. Le gain de temps est donc réel à ce niveau aussi.

## 2. Discussion et perspectives

Nous avons développé GeneTree, une application au service du généticien en particulier en génétique oncologique. Elle permet de façon simple et sécurisée de saisir une seule fois les informations du patient, éventuellement par le patient lui-même, et d'exporter l'arbre, le compte rendu et les scores de risques.

Cette application est, à notre connaissance, la seule à proposer toutes ces fonctionnalités. De plus, son utilisation actuelle en pratique clinique dans deux centres différents souligne son intérêt. Le gain de temps est estimé au total à plus de 30 minutes par patients c'est-à-dire plusieurs heures potentiellement gagnées par semaine. De plus, l'utilisation de l'interface patient offre au patient la possibilité de s'investir dans la consultation, de devenir acteur.

Cet outil, qui a été couronné par plusieurs prix (prix Thierry Frebourg 2022, prix BioInfo Diag 2022, prix poster SIGF 2022) est aujourd'hui utilisé quotidiennement au service de génétique clinique du CHRU de Nancy, ainsi que dans le service d'onco-

généétique du CLCC de Montpellier. GeneTree a également été lauréat du prix universitaire 2022 de l'Incubateur Lorrain. Ce projet est donc entré en incubation en Mars 2023, j'ai été recruté en binôme avec mon épouse Manon Ravel. Cette incubation vise à déterminer s'il est possible de créer une entreprise autour de ce produit et d'en définir les paramètres.

A son terme, le projet évoluerait vers une application plus complète avec un serveur permettant le stockage de données patients et la sécurisation des échanges entre le patient et le médecin.

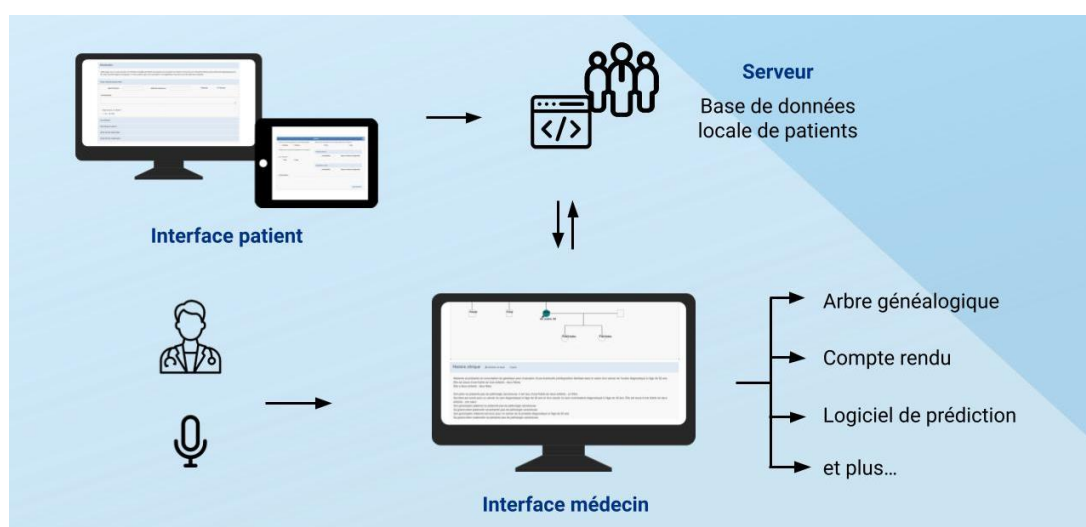


Figure 11. Représentation du projet GeneTree.

Le projet définitif comprend la mise en place d'un serveur qui permet de sauvegarder les données au sein d'un centre, de sécuriser les échanges avec le patient et de permettre des échanges entre centres et au sein du système informatique du centre.

Cette idée de limiter la redondance en médecine serait tout à fait généralisable à d'autres spécialités. On peut évoquer par exemple les questionnaires effectués de façon systématique dans ce nombreuses autres spécialités (consultation d'anesthésie, etc.).

Si l'on se place de façon plus globale dans la thématique de la génétique, l'utilisation d'un modèle monogénique est très performante dans plusieurs pathologies, y compris, comme développé plus haut dans l'identification d'un syndrome de prédisposition aux cancers. Néanmoins, il n'est pas rare qu'une cause monogénétique ne soit pas identifiée même dans des situations où une prédisposition forte au cancer est supposée. Cela peut être dû à des gènes encore non associés à une prédisposition génétique au cancer et donc non analysés, des variations difficiles à détecter avec les approches classiques d'exploration du génome (intronique profond, promoteur) ou d'autres mécanismes plus complexes.

Symétriquement, la majorité des pathologies monogénétiques ont une grande variabilité phénotypique. C'est un point qui a d'ailleurs été montré dans mes travaux ultérieurs sur la narcolepsie (Ollila et al., 2015) et plus récents (*STUB1*, *SARS1*, *GCH1*) (Ravel, Benkirane, et al., 2021; Ravel, Dreumont, et al., 2021; Ravel et al., 2023, 2022).

Cette variabilité d'expression (ou variabilité phénotypique) que ce soit à l'échelle d'un individu ou d'une cellule est un des aspects étudiés par les approches de biologie des systèmes. C'est le cas du projet ACSN (Monraz Gomez et al., 2019) qui vise à aborder la complexité du cancer par la représentation systématique des voies de signalisation impliquées dans le cancer sous la forme de cartes de réseaux complètes. Un des aspects majeurs de cet atlas, la mort cellulaire programmée, est traitée dans le chapitre suivant.

### 3 La mort cellulaire régulée

Ce travail a fait l'objet d'une communication écrite publiée :

Ravel, J.-M.; Monraz Gomez, L.C.; Sompairac, N.; Calzone, L.; Zhivotovsky, B.; Kroemer, G.; Barillot, E.; Zinovyev, A.; Kuperstein, I. *Comprehensive Map of the Regulated Cell Death Signaling Network: A Powerful Analytical Tool for Studying Diseases*. *Cancers* 2020, 12, 990. <https://doi.org/10.3390/cancers12040990> (Ravel et al., 2020)(Ravel et al., 2020)

#### 1. Contexte

Le terme « cancer » englobe un groupe de maladies se caractérisant par la multiplication et la propagation anarchiques de cellules anormales. Des outils et technologies expérimentaux et informatiques fournissent une avalanche de données sur la myriade de manifestations du cancer aussi bien au niveau cellulaire, que tissulaire ou individuel. Hanahan et Weinberg ont proposé en 2000 six caractéristiques du cancer distinctives et complémentaires qui favorisent la croissance tumorale et la dissémination métastatique, « the hallmark of cancer » (Hanahan & Weinberg, 2000). Depuis, ces critères ont été mis à jour et au moins huit *hallmark* sont aujourd'hui décrits : les capacités acquises pour maintenir la signalisation proliférative, échapper aux suppresseurs de croissance, résister à la mort cellulaire, permettre l'immortalité répllicative, induire/accéder à la vascularisation, activer l'invasion et les métastases, reprogrammer le métabolisme cellulaire et éviter la destruction immunitaire (Hanahan, 2022). La dérégulation du métabolisme cellulaire et l'évitement de la destruction immunitaire sont maintenant également considérées comme des caractéristiques essentielles du cancer, au même titre que les caractéristiques initialement décrites.

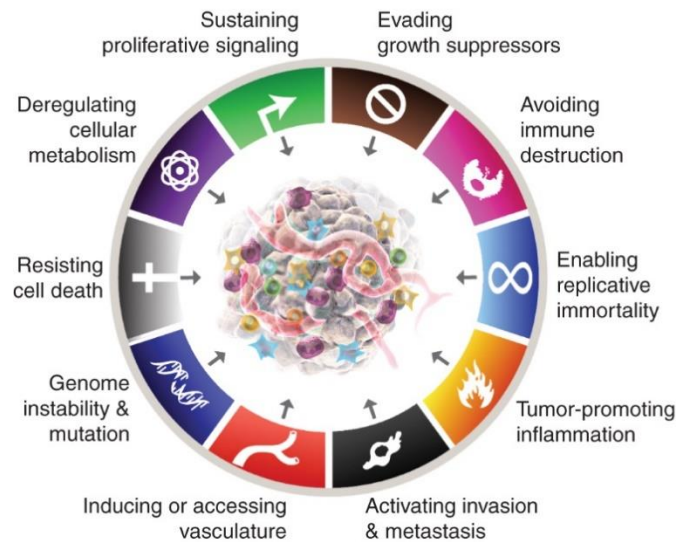


Figure 12. Les caractéristiques du cancer, version 2022.

Les caractéristiques du cancer englobent actuellement huit capacités et deux caractéristiques

La cancérogenèse est ainsi due à un fonctionnement aberrant d'un réseau complexe d'interactions moléculaires qui affecte simultanément de multiples fonctions cellulaires. Par conséquent, la qualité de l'analyse des données dépend fortement de la qualité des réseaux de signalisation. C'est à partir de ce constat qu'a été développé l'Atlas du réseau de signalisation du cancer (ACSN, *Atlas of Cancer Signalling Network*), une carte interactive et complète des mécanismes moléculaires impliqués dans le cancer. Cette ressource comprend des outils de navigation, de visualisation et d'analyse des données moléculaires dans le contexte des cartes de réseaux de signalisation. Le contenu d'ACSN couvre les principaux mécanismes impliqués dans la progression du cancer, notamment la réparation de l'ADN, la survie cellulaire, l'apoptose, le cycle cellulaire, l'EMT (Transition épithélio-mésenchymateuse) et la motilité cellulaire (Figure 13. Image du portail ACSN).

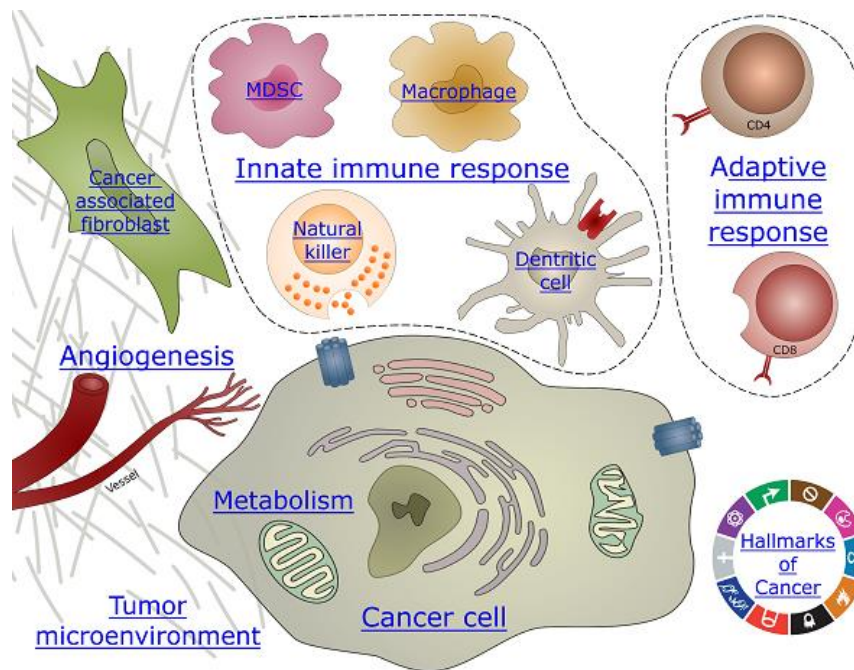


Figure 13. Image du portail ACSN.

Cette image représente les différentes cartes regroupées dans le projet ACSN. On note en particulier l'étude non seulement de la cellule cancéreuse mais également de son environnement. Jean-Marie Ravel et Maria Kondratova, Copyright (c) 2018.

L'objectif de ce projet était la création d'une carte décrivant les mécanismes de morts cellulaires régulées impliqués dans le cancer.

## 2. Introduction

Les mécanismes moléculaires conduisant à la mort cellulaire sont perturbés dans de nombreuses maladies (Kang et al., 2019; Pentimalli et al., 2018). Deux scénarios principaux scellent le destin final d'une cellule. Les cellules peuvent soit mourir d'une mort cellulaire régulée (RCD, *Regulated Cell Death*), soit succomber à une mort cellulaire accidentelle (Ouyang et al., 2012; Stevens et al., 2013). La mort cellulaire accidentelle, ou non régulée, souvent appelée nécrose, est un processus soudain qui combine la rupture de la membrane cellulaire et la libération du contenu cellulaire dans le milieu environnant (Kaczmarek et al., 2013). En revanche, la mort cellulaire régulée implique une contribution active des cellules et peut être exécutée par une panoplie de voies différentes qui incluent l'apoptose, la nécroptose, l'autophagie, la ferroptose, la parthanatose et la pyroptose (Galluzzi et al., 2018; Yatim et al., 2017).

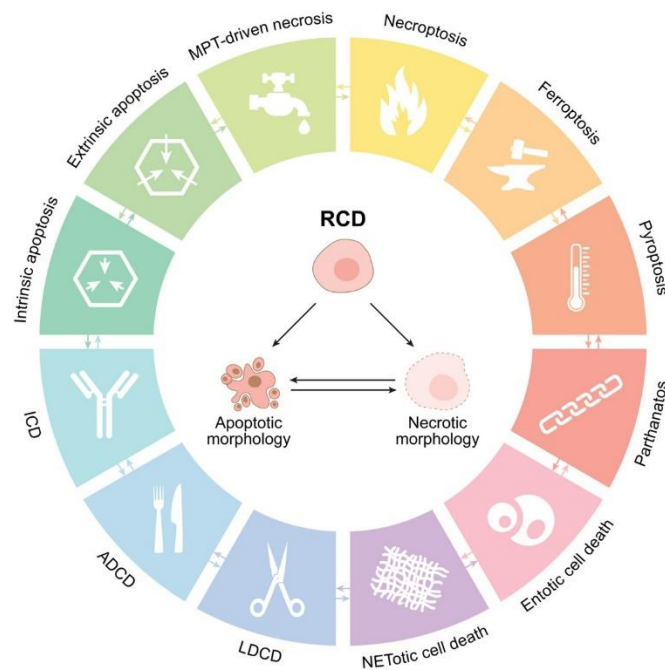


Figure 14. Programmes majeurs de mort cellulaire

Les cellules exposées à des perturbations irrémédiables du microenvironnement intracellulaire ou extracellulaire peuvent activer l'une des nombreuses cascades de transduction de signaux qui conduisent finalement à leur mort. Chacun de ces modes de mort cellulaire régulée (RCD) est initié et propagé par des mécanismes moléculaires qui présentent un degré considérable d'interconnectivité. En outre, chaque type de mort cellulaire régulée peut se manifester par un spectre complet de caractéristiques morphologiques allant de la nécrose totale à l'apoptose totale, et par un profil immunomodulateur allant de l'anti-inflammatoire et du tolérogène au pro-inflammatoire et à l'immunogène. ADCD : mort cellulaire dépendante de l'autophagie, ICD : mort cellulaire immunogène, LDCD : mort cellulaire dépendante du lysosome, MPT : transition de perméabilité mitochondriale (Galluzzi et al., 2018).

L'un des modes de RCD les plus étudiés est l'apoptose, un processus d'élimination cellulaire « propre » considéré comme non inflammatoire et non immunogène. Toutefois, dans certaines circonstances, les cellules mourant par apoptose peuvent induire une réponse immunitaire contre les antigènes des cellules mortes (Yatim et al., 2017). L'apoptose est un processus hautement régulé, dépendant de l'énergie, qui nécessite un apport d'ATP et qui dépend donc d'un métabolisme bioénergétique fonctionnel. La gravité du stress environnemental, des dommages, des perturbations génétiques ou épigénétiques dicte le choix entre le programme de survie ou de mort et le mode d'exécution de la mort cellulaire, qu'il s'agisse de l'apoptose ou d'un autre sous-programme létal (Ellis et al., 1991). Le défi consiste à intégrer toutes les voies distinctes d'initiation, de signalisation et d'exécution en une seule carte transparente. Le RCD est coordonné par de nombreuses voies de signalisation, y compris des signaux trophiques relayés par AKT et mTOR, des kinases de stress, notamment ERK et JNK, ou des facteurs de transcription tels que NF-

xB. La RCD dépend également des mitochondries et du métabolisme du glucose qui dictent l'état de l'approvisionnement en énergie. En outre, il existe un niveau complexe de régulation post-transcriptionnelle par des ARN non codants, dont les miARN, et des modifications post-traductionnelles des protéines liées à l'apoptose. D'un point de vue morphologique, l'apoptose se caractérise par le rétrécissement des cellules, la condensation et la fragmentation des noyaux, le détachement des cellules des tissus environnants et l'exposition à la surface de signaux tels que la phosphatidylsérine qui facilite l'élimination des cellules mourantes par les macrophages.

Une autre forme de RCD bien étudiée, l'autophagie, est activée dans diverses situations de stress, en particulier en réponse à une pénurie d'énergie dans un contexte de famine ou de pénurie de facteurs de croissance (Xie et al., 2016). L'autophagie assure l'homéostasie cellulaire en recyclant les organites cellulaires et les macromolécules, assurant ainsi la cytoprotection et réduisant la propension des cellules à subir un RCD. Dans certaines circonstances, cependant, l'autophagie peut se transformer en un processus d'auto-élimination conduisant à la mort cellulaire. La nécroptose est un mode de mort cellulaire médié par la signalisation, dont les caractéristiques morphologiques sont similaires à celles de la nécrose. La nécroptose s'accompagne d'un "gonflement" des cellules et des organites et d'une perte rapide de l'intégrité de la membrane plasmique. La ferroptose est un mode de mort cellulaire induit par une accumulation de peroxydes lipidiques due à un dérèglement du métabolisme du fer qui facilite la génération d'espèces réactives de l'oxygène (ROS) et la peroxydation consécutive des lipides membranaires (Xie et al., 2016). Le parthanatose est un mode de mort cellulaire dépendant de la poly [ADP-ribose] polymérase 1 (PARP1), induit principalement par des facteurs intracellulaires tels que les lésions de l'ADN. Lorsqu'elle est activée par des cassures de l'ADN simple brin, la PARP1 synthétise des chaînes de poly (ADP-ribose) (PAR) qui sont attachées de manière covalente aux protéines nucléaires tout en consommant de l'ADP et du NAD<sup>+</sup>. PAR agit alors comme un signal d'activation pour d'autres enzymes de réparation de l'ADN. Cependant, en cas de dommages excessifs à l'ADN, les processus de réparation de l'ADN sont surstimulés, ce qui entraîne la consommation d'ATP et l'épuisement du NAD, déclenchant l'exécution cellulaire. Le parthanatose est associée à la fission mitochondriale, son principal phénotype (Fatokun et al., 2014). Enfin, la pyroptose est une modalité de

mort cellulaire dépendante de la caspase-1 qui peut être activée par des signaux extracellulaires (tels que le lipopolysaccharide bactérien, LPS) qui se lie à des récepteurs (tels que NLRC4), induisant une dimérisation du récepteur qui déclenche le recrutement de la protéine PYCARD et donc l'activation de la caspase-1.

Ces très nombreuses connaissances sur les mécanismes de mort cellulaire régulée sont dispersées dans des milliers de publications scientifiques. Cette situation complique non seulement la compréhension globale de la régulation de ce processus ubiquitaire, mais aussi l'application d'approches de biologie des systèmes. Les efforts de synthèse systématique et de formalisation des connaissances sur les interactions moléculaires dans un format structuré et lisible par ordinateur ont déjà été appliqués à de nombreuses bases de données de voies. Plusieurs bases de données contiennent des collections de diagrammes de signalisation qui se chevauchent partiellement et qui représentent de manière assez détaillée les différents modes de mort cellulaire, les mitochondries et le métabolisme du glucose (Croft et al., 2014; Hoffmann, 2008; M. Kanehisa et al., 2012; M. T. K. E. G. G. database Kanehisa, 2002; Perfetto et al., 2016). Toutefois, le facteur limitant de ce type de représentation de la signalisation est l'absence d'une vue intégrée des voies de signalisation.

### 3. Matériels et méthodes

#### a. Construction de la carte

Le réseau de signalisation de la régulation de la mort cellulaire régulée a été construit manuellement et est basé sur des informations extraites de la littérature récupérée dans la base de données PubMed. La carte a été composée à l'aide de l'outil CellDesigner (Funahashi et al., 2007) et de la syntaxe visuelle standard *Systems Biology Graphical Notation* (SBGN) (Le Novère et al., 2009) pour représenter les mécanismes de biologie moléculaire. Le logiciel crée une représentation structurée en langage de balisage de la biologie des systèmes (SBML), adaptée à une analyse informatique ultérieure (Le Novère et al., 2009). Chaque entité de la carte est annotée par des références correspondantes et

des notes spécifiques. Dans la mesure du possible, les références des articles *princeps* sont ajoutées.

La carte de mort cellulaire régulée a été élaborée selon une méthodologie décrite précédemment (Kondratova et al., 2018; Kuperstein et al., 2015). Les cartes sont dessinées en utilisant le langage *Process Description* de la syntaxe *Systems Biology Graphical Notation* (SBGN). L'unité élémentaire de la carte est une réaction biochimique, formée de réactifs, de produits et de divers types de régulateurs (Figure 5. Représentation d'un modèle simplifié en utilisant la notation SBGN. .

#### b. Règles de sélection de la littérature

Les règles de sélection de la littérature sont les suivantes : cinq à dix articles de synthèse dans le domaine ont d'abord été choisis pour extraire les voies de consensus et les réglementations acceptées dans le domaine et pour approcher la littérature suggérée dans ces synthèses. Ces articles ont permis de construire la « colonne vertébrale » du réseau et de définir les limites des processus biologiques (Figure 15. Processus de construction de la carte de la mort cellulaire régulée.). En outre, des informations provenant d'articles originaux, de préférence des études les plus récentes, sont analysées et ajoutées afin d'élargir et d'enrichir le réseau. Dans la majorité des cas, la décision d'ajouter un événement ou une régulation biochimique ou un processus à la carte doit être étayée par au moins deux études indépendantes réalisées par des équipes différentes.

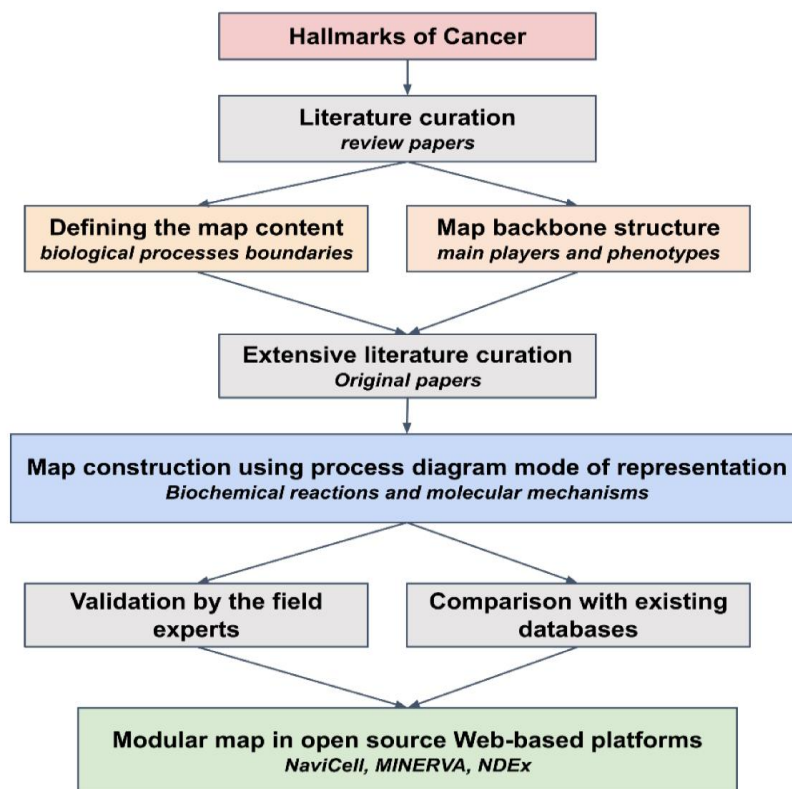



Figure 15. Processus de construction de la carte de la mort cellulaire régulée.










Le schéma décrit les différentes étapes de la construction de la carte. Le travail débute par la collecte d'informations spécifiques au cancer et liées à la mort cellulaire régulée (RCD). L'objectif est de déterminer les interactions - au niveau moléculaire - à partir de publications scientifiques et de bases de données. Ensuite, la carte est construite d'après ces informations et chaque réaction annotée en organisant les interactions en plusieurs modules. Finalement, la carte est validée par les experts du domaine, comparée avec d'autres bases de données et éventuellement complétée avant d'être mise en ligne.

### c. Format d'annotation de la carte

Chaque entité et réaction de la carte est annotée à l'aide du format d'annotation NaviCell (Figure 16. Page d'annotation d'une entité de la carte RCD au format NaviCell).

**Protein Caspase9\*** 

**Identifiers**  
 caspase 9, apoptosis-related cysteine peptidase  
[HUGO:CASP9](#) [HGNC:1511](#) [ENTREZ:842](#) [UNIPROT:P55211](#)

**Maps\_Modules**  
 MODULE:ANTIOXIDANT\_RESPONSE   
 MODULE:APOPTOSIS   
 MODULE:STARVATION\_AUTOPHAGY   
 MODULE:CASPASES   
 MODULE:DEATH\_RECEPTOR\_PATHWAYS   
 MODULE:DEPENDANCE\_RECEPTORS   
 MODULE:MOMP\_REGULATION   
 MODULE:RCD\_GENES   
 MODULE:ER\_STRESS 

**References**  
[PMID:19931333](#)  
[PMID:26618107](#)  
[PMID:12011067](#)  
 DCC interacts with both caspase-3 and caspase-9 and drives the activation of caspase-3 through caspase-9  
[PMID:19465923](#)  
 The Patched dependence receptor triggers apoptosis through a DRAL-caspase-9 complex.  
[PMID:22679284](#)  
 Patched dependence receptor triggers apoptosis through ubiquitination of caspase-9.

Figure 16. Page d'annotation d'une entité de la carte RCD au format NaviCell

L'annotation comprend trois sections : « Identifiers », « Maps\_Modules » et « References ». La section « Identifiers » fournit des liens vers les descriptions des entités correspondantes dans les bases de données HGNC, UniProt, Entrez, GeneCards, REACTOME, KEGG et Wikipedia. La section « Maps\_Modules » contient des liens vers les modules dans lesquels l'entité est trouvée. La section « References » comprend des notes ajoutées manuellement et des liens vers les publications pertinentes dans PubMed. Les annotations peuvent être associées à une entité moléculaire (telle qu'une protéine) ainsi qu'à sa modification particulière (telle qu'une modification post-traductionnelle particulière). NaviCell fournit des mécanismes de référencement croisé de la carte RCD avec d'autres bases de données moléculaires et de voies, telles que KEGG PATHWAY, Reactome, Atlas of Genetics and Cytogenetics in Oncology and Haematology, GeneCards et Wiki Genes.

d. Génération de la carte NaviCell avec NaviCell Factory

Un système d'étiquetage systématique a été appliqué pour l'annotation de chaque entité avec des étiquettes de nom de module. Les étiquettes sont converties en liens par *NaviCell Factory* au cours du processus de génération de la version en ligne de la carte.

La carte de CellDesigner annotée au format NaviCell est convertie en une interface Web NaviCell, qui est un ensemble de pages html avec un code JavaScript intégré qui peut être lancé dans un navigateur Web pour une utilisation en ligne. Les identifiants HGNC dans le formulaire d'annotation permettent d'utiliser l'outil NaviCell pour la visualisation des données omiques. Un guide détaillé de l'utilisation de l'usine NaviCell intégrée dans le plugin Cytoscape de BiNoM (Bonnet et al., 2013) est fourni à l'adresse <https://navicell.curie.fr/doc/NaviCellMapperAdminGuide.pdf>.

#### e. Navigation dans la carte

Le contenu de la carte de la mort cellulaire régulée est fourni sous la forme d'une carte interactive. La visualisation et l'exploration de la carte RCD sont facilitées par l'environnement Web NaviCell (<https://navicell.curie.fr>). NaviCell utilise le moteur et les principes de Google Maps pour naviguer dans le contenu de la carte. Ainsi, la logique de navigation (défilement, zoom) et les caractéristiques (marqueurs, fenêtres d'appel et barre de zoom) sont reprises de l'interface de Google Maps™. Une recherche peut être effectuée à l'aide de la fenêtre de recherche ou en sélectionnant les entités d'intérêt dans le panneau de sélection. Le zoom dans NaviCell simplifie la navigation dans les grandes cartes d'interactions moléculaires en offrant plusieurs niveaux d'affichage de la carte. Pour faciliter l'exploration, chaque niveau de zoom sur la carte expose une certaine profondeur de détail. En outre, la navigation est également simplifiée par la structure modulaire de

la carte où chaque module peut être visualisé individuellement. Le passage de la carte globale aux cartes modulaires se fait à l'aide de l'icône de la carte (globe) (Figure 17. Navigation dans la carte de la mort cellulaire régulée, zooms et annotations des entités.).

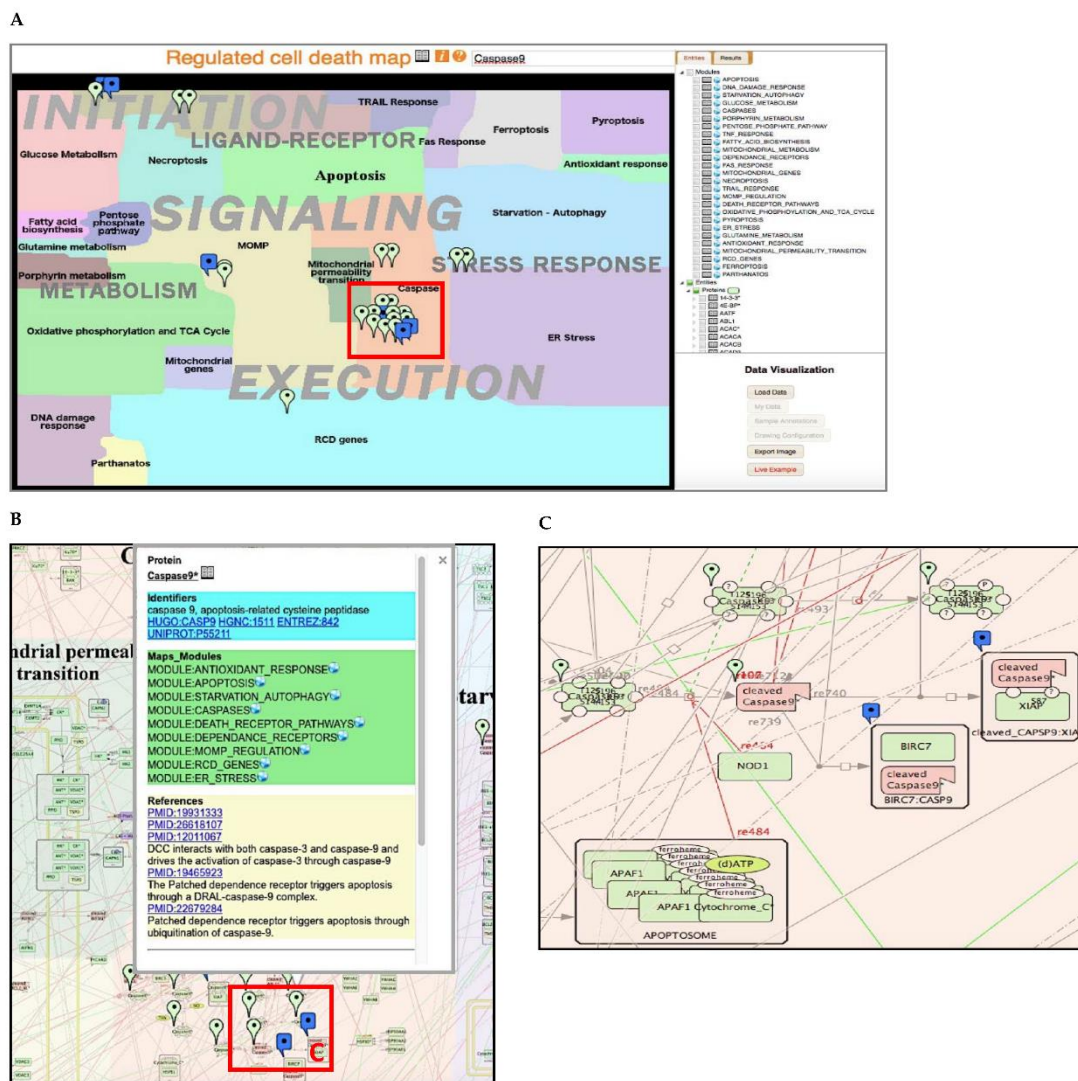


Figure 17. Navigation dans la carte de la mort cellulaire régulée, zooms et annotations des entités.

(A) Interface cartographique dans la plateforme Google Maps alimentée par NaviCell avec une vue de dessus de la carte de la mort cellulaire régulée. L'interface comprend la fenêtre de la carte, le panneau de sélection, le panneau d'analyse des données et le panneau supérieur. L'interrogation de la carte est possible via la fenêtre de recherche ou en cochant l'entité dans une liste d'entités dans le panneau de sélection qui laissera tomber des marqueurs sur toute la carte (par exemple, Caspase 9). (B) Zoom d'un fragment de la carte et fenêtre d'appel. Un clic sur un marqueur ouvre une fenêtre d'appel contenant trois sections : « Identifiants » avec des liens vers des bases de données externes ; « Modules de la carte » avec des liens vers les modules fonctionnels où se trouve l'entité d'intérêt ; « Références » avec des liens vers PubMed et des notes d'annotation. Un clic sur l'icône « globe » ouvre la carte correspondante. Un clic sur une icône "livre" ouvre une page d'annotation détaillée. (C) Zoom d'un fragment de module montrant le niveau le plus détaillé de la

représentation des mécanismes moléculaires. La participation de la molécule sélectionnée (par exemple, ici la caspase 9) à diverses réactions et complexes est visible.

#### 4. Présentation de la carte

##### a. Caractéristiques générales de la carte

La mort cellulaire régulée est caractérisée par de nombreuses interactions entre les mécanismes moléculaires. Afin de représenter ces détails de manière organisée et lisible, nous avons traité cette complexité de deux manières. Premièrement, la carte a une structure hiérarchique et est construite autour de trois « couches » successives (initiation, signalisation, exécution). La carte est également divisée en méta-modules, puis en modules fonctionnels et, enfin, au niveau le plus détaillé, de réactions biochimiques (Figure 17. Navigation dans la carte de la mort cellulaire régulée, zooms et annotations des entités. et Tableau 1. Structure et contenu de la carte RCD.). Deuxièmement, chaque entité est étiquetée avec les modules auxquels elle appartient. Ce système permet de retracer la participation des entités à différents processus et aide à retrouver une structure de base du diagramme de réseau.

La carte RCD rassemble au total 1008 protéines, 260 gènes, 93 miRNA et 2020 réactions. Elle est basée sur 800 publications scientifiques originales. La carte est divisée en trois méta-modules et 26 modules fonctionnels distincts (Tableau 1. Structure et contenu de la carte RCD.).

##### b. Structure de la carte de la mort cellulaire régulée

La mort cellulaire régulée est un processus dynamique, dans lequel la propagation et la durée des signaux déterminent le destin des cellules (« survie » ou « destruction »). Différents acteurs peuvent partager les mêmes modulateurs ou les mêmes cibles. Le processus de RCD peut être divisé en trois étapes principales, l'initiation, la signalisation et l'exécution, que nous avons définies comme les couches de la carte (Figure 17. Navigation dans la carte de la mort cellulaire régulée, zooms et annotations des entités. et Tableau 1. Structure et contenu de la carte RCD.).

Tableau 1. Structure et contenu de la carte RCD.

<i>Layer/Module</i>	Chemical Species (Entities)	Proteins	Genes	RNAs	asRNAs	Reactions
<i>Initiation (reversible)</i>						
<i>STRESS RESPONSE</i>						
<i>ANTIOXIDANT RESPONSE</i>	213	151	4	9	12	95
<i>DNA DAMAGE RESPONSE</i>	128	59	3	3	5	80
<i>ER STRESS</i>	382	161	45	44	4	291
<i>STARVATION-AUTOPHAGY</i>	438	165	19	31	31	290
<i>LIGAND RECEPTOR</i>						
<i>DEATH RECEPTOR PATHWAYS</i>	640	287	23	27	32	469
<i>TRAIL RESPONSE</i>	36	25	0	0	0	18
<i>FAS RESPONSE</i>	29	24	0	0	0	14
<i>TNF RESPONSE</i>	87	39	3	3	0	57
<i>DEPENDENCE RECEPTORS</i>	78	52	0	0	0	36
<i>METABOLISM</i>						
<i>CELL METABOLISM</i>						
<i>FATTY ACID BIOSYNTHESIS</i>	78	34	1	3	2	33
<i>GLUCOSE METABOLISM</i>	190	112	0	0	0	104
<i>GLUTAMINE METABOLISM</i>	42	20	0	0	0	20
<i>PENTOSE PHOSPHATE PATHWAY</i>	64	12	2	2	2	39
<i>PORPHYRIN METABOLISM</i>	51	17	0	0	0	26
<i>MITOCHONDRIAL METABOLISM</i>						
<i>OXIDATIVE PHOSPHORYLATION AND TCA CYCLE</i>	232	179	0	0	0	109
<i>MITOCHONDRIAL GENES</i>	114	116	2	30	0	36
<i>Signaling (rewirable)</i>						
<i>APOPTOSIS</i>	584	246	34	40	57	401
<i>NECROPTOSIS</i>	242	118	0	0	0	188
<i>FERROPTOSIS</i>	193	65	17	19	0	110
<i>PARTHANATOS</i>	80	17	1	0	0	64
<i>PYROPTOSIS</i>	112	36	1	0	0	79
<i>Execution (irreversible)</i>						
<i>MOMP REGULATION</i>	630	319	19	48	17	396
<i>MITOCHONDRIAL PERMEABILITY TRANSITION</i>	57	36	0	0	0	33
<i>CASPASES</i>	318	158	5	5	12	221
<i>RCD GENES</i>	633	135	168	179	90	365
<i>RCD GLOBAL MAP</i>	2657	1008	215	260	93	2020

La couche « Initiation » décrit les signaux d'entrée et les mécanismes qui déclenchent la mort cellulaire régulée. Les trois principales entrées sont regroupées dans les modules « Ligand-Récepteur », « Métabolisme » et « Réponse au stress ». La couche « Signalisation » est le niveau où la décision entre la mort cellulaire et la survie est prise, ainsi que la sélection du type de mort cellulaire. Elle regroupe tous les modes de mort cellulaire précédemment décrits et leurs processus d'inhibition. C'est aussi l'étape d'interaction où les « déclencheurs » transmettent leur signal à un messager secondaire en fonction du contexte cellulaire. Ainsi, un même élément déclencheur peut provoquer différents types de résultats en fonction de l'absence ou de la présence d'autres entités. La couche « Exécution » représente l'étape finale au cours de laquelle le processus dépasse le point de non-retour et les composants vitaux de la cellule sont dégradés de manière irréversible. La perméabilisation de la membrane externe des mitochondries (MOMP), la transition de perméabilité des mitochondries (MPT) et les caspases sont les principaux acteurs de cette étape. Chaque module de la couche « signalisation » déclenche une combinaison spécifique de modules dans la couche « exécution ». Par exemple, l'initiation de la parthanatose déclenche une fragmentation de l'ADN à grande échelle, tandis que le déclenchement de l'apoptose conduit préférentiellement à l'activation des caspases et à la MOMP.

En outre, la structure modulaire de la carte permet d'explorer chaque mécanisme moléculaire séparément, sous la forme de cartes indépendantes. La carte est en effet divisée en méta-modules fonctionnels, eux-mêmes subdivisés en modules fonctionnels plus petits, créant ainsi 26 modules fonctionnels. Par exemple, le méta-module traitant des "voies des récepteurs de la mort" contient des informations sur la réponse à TRAIL, la ligature des récepteurs Fas, la réponse au TNF et la signalisation des récepteurs de dépendance (Tableau 1. Structure et contenu de la carte RCD.).

### c. Contenu de la carte de la mort cellulaire régulée

La carte rassemble des informations sur toutes les principales voies de mort cellulaire régulée, facilitant ainsi l'illustration de leurs interactions.

L'apoptose est classiquement divisée en « apoptose extrinsèque », représentée à gauche de la Figure 18. Représentation simplifiée de l'apoptose, et en « apoptose intrinsèque », à droite (Galluzzi et al., 2018).

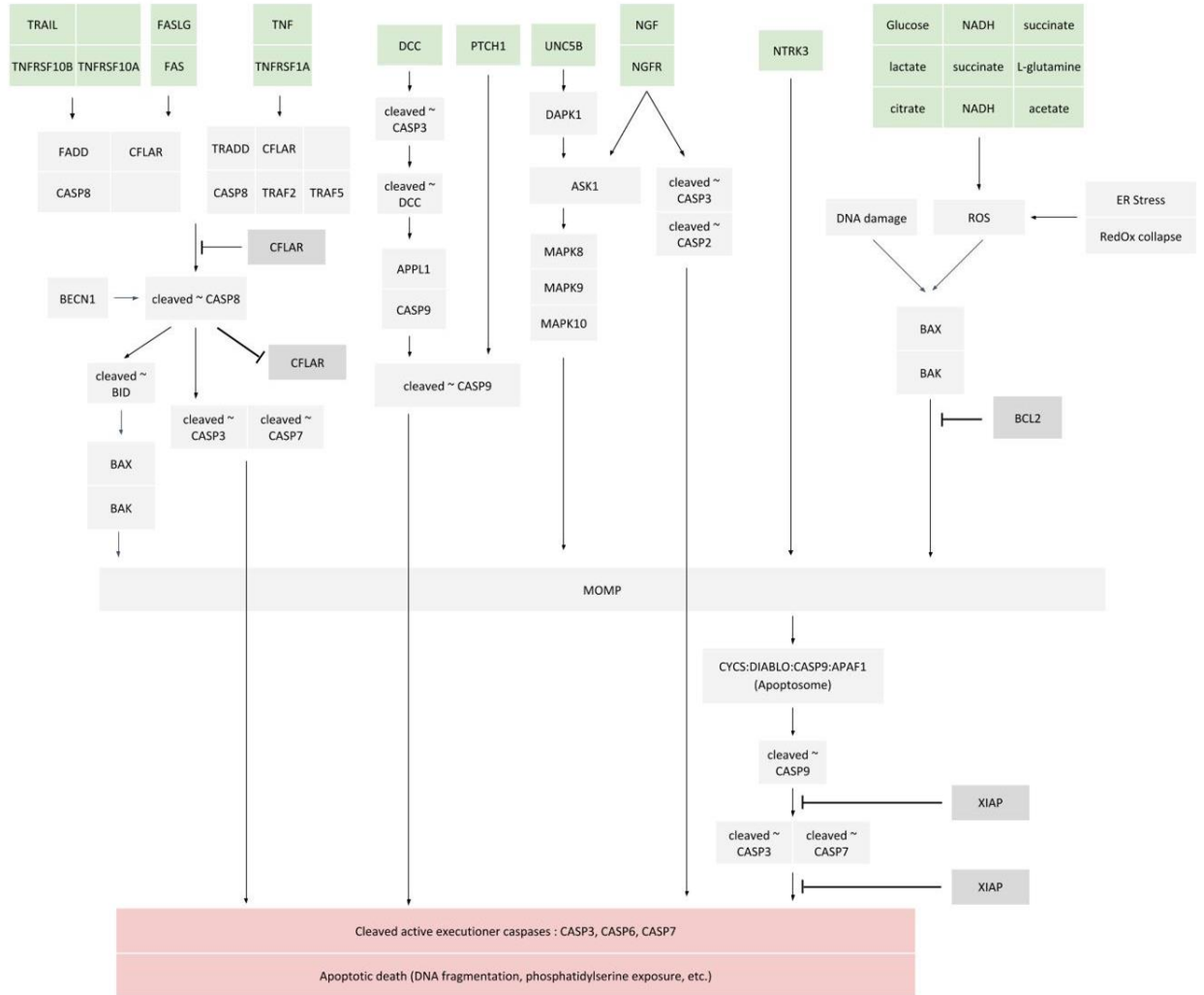


Figure 18. Représentation simplifiée de l'apoptose

Représentation des modes de mort cellulaire régulée sous forme de diagrammes de flux d'activité comprenant les initiateurs (en vert), les acteurs clés et les effecteurs (en rouge).

L'apoptose extrinsèque est initié par les récepteurs de mort (*death receptors*), dont la liaison au ligand déclenche le signal, et les récepteurs de dépendance (*dependence receptors*), dont le retrait du ligand active le signal (Ashkenazi & Dixit, 1998). Ce signal d'initiation conduit à une activation rapide des caspases (en particulier CASP3 et CASP7) et finalement à la mort apoptotique (Dickens et al., 2012). L'apoptose intrinsèque conduit de préférence à la perméabilisation de la membrane externe de la mitochondrie (MOMP), souvent en raison de l'oligomérisation de BAX et BAK ensemble (Czabotar et al., 2014) et de leur insertion dans la membrane mitochondriale. L'activation CASP9-dépendante

de l'apoptosome dans le cytoplasme résulte de la libération mitochondriale de CYCS (anciennement connu sous le nom de cytochrome c) et de SMAC (également connu sous le nom de DIABLO). L'activation de CASP9 conduit à l'activation protéolytique de CASP3 et CASP7 et à l'apoptose.

Le déclenchement de la nécroptose est guidé par les protéines RIPK1 et RIPK3 (Vandenabeele et al., 2010). Lorsque CASP8 est inhibé, en particulier par c-FLIP (codé par *CFLAR*), RIPK3 est activé par RIPK1. Le recrutement de RIPK3 et de MLKL dans le complexe du nécrosome déclenche les caractéristiques de la nécroptose (augmentation du Ca<sup>2+</sup> cytosolique, production de ROS, suractivation de PARP1 et perméabilisation de la membrane du lysosome) et finalement la mort par nécroptose (Figure 19. Nécroptose.).

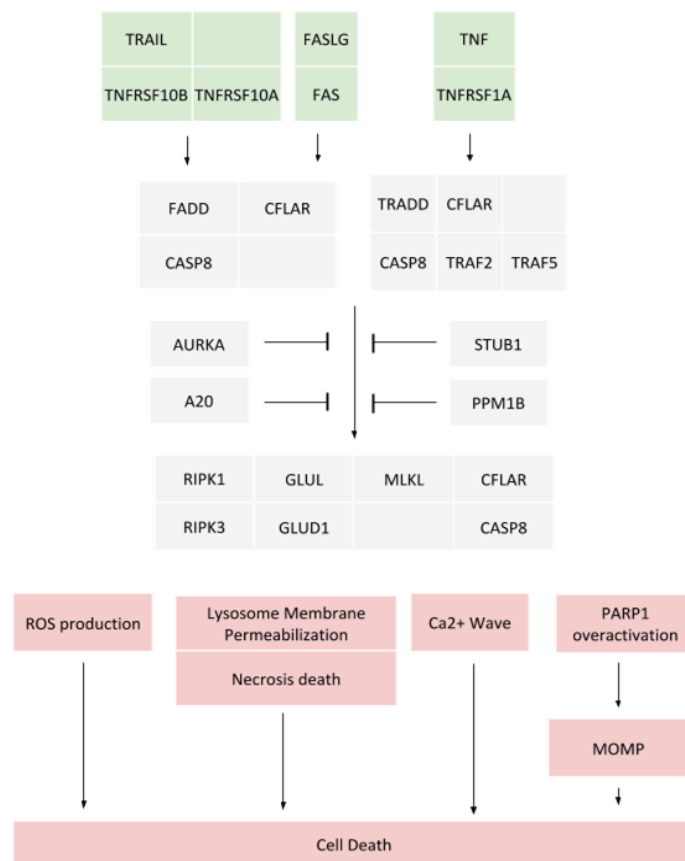


Figure 19. Nécroptose.

Représentation des modes de mort cellulaire régulée sous forme de diagrammes de flux d'activité comprenant les initiateurs (en vert), les acteurs clés et les effecteurs (en rouge).

La mort cellulaire dépendante de l'autophagie et des lysosomes est influencée par l'état énergétique de la cellule et couplée à la formation d'autophagosomes et d'autophagolysosomes (Galluzzi et al., 2012). Les ROS, la privation d'acides aminés et les lésions de l'ADN entraînent une diminution du 5' adénosine triphosphate (ATP) cytosolique et une augmentation conséquente du 5' adénosine monophosphate (AMP), déclenchant ainsi l'activation de la kinase dépendante de l'AMP (AMPK). Cela déclenche (1) l'inhibition du complexe pro-survie mTORC1 ; (2) l'activation de BECN1 (Beclin 1) et entraîne la formation d'autophagosomes (Figure 20. Autophagie et mort cellulaire liée au lysosome.).

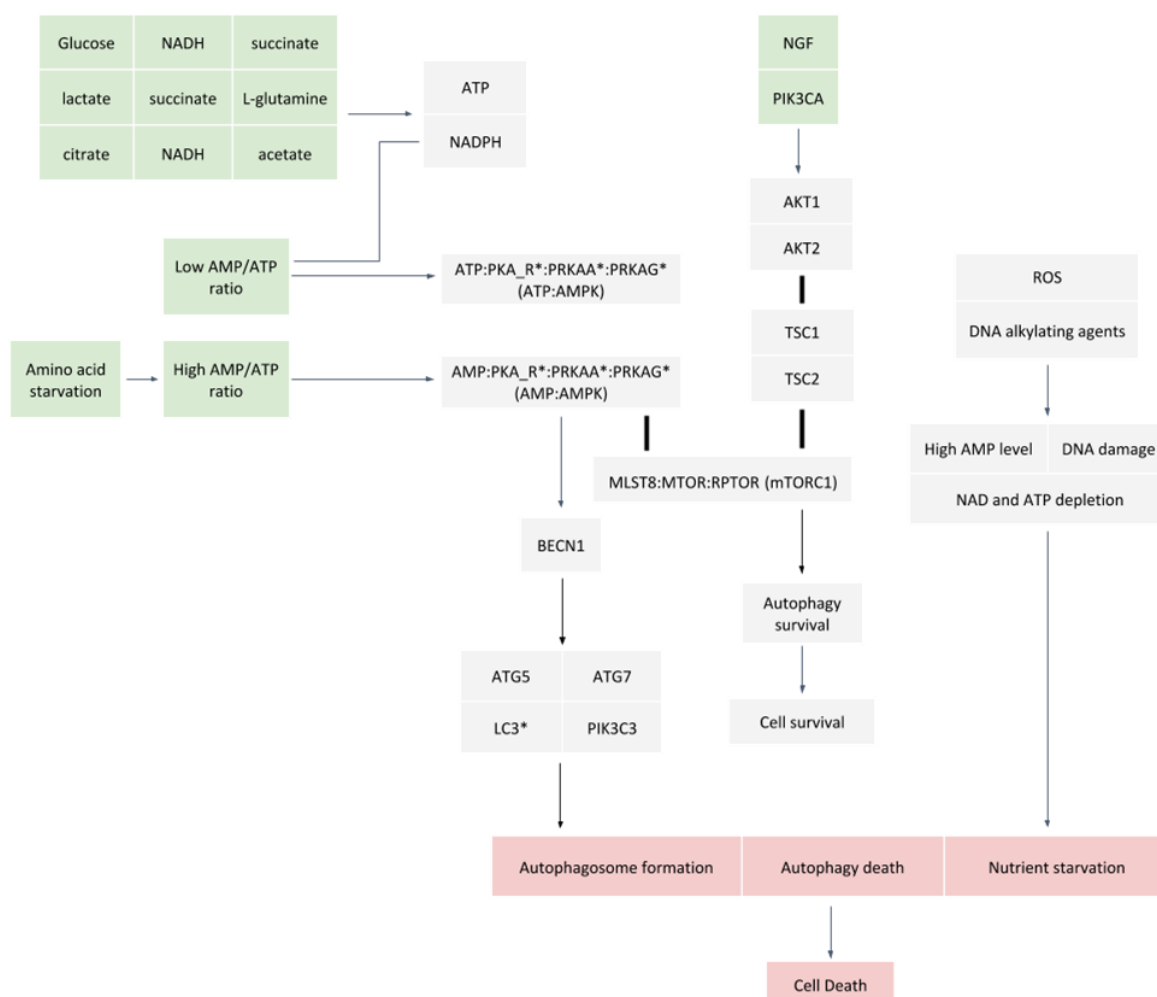


Figure 20. Autophagie et mort cellulaire liée au lysosome.

Représentation des modes de mort cellulaire régulée sous forme de diagrammes de flux d'activité comprenant les initiateurs (en vert), les acteurs clés et les effecteurs (en rouge).

La caractéristique principale de la parthanatose est la suractivation de PARP1 suite à des lésions de l'ADN par alkylation ou à l'exposition à des ROS (David et al., 2009; Fatokun et al., 2014). L'activation excessive de PARP1 conduit à l'accumulation de

poly(ADP-ribose) (PAR), déclenchant *in fine* la libération d'AIMF1 des mitochondries et la fragmentation de l'ADN par AIFM1. PARP1 est également responsable de l'épuisement du NAD<sup>+</sup> et de l'ATP, ce qui provoque un effondrement fatal du stock d'énergie cellulaire et du métabolisme redox (Figure 21. Parthanatose).

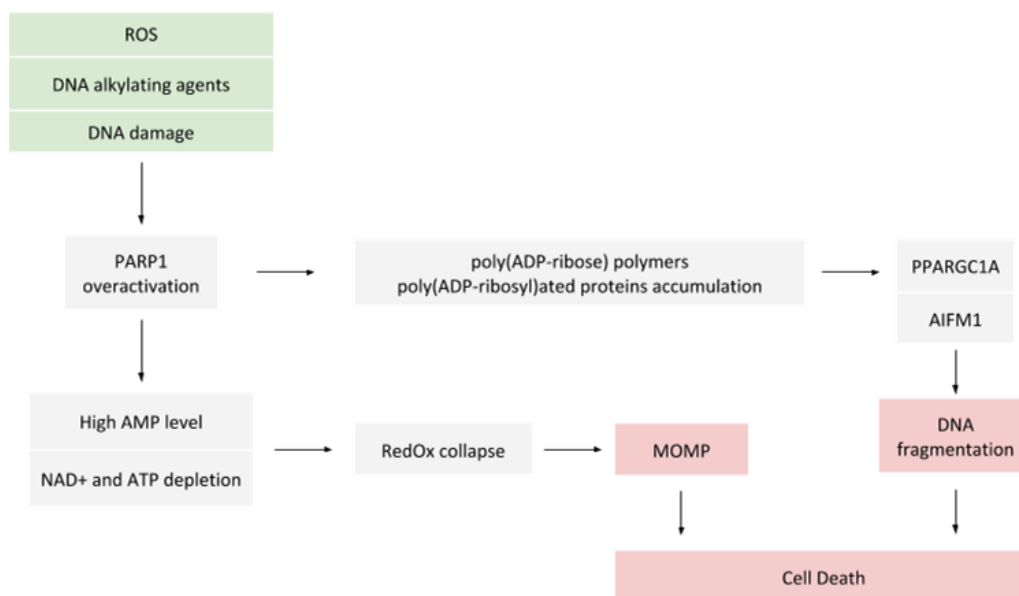


Figure 21. Parthanatose.

Représentation des modes de mort cellulaire régulée sous forme de diagrammes de flux d'activité comprenant les initiateurs (en vert), les acteurs clés et les effecteurs (en rouge).

La pyroptose est déclenchée par des perturbations de l'immunité innée (Mamik & Power, 2017; Shi et al., 2017). Des signaux variés, tels que des produits viraux et bactériens (les lipopolysaccharides, LPS, est l'un des principaux signaux), déclenchent la formation d'inflammasomes suivie de l'activation de caspases spécifiques, notamment CASP1, CASP4 et CASP5. Ces caspases clivent des substrats tels que la Gasdermin D (GSDMD), ce qui conduit à la pyroptose. De plus, les interleukines pro-inflammatoires IL1-Beta et IL-18 sont également souvent sécrétées (Figure 22. Pyroptose (à gauche), ferroptose (au centre) et nécrose liée à la perméabilité mitochondriale).

La ferroptose repose sur la dérégulation du système redox (Latunde-Dada, 2017; Magtanong et al., 2016). Lorsque les enzymes antioxydantes, principalement la Glutathion peroxydase 4 (GPX4), sont dépassées, les lipides membranaires sont oxydés en peroxydes, ce qui entraîne des altérations fatales de la perméabilité membranaire, des vagues de calcium, des modifications du potentiel membranaire et la mort cellulaire (Figure 22.

Pyroptose (à gauche), ferroptose (au centre) et nécrose liée à la perméabilité mitochondriale).

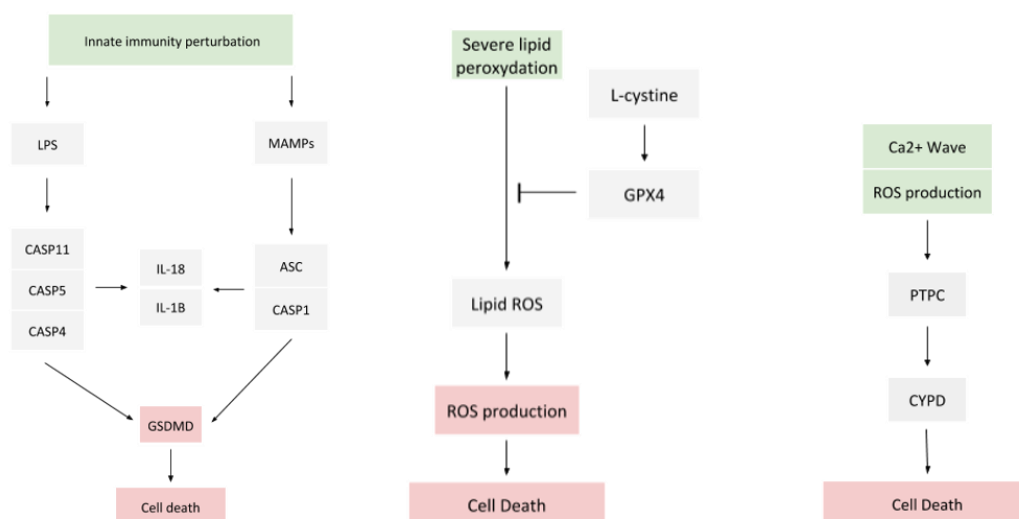


Figure 22. Pyroptose (à gauche), ferroptose (au centre) et nécrose liée à la perméabilité mitochondriale (à droite).

Représentation des modes de mort cellulaire régulée sous forme de diagrammes de flux d'activité comprenant les initiateurs (en vert), les acteurs clés et les effecteurs (en rouge).

#### d. L'interaction entre les modes de mort cellulaire et les points de commutation

À partir du réseau détaillé et des modules de la carte, il est possible de déduire un schéma des principaux acteurs et de l'interaction entre eux (Figure 23. Interaction entre les modes de mort cellulaire régulée). Pour simplifier et clarifier le message, la signalisation est représentée sous la forme d'un diagramme de flux d'activité indiquant soit l'activation, soit l'inhibition. Les entrées au sommet de la couche d'initiation dictent le choix de la voie. Cependant, la régulation de la voie choisie n'est pas triviale, car le même régulateur peut influencer différents modes de mort cellulaire, en fonction de la présence d'activateurs et d'inhibiteurs. Il est également important de noter que la formation de complexes de haut poids moléculaire est une étape essentielle pour l'initiation de tous les modes de mort cellulaire. De plus, une coopération entre les modes de mort cellulaire peut se produire. Une fois le choix du mode de mort cellulaire effectué, la signalisation se propage à la couche d'exécution, aboutissant à différentes formes de mort cellulaire, chacune ayant un phénotype particulier (Figure 23. Interaction entre les modes de mort cellulaire régulée).

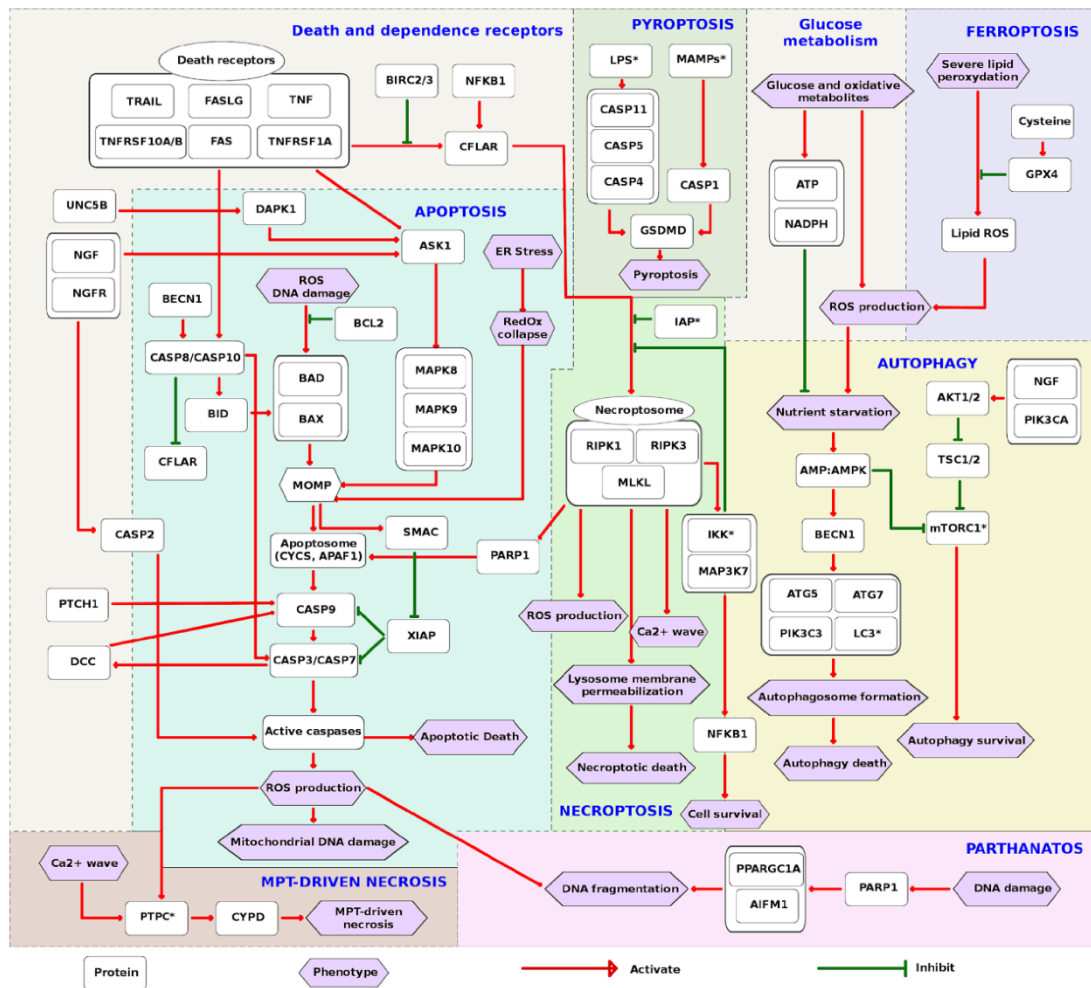


Figure 23. Interaction entre les modes de mort cellulaire régulée

Les molécules d'entrée, les principaux acteurs et les phénotypes résultants ont été sélectionnés pour chaque module de la carte RCD. Le choix a été basé sur le nombre d'interactions sur la carte et sur des informations tirées de la littérature. Chaque carte de module a été élaguée manuellement, tout en gardant intacts les acteurs principaux, les molécules fonctionnellement importantes et leurs interactions correspondantes. Les graphiques résultants ont été intégrés manuellement et l'interaction entre les modes de mort cellulaire a été ajoutée. Le diagramme a été converti en diagramme de flux d'activité en remplaçant les arêtes de réaction par des arêtes d'influence, soit d'activation, soit d'inhibition, afin d'obtenir la présentation finale. Certaines entités moléculaires sont regroupées en une entité générique représentant soit un complexe moléculaire, soit une cascade de réactions. La nomenclature HGNC a été utilisée pour nommer les molécules, à l'exception des entités marquées par des étoiles, pour lesquelles des noms communs ont été utilisés. Les modes RCD sont mis en évidence par le fond coloré semi-transparent. La plupart des voies sont tracées du haut (initiation) vers le bas de la figure (exécution). Certaines incidences de la diaphonie entre les voies sont clairement illustrées. La production de ROS est une entrée et une sortie cruciales du RCD. Le récepteur de mort peut déclencher une grande variété de réponses, même la survie cellulaire (via l'induction de IAP, NFKB1). L'apoptose, la nécroptose et l'autophagie sont des processus interconnectés (grâce aux implications multifonctionnelles de CFLAR, PARP1, IAP, BECN1). L'approvisionnement en énergie est un régulateur majeur de la survie de la cellule. IAP : protéines inhibitrices de l'apoptose ; LPS : lipopolysaccharides ; MAMPs : microbe-associated molecular patterns ; PTPC : permeability transition complex ; MPT : mitochondrial permeability transition ; LC3 : microtubule-associated protein 1 light chain 3 beta (MAP1LC3B).

L'interaction entre les ligands et les récepteurs de mort (*Death receptor*) aboutit à l'initiation de la signalisation. En effet, après interaction du TNF avec son récepteur, RIPK1 et CFLAR sont recrutés dans le complexe de signalisation des récepteurs de mort (DISC) grâce à TRADD ainsi que CASP8. Trois voies principales émergent, en fonction des acteurs recrutés et du degré d'oligomérisation de CASP8 : l'apoptose, la nécroptose et la signalisation pro-survie. CFLAR empêche le déclenchement de l'apoptose. Dans cette situation, RIPK1 est activé par polyubiquitination. RIPK1 polyubiquitaire active ensuite IKK et MAP3K7, ce qui entraîne l'activation du facteur de transcription pro-survie NF- $\kappa$ B. De plus, la phosphorylation de RIPK1 par IKK et MAP3K7 empêche son interaction avec FADD et CASP8, favorisant ainsi l'apoptose indépendante de RIPK1 (Geng et al., 2017). Les membres de la famille des protéines inhibitrices de l'apoptose (IAP) jouent également un rôle important dans ces interactions. Par exemple, une IAP appelée XIAP inhibe les caspases de manière constitutive à l'état physiologique. D'autres IAP comme BIRC2 et BIRC3 conduisent à une augmentation des facteurs anti-apoptotiques et rendent ubiquitaires également RIPK1, déclenchant ainsi une signalisation NF- $\kappa$ B pro-survie (Silke & Vucic, 2014; Witt & Vucic, 2017). Cependant, certaines protéines inhibitrices de l'apoptose (IAP) provoquent la dé-ubiquitination de RIPK1 dépendante du CYLD, ce qui entraîne la libération de RIPK1 du complexe du nécrosome, et conduit à l'apoptose extrinsèque (Hitomi et al., 2008).

L'apoptose dépend fortement de l'apport d'énergie. Par conséquent, en l'absence d'ATP, les cellules ont tendance à subir une mort nécrotique plutôt qu'apoptotique. Dans des circonstances spécifiques, lorsque l'AMPK est activée, BECN1 peut favoriser le clivage de CASP8 et l'activation de l'apoptose. Il est intéressant de noter que les mécanismes apoptotiques et nécrotiques peuvent être activés conjointement si une combinaison de stimuli est appliquée à la cellule ou si la signalisation apoptotique est activée alors que les mitochondries sont perméabilisées et libèrent des facteurs qui favorisent la nécroptose, tels que le CYPD (Galluzzi et al., 2014). L'autophagie maintient normalement l'état énergétique de la cellule (Chaabane et al., 2013; Galluzzi et al., 2012), mais une autophagie prolongée peut entraîner la mort cellulaire (Tsuchihashi et al., 2015).

Les espèces oxydantes réactives (ROS) sont largement impliquées dans la plupart des modes de RCD. Comme elles peuvent agir à la fois en tant qu'initiateur et exécuteur, en fonction de la sous-routine RCD, les ROS ont tendance à augmenter les interactions entre les voies de mort cellulaire. L'accumulation de ROS favorise le recrutement de BAX/BAK dans les mitochondries et agit donc comme un initiateur majeur de l'apoptose intrinsèque. Les ROS sont également essentiels à la libération du cytochrome C (CYCS) des mitochondries. De même, la parthanatose peut être la conséquence d'un stress oxydatif prolongé. Bien que le mécanisme précis reste inconnu, il a été démontré que l'activation des caspases génère des ROS via la chaîne de transport d'électrons mitochondriale (K. Choi et al., 2010). De même, la ferroptose est basée sur la génération d'une quantité excessive de ROS.

e. Comparaison de la carte RCD avec des ressources de voies similaires

Un certain nombre de mécanismes de mort cellulaire régulés sont décrits dans plusieurs ressources de voies existantes telles que les bases de données de voies KEGG (M. Kanehisa et al., 2012) et REACTOME (Croft et al., 2014).

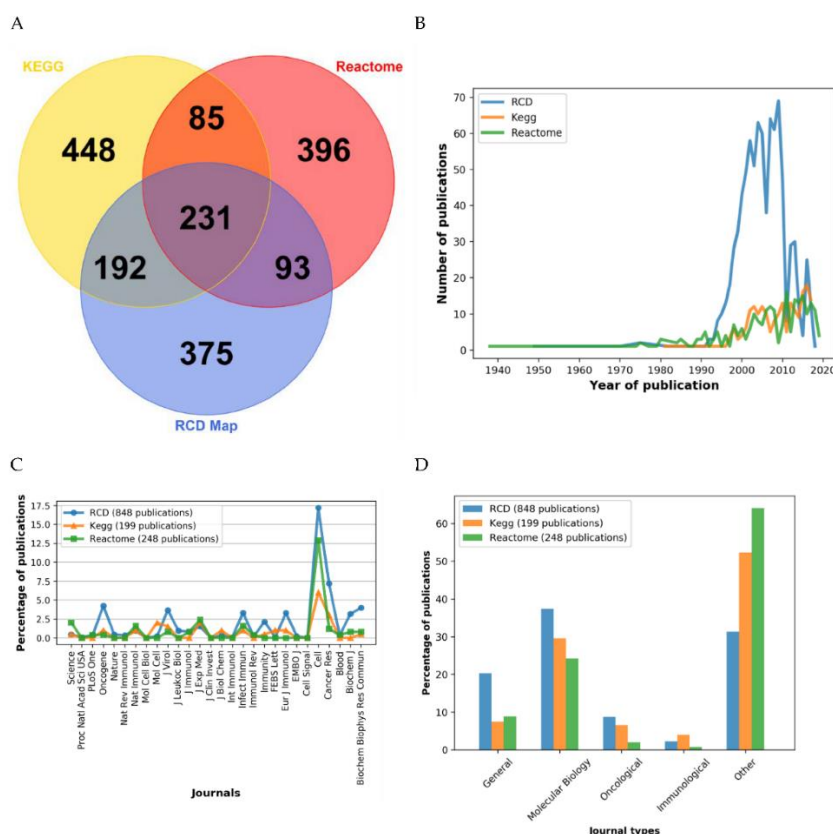


Figure 24. Comparaison de la carte RCD avec d'autres ressources sur les voies.

(A) Comparaison du contenu des noms HUGO dans les voies sélectionnées dans les bases de données REACTOME et KEGG avec la carte RCD. Le diagramme de Venn montre le chevauchement des noms HUGO entre les différentes bases de données. Distribution (B) des années de publication et (C) des revues annotant les voies sélectionnées dans les bases de données REACTOME, KEGG et RCD. (D) Distribution relative des types de journaux utilisés pour l'annotation des voies sélectionnées dans REACTOME, KEGG et RCD.

Il existe d'autres ressources liées à la mort cellulaire dans ce domaine, notamment DeathBase (<http://www.deathbase.org>), une base de données sur les protéines impliquées dans la mort cellulaire (Doctor et al., 2003) et ApoptoProteomics (<http://apoptoproteomics.uio.no>), une base de données intégrée pour l'analyse des données protéomiques qui contient également des informations précieuses sur la correspondance entre les protéines de la mort cellulaire et les médicaments anticancéreux (Arntzen & Thiede, 2012). Enfin, il existe une plateforme en ligne spécifique à la mort cellulaire programmée de la levure (<http://ycelldeath.com/index.html>) qui recueille divers types d'informations sur les trois sortes de mort cellulaire programmée de la levure, à savoir l'apoptose (Wanichthanarak et al., 2013), la nécrose et l'autophagie. Les référentiels de voies spécifiques à un type de mort cellulaire sont rares. Jusqu'à présent, il n'existe que la base de données *Autophagy Regulatory Network* qui rassemble des informations sur les interactions protéine-protéine liées au processus d'autophagie et qui relie également ces composants et régulateurs aux voies de signalisation déjà existantes dans les ressources publiques (Türei et al., 2015). À notre connaissance, il n'existe pas d'autres bases de données spécifiques à d'autres processus de mort cellulaire.

La majorité des ressources susmentionnées ne représentent pas les processus de mort cellulaire sous la forme de diagrammes de description des processus. En fait, les bases de données les plus proches du mode de représentation de la carte RCD et donc les plus comparables, sont KEGG et REACTOME, où les voies sont représentées à l'aide d'une syntaxe similaire à celle utilisée pour générer la carte RCD. La carte RCD a été comparée aux voies pertinentes décrivant les processus liés à la mort cellulaire à partir des bases de données de voies KEGG et REACTOME. Au total, ces voies sélectionnées couvrent 956 noms de gènes uniques dans KEGG et 805 dans REACTOME (Figure 24. Comparaison de la carte RCD avec d'autres ressources sur les voies.).

Certains modules représentent les processus à un niveau plus détaillé en termes de nombre de noms de gènes, par rapport aux voies correspondantes des bases de données KEGG et REACTOME. Par exemple, la voie de la mort cellulaire programmée dans REACTOME contient 178 gènes, ce qui correspond aux modules Apoptose et Nécroptose dans la carte RCD, qui couvrent ensemble 284 gènes. En revanche, le module de nécroptose est légèrement plus enrichi dans la voie KEGG que dans la carte RCD (165 dans KEGG contre 154 dans ACSN). Cependant, dans KEGG, la voie parthanatose (notamment les groupes de gènes d'histones) est incluse dans le module de nécroptose. Au contraire, dans la carte RCD, la voie parthanatose est décrite comme un module indépendant et est mise à jour avec les publications les plus récentes relatives au cancer. Certaines voies métaboliques sont plus représentées dans d'autres bases de données que dans la carte RCD. C'est le cas de la voie des pentoses phosphates et du métabolisme des porphyrines (qui comprend en fait le métabolisme de la chlorophylle) dans KEGG, ainsi que du métabolisme des acides gras dans REACTOME (<https://www.mdpi.com/2072-6694/12/4/990/s1>). Cependant, il est important de souligner que les réactions métaboliques n'entrent pas dans le champ d'application de la carte RCD à moins qu'elles ne modulent des processus liés à la mort cellulaire.

323 des 805 gènes des voies REACTOME et 423 des 956 gènes de la voie KEGG sont également présents dans la carte RCD (Figure 24. Comparaison de la carte RCD avec d'autres ressources sur les voies.). Fait remarquable, 375 gènes sont uniques pour la carte RCD. Ces entités uniques sont réparties de manière homogène sur l'ensemble de la carte. Les modes de RCD décrits plus récemment, tels que la ferroptose et la pyroptose, qui sont soigneusement représentés dans la carte RCD, sont presque totalement absents de KEGG et de REACTOME. Dans l'ensemble, nous concluons que le contenu de la carte RCD n'est pas redondant avec d'autres bases de données de voies et qu'une proportion significative d'entités est unique à la carte RCD.

Nous avons également comparé les publications utilisées pour annoter la carte RCD et les voies sélectionnées correspondantes dans les ressources KEGG et REACTOME. 815 articles sur les 829 utilisés pour annoter la carte du RCD sont uniques et ne sont pas référencés dans les annotations des voies sélectionnées dans KEGG ou Reactome

(<https://www.mdpi.com/2072-6694/12/4/990/s1>, figure S5B). Bien que l'âge médian des références bibliographiques dans la carte RCD soit plus élevé que dans les voies correspondantes des deux autres bases de données, 25 % des articles référencés dans la carte RCD ont été publiés relativement récemment, entre 2010 et 2018 (Figure 24. Comparaison de la carte RCD avec d'autres ressources sur les voies. B).

Enfin, les types de journaux représentés dans les trois bases de données ont été comparés. L'éventail des revues utilisées pour l'annotation de la carte RCD et des deux autres bases de données est relativement similaire. Toutefois, la répartition des articles provenant de différents types de revues est assez différente. Dans les annotations de la carte RCD, les citations de revues dans le domaine de la biologie moléculaire sont surreprésentées par rapport aux deux autres bases de données. Cela indique que la carte RCD s'attarde sur les mécanismes moléculaires plutôt que sur la description phénotypique du processus (Figure 24. Comparaison de la carte RCD avec d'autres ressources sur les voies. C, D).

Dans l'ensemble, les résultats des comparaisons de bases de données indiquent que la ressource RCD est spécifique à un sujet et couvre tous les modes connus de mécanismes de mort cellulaire et leur interaction. La présentation soignée et l'organisation visuelle des connaissances biologiques affichées dans la carte en font une ressource attrayante pour l'analyse des données.

## 5. Première application : Alzheimer et cancer du poumon

La maladie d'Alzheimer est la maladie neurodégénérative la plus fréquente, dans laquelle les neurones meurent progressivement (Sperling et al., 2011). Néanmoins, les processus moléculaires qui entraînent la mort des cellules neuronales restent imprécis. Il est également intéressant de noter qu'une comorbidité inverse sur le plan épidémiologique a été décrite entre la maladie d'Alzheimer et le cancer du poumon non à petite cellule (Sánchez-Valle et al., 2017). Nous avons donc comparé les voies de signalisation impliquées ces deux pathologies.

Les données d'expression génique de trois études différentes ont été analysées en se concentrant sur l'hippocampe, qui est la zone du cerveau la plus touchée par cette maladie (Raskin et al., 2015). En outre, le transcriptome de trois cohortes de cancer du poumon non à petites cellules (CPNPC) a été étudié (GEO : GSE36980, GSE48350, GSE5281, GSE19188, GSE19804, GSE33532).

La méthode ROMA a été appliquée pour quantifier les scores d'activité des modules pour les deux maladies en utilisant la définition des modules de la carte RCD comme détaillé dans Martignetti et al. (Martignetti et al., 2016). Les analyses ont été réalisées, pour chacune des maladies de façon séparée, en comparant l'activité des patients à celle des contrôles. Les valeurs d'activité des modules ont été représentées sur la carte RCD à l'aide du plugin BiNoM de Cytoscape pour visualiser les différences entre les deux maladies (Bonnet et al., 2013).

Les scores ROMA, étudiant l'activité des modules, de la maladie d'Alzheimer et du cancer du poumon non à petites cellules présentaient plutôt une tendance inverse (Figure 25. Visualisation des scores d'activité moyens des modules ROMA à l'aide de données d'expression provenant d'échantillons d'hippocampe de patient atteint de la maladie d'Alzheimer (AD) et d'échantillons de cancer du poumon (LC) non à petites cellules (CPNPC) dans la carte RCD.). En ce qui concerne les types de mort cellulaire, les modules correspondant à la réponse TRAIL, à la pyroptose et aux récepteurs de dépendance étaient plus actifs dans la maladie d'Alzheimer (Figure 25. Visualisation des scores d'activité moyens des modules ROMA à l'aide de données d'expression provenant d'échantillons d'hippocampe de patient atteint de la maladie d'Alzheimer (AD) et d'échantillons de cancer du poumon (LC) non à petites cellules (CPNPC) dans la carte RCD.A). En revanche, dans le NSCLC, les modules ligand-récepteur (réponse TNF, réponse TRAIL et réponse FAS), ainsi que certains modules de la couche de signalisation (Pyroptose et Récepteurs de dépendance), sont moins actifs (Figure 25. Visualisation des scores d'activité moyens des modules ROMA à l'aide de données d'expression provenant d'échantillons d'hippocampe de patient atteint de la maladie d'Alzheimer (AD) et d'échantillons de cancer du poumon (LC) non à petites cellules (CPNPC) dans la carte RCD.A).

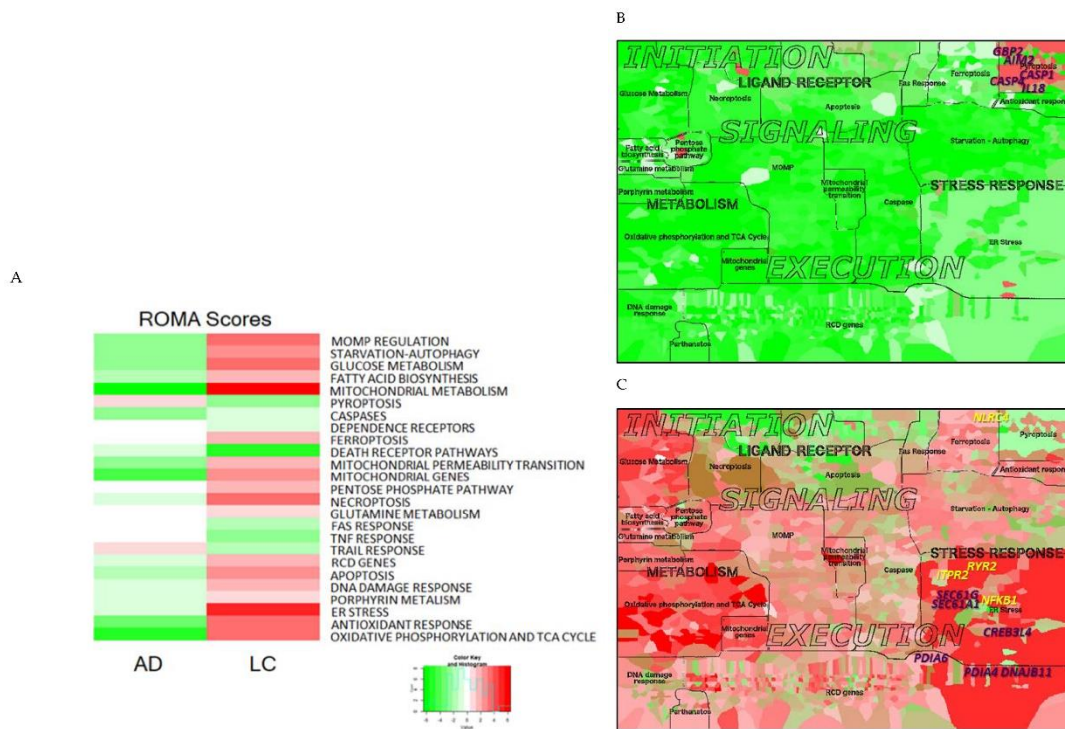


Figure 25. Visualisation des scores d'activité moyens des modules ROMA à l'aide de données d'expression provenant d'échantillons d'hippocampe de patient atteint de la maladie d'Alzheimer (AD) et d'échantillons de cancer du poumon (LC) non à petites cellules (CPNPC) dans la carte RCD.

(A) Carte de chaleur (*heatmap*) représentant les scores ROMA pour les deux maladies (chacune par rapport à ses témoins normaux). Coloration de la carte RCD avec les scores ROMA des données de la maladie d'Alzheimer (B) et du cancer du poumon non à petites cellules (C). Les valeurs tracées correspondent au score relatif du module ROMA par rapport aux témoins (comme en A). Les gènes contribuant le plus sont représentés par leur emplacement sur la carte, les gènes contribuant positivement en violet et les gènes contribuant négativement en jaune.

Il convient de noter que la plupart des modules correspondant au module de pyroptose semblent être plus actifs dans la maladie d'Alzheimer. En outre, dans le NSCLC, plusieurs modules liés au métabolisme (notamment le métabolisme du glucose, la phosphorylation oxydative et le cycle du citrate), ainsi que le stress du réticulum endoplasmique, étaient plus actifs (Figure 25. Visualisation des scores d'activité moyens des modules ROMA à l'aide de données d'expression provenant d'échantillons d'hippocampe de patient atteint de la maladie d'Alzheimer (AD) et d'échantillons de cancer du poumon (LC) non à petites cellules (CPNPC) dans la carte RCD. B et C). Cette activité des modules liés au métabolisme confirme des études antérieures (Hokama et al., 2014; Liang et al., 2008), notamment sur la différence entre la maladie d'Alzheimer et le cancer du poumon. En effet, la comparaison utilisant de la carte RCD entre la maladie d'Alzheimer et le cancer du poumon non à petites cellules est conforme aux spéculations sur la comorbidité inverse

entre les deux maladies (Musicco et al., 2013; Ou et al., 2012; Sánchez-Valle et al., 2017), ainsi qu'aux études épidémiologiques suggérant que le cancer du poumon non à petites cellules est moins fréquent chez les patients atteints de la maladie d'Alzheimer que chez les personnes sans maladie d'Alzheimer appariées selon l'âge (Tabarés-Seisdedos & Baudot, 2016).

L'étape suivante a consisté à identifier les gènes les plus importants pour chaque maladie. C'est-à-dire ceux qui contribuaient le plus aux différences d'activités observées. Pour ce faire, le coefficient de corrélation entre l'expression des gènes et l'activité d'un module a été calculé pour chaque étude. Les gènes qui présentaient un coefficient de corrélation d'au moins 0,5 (en valeur absolue), avec une valeur p significative et qui apparaissaient dans au moins deux des trois ensembles de données ont été sélectionnés. Une liste de gènes contribuant positivement ou négativement à chaque module a ainsi été établie (Tableau 2. Gènes contribuant le plus à chaque module pour les ensembles de données sur la maladie d'Alzheimer et le cancer du poumon.).

Tableau 2. Gènes contribuant le plus à chaque module pour les ensembles de données sur la maladie d'Alzheimer et le cancer du poumon.

Vert : gènes contribuant positivement, rouge : gènes contribuant négativement.

Module	Top Contributing Genes AD	Top Contributing Genes LC
MOMP regulation	<i>ATG5, VDAC3, CUL2, CYCS, HK1, MAPK6, BMF, RPS6KA1, MAPK10, VDAC2</i>	<i>DIABLO, RPS6KA1, CISH, CSNK2A1, C1QBP, FOXO1, PPP1CC, BCL2L11, SLC25A5, VDAC2</i>
Starvation Autophagy	<i>CAB39, ATG5, MAPK6, MAPK8, ACACA, PINK1, RPS6KA1, CASP8, MTOR, ATG12</i>	<i>PARP1, SESN1, LDHA, BCL2L11, PI3KC3, RPS6KA1, PINK1, ATG3, ATG12</i>
Glucose Metabolism	<i>PFKM, PRKACB, PRKAA2, LDHA, NOTCH1, H6PD, ALDOA</i>	<i>PGK1, LDHA, ANAPC11M, ALDOA, LDHC, PKM, ANAPC2</i>
Fatty acid biosynthesis	<i>PRKAA2, PPAT, CAB39, PRPS1, PRKACB, ACACA</i>	<i>ACLY, ACSS2, MLST8, PRPS1, PRPS2</i>
Mitochondrial metabolism	<i>SLC25A14, NDUFA9, FH, DLD, NDUFAF1, HK3, TNFRSF1A, ALOX15B</i>	<i>UQCRRH, ATP5G1, SLC25A10, CYC1, TACO1, NDUFAB1, PFKB2, RIPK3, SESN1, PLA2G4C</i>
Pyroptosis	<i>IL18, CASP4, GBP2, CASP1, AIM2</i>	<i>CASP1, GBP2, IL18, CASP4, IL1B</i>
Caspases	<i>YWHAG, MAPK8, MAPK6, CAB39, LMNA, HIP1, RIPK3, MTOR</i>	<i>G0S2, BIRC3, PIDD, FOXO1, RNF41</i>
Dependence receptors	<i>AATF, APPL2, PPP2CB, NGF, CARD8, NSMAF</i>	<i>PIK3CA, NEDD4, TRAF6, PPP2CB, NSMAF</i>
Ferroptosis	<i>NFE2L2, NQO1, FTL, ACSF2, SLC3A2, FTH1</i>	<i>GCLC, GCLM, GSR, MAFG, GSS, SLC3A2</i>
Death receptor pathways	<i>B4GALT6, GUCY1B3, CUL3, RBX1, RIPK3, HIF1A, NDUT5, CASP3</i>	<i>CYLD, PARP1, BRCA1, H2AFX, PRKDC, CARD8, STAT5A, FOXO1, MAGED1, TNFAIP3</i>
Mitochondrial permeability transition	<i>SLC25A4, HK1, VDAC3, VDAC1, PPID, CKMT1A, VDAC2</i>	<i>PPID, CAPN2, VDAC2, SLC25A5</i>
Mitochondrial genes	<i>ATP5C1, ATP5H, ATP5L, AIFM1, ATP5G1, NDUFB6, NDUFC2, ATP5A1, COX7A2, NDUFS4</i>	<i>ATP5G1, NDUFAB1, COX5B, ATP5B, NDUFB9, ATP5G3, COX6A1, COX6C, TA CO1, COX8A</i>

Module	Top Contributing Genes AD	Top Contributing Genes LC
Pentose phosphate pathway	ACACB, NME4	NME2, ACACB, RRM1, ACSS2, RRM2B, NME4, TXN, TP53, NME7
Necroptosis	B4GALT6, CUL3, RBX1, RIPK3, HK3, CASP10, PPP1CC	GUCY1A3, H2AFX, PARP1, CFLAR, RIPK3, GUCY1A2, DNML1, MLKL, NCF1
Glutamine metabolism	GLS, GLS2	MLKL, RIPK3, RIPK1, PRPS1
FAS response	CFLAR, CASP8, MLKL, GALNT14, CASP10	PTPN13, TNFSF10, MLKL, CFLAR, RIPK1, SQSTM1
TNF response	CYLD, MAP3K7, BIRC2, NSMAF, TAB2	CYLD, FLAD1, RNF11, PPP6C, CIAPIN1, RIPK1, CFLAR
TRAIL response	CYLD, CASP8, MLKL, GALNT14, CASP10	MLKL, TNFAIP3, CFLAR, RIPK1, TNFRSF10D
RCD genes	FBXO45, PKM, CDC42, ATG12, HSP90AA1, SIRT1, HIF1A, CASP3, BIRC2	KLF8, BCL2L11, STAT5A, PGK1, HUWE1, ALAS1, ENO1, NPM1, HSP90B1, ELK1
Apoptosis	RHOA, TCEB1, CYCS, SKP1, CCNB1, NKRF, BECN1, YWHAG, RPS6KA1, HIP1	DIABLO, STAT5A, RIPK3, CISH, RPS6KA1, HUWE1, C1QBP, BCL2L11, CDK6, PPP3R1
DNA damage response	PPP2R1A, CAB39, CSNK2A1, NKRF, HUWE1, RNF8, PRKDC	H2AFX, RAD50, MRE11A, ACSS2, HUWE1, BARD1, NKRF, CANK2A1, PPARGC1A, APBB1
Porphyrin metabolism	UROD, COX10, UROS, CPOX, COX15	CPOX, ALAS1, HMBS, ALAS2
ER stress	DERL1, DERL2, HSPA9, ATP2A1, MAP3K5, MAP3K4, DNAJC3, VCP, EIF2S3	NLRC4, PDIA6, RYR2, NFKB1, PDIA4, DNAJB11, SEC61G, SEC61A1, CREB3L4, ITPR1
Antioxidant response	NDUFS1, AIFM1, RBX1, CUL3, NDUFC2, NDUFAF1, COX7A2, NDUFS4, NDUFB5, COX6B2	UQCRH, MCL1, NDUFAB1, CYC1, COX5B, STAT5A, NDUFB9, COX6B1, COX4I2, COX6C
Oxidative phosphorylation and TCA cycle	NDUFS1, ATP5H, SLC25A14, PARK2, IDH3B, NDUFC2, FH, ATP5G1, IDH3A, SLC1A5	ATP5B, NDUFAB1, COX5B, ATP5G3, COX8A, COX6C, COX6B1, FH, NDUFB10, COX5A

La pyroptose a déjà été associée à la maladie d'Alzheimer et associée à l'inflammation aseptique chronique (Mamik & Power, 2017; Patel et al., 2017). Des données expérimentales ont d'ailleurs établi un lien entre la pyroptose dans la maladie d'Alzheimer et l'inflammasome lié à NLRP1 (Kaushal et al., 2015) ou l'inflammasome lié à NLRP3 (Schmid-Burgk et al., 2016). Cependant, la plupart des études établissant une corrélation entre la maladie d'Alzheimer et la pyroptose ont été réalisées sur des modèles de rongeurs. Ici, nous avons identifié *IL18*, *CASP4*, *GBP2*, *CASP1* et *AIM2* comme les gènes contribuant le plus au module de pyroptose. Le gène *IL18* est surexprimé dans le cerveau des patients atteints de la maladie d'Alzheimer (Ojala et al., 2009). L'expression de *CASP4* semble d'ailleurs être un médiateur des réponses inflammatoires dans la pathologie de la maladie d'Alzheimer (Kajiwara et al., 2016). *CASP1*, ainsi que d'autres gènes codant pour les caspases, est surexprimé chez les patients atteints de la maladie d'Alzheimer (Pompl et al., 2003). Dans des modèles de souris, on a constaté que l'AIM2 favorise la sécrétion d'IL1B par les neurones, ce qui pourrait également jouer un rôle dans la pathologie de la maladie d'Alzheimer (Saresella et al., 2016). Cependant, l'expression de GBP2 n'a pas encore été évaluée pour son rôle potentiel dans la pathologie de la maladie d'Alzheimer. Récemment, Saresella et ses collègues ont constaté que les composants de l'inflammasome (NLRP1, NLRP3, PYCARD, CASP1, CASP5 et CASP8) et les effecteurs en aval (IL1B, IL18) étaient régulés à la hausse dans les cellules mononucléaires du sang périphérique de patients atteints de la maladie d'Alzheimer modérée et sévère (Giampietri et al., 2015). Tous ces résultats confirment ceux que nous avons obtenus en utilisant la carte RCD ; néanmoins, de futures recherches sont nécessaires pour confirmer si la neuro-inflammation conduit à la pyroptose au cours de la pathologie de la maladie d'Alzheimer.

Dans le cancer, le stress du réticulum endoplasmique (RE) a été identifié comme une réponse adaptative qui favorise soit la croissance, soit l'apoptose (Avril et al., 2017; Giampietri et al., 2015). En outre, le stress du RE a également été associé à la résistance à la chimiothérapie (Salaroglio et al., 2017). Dans notre étude, nous avons identifié les gènes *PDIA6*, *PDIA4*, *DNAJB11*, *SEC61G*, *SEC61A1* et *CREB3L4* comme des gènes contribuant positivement et *ITPR1*, *RYR2*, *NFKB1* et *NLRC4* comme des gènes contribuant négativement au module de stress ER. Il a par exemple été démontré que *PDIA6* et *PDIA4* sont surexprimés dans les biopsies de CBNPC résistantes à la

chimiothérapie par le cisplatine, et que leur inhibition peut inverser la résistance aux médicaments (Tufo et al., 2014). Le gène *NFKB1* a été décrit comme un acteur clé de la voie du stress ER et des mécanismes de survie du cancer (Logue et al., 2013; Xia et al., 2014). *NLRC4* est régulé à la baisse dans les cas de cancer du poumon (Välk et al., 2010). *NLRC4* contient un domaine de recrutement des caspases (CARD) par lequel il peut réguler l'apoptose via les voies de signalisation NF- $\kappa$ B (Välk et al., 2010), ce qui suggère un lien possible entre les gènes *NLRC4* et *NFKB1* dans ce module. En revanche, il n'existe pas de rapports cohérents sur l'implication possible de *DNAJB11*, *SEC61G*, *SEC61A* et *CREB3L4* dans le cancer du poumon non à petites cellules.

## 6. Deuxième application : sous types de cancer de l'ovaire

Le cancer de l'ovaire, le deuxième cancer le plus répandu de l'appareil génital féminin, avec un taux de mortalité élevé (Gilks & Prat, 2009; McCluggage, 2011), englobe un groupe hétérogène de tumeurs car différentes par leurs caractéristiques génétiques, morphologiques et pathologiques qui sont impliquées dans l'initiation et la progression de la maladie (Etemadmoghadam et al., 2009; Konecny et al., 2014). Nous avons analysé les données transcriptomiques et les données sur le nombre de copies d'une cohorte de 489 cancers ovariens sévères de haut grade accessible via *The Cancer Genome Atlas* (TCGA) (Bell et al., 2011; Tomczak et al., 2015). Ces tumeurs ont été regroupées en quatre sous-types : différencié, immunoréactif, mésoenchymateux et prolifératif, comme indiqué ailleurs (Bell et al., 2011).

Sur la base de ces définitions de sous-types, nous avons appliqué l'analyse ROMA aux données d'expression de l'ARNm à cet ensemble de données, tout en comparant les quatre sous-types de cancer de l'ovaire. La visualisation des scores ROMA dans le contexte de la carte RCD a révélé des différences importantes entre ces sous-catégories (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire).

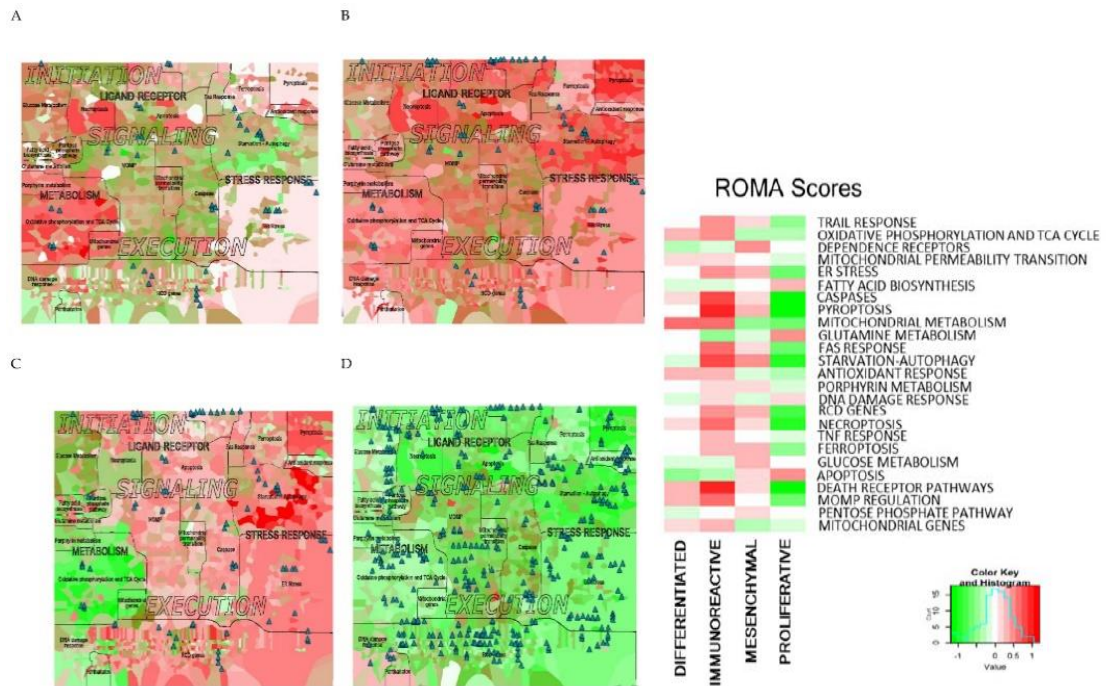


Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire

(A) différencié, (B) mésenchymateux, (C) immunoréactif, (D) sous-type prolifératif. La couleur de fond représente la valeur d'expression par protéine (rouge : régulée à la hausse, vert : régulée à la baisse). Les glyphes (triangles) représentent les gains en nombre de copies de gènes supérieurs à 4 copies. (E) Carte *heatmap* des scores ROMA par sous-type.

Le sous-type différencié présente plusieurs modules actifs dans le domaine du métabolisme (métabolisme mitochondrial, métabolisme du glucose, phosphorylation oxydative et cycle TCA). En revanche, les modules liés à l'apoptose et à l'autophagie semblaient sous-activés (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire A). Comme on pouvait s'y attendre (Konecny et al., 2014), le sous-type immunoréactif était caractérisé par une régulation à la hausse de la majorité des modules liés aux voies des récepteurs de la mort et aux exécuteurs liés à la caspase (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire B). En outre, nous avons observé l'activation des modules initiés par l'inflammation, tels que la pyroptose et, dans une moindre mesure, l'autophagie (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire C). Le type mésenchymateux, le plus agressif, a montré une forte activité dans certains modules d'exécution de la mort cellulaire, avec une suppression notable de la phosphorylation oxydative et du cycle TCA ainsi que de la réponse antioxydante (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire B). Le sous-type prolifératif présentait une activité plus élevée des modules liés à la

biosynthèse des acides gras, au métabolisme de la glutamine et à l'intégrité nucléaire (y compris la réponse aux dommages de l'ADN), ce qui correspond à l'activité répliquative élevée de ce sous-type de cancer. En revanche, la majorité des modules fonctionnels responsables de la régulation et de l'exécution de la mort cellulaire étaient fortement régulés à la baisse dans le sous-type prolifératif (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire E).

La visualisation des données de variation de nombre de copies des différents sous-types de cancer de l'ovaire dans le contexte de la carte de mort cellulaire régulée ici décrite permet de voir que le type prolifératif correspond à un plus grand nombre de gains de variation de nombre de copies (*CNV*), indiquant des aberrations génomiques plus fréquentes, par rapport aux autres sous-types (Figure 26. Visualisation des données d'expression et génomiques des groupes de cancer de l'ovaire), ce qui représente un facteur pronostique négatif majeur (Bell et al., 2011; Etemadmoghadam et al., 2009). Par exemple, *YWHAZ* favorise les métastases du cancer de l'ovaire en modulant la glycolyse (Shi et al., 2018). Les variations de nombre de copies de *WISP1* ont été associées à l'adénocarcinome de l'endomètre (Tang et al., 2011) tandis que *NDRG1* a été rapporté comme suppresseur de tumeur dans la carcinogenèse ovarienne à travers des sous-types ovariens distincts (Wang et al., 2014; D. Zhang et al., 2016). Il n'est donc pas surprenant de trouver ces gènes en tête de liste à travers les sous-types de cancer de l'ovaire.

## 7. Discussion

Les différents modes de mort cellulaire régulée partagent de nombreuses molécules communes entre les initiateurs, la signalisation et les acteurs de l'exécution. Il existe souvent un ensemble de signaux (*clues*), tels que l'état énergétique de la cellule, les signaux extra-cellulaires, ainsi que le type de dommage à l'ADN, qui, en combinaison, dicteront un choix de scénarios de mort cellulaire. La manière dont cette coordination s'effectue est dans une large mesure mal comprise. Les représentations linéaires et déconnectées des mécanismes de mort cellulaire régulés sont loin d'être satisfaisantes. Dans ce travail, nous avons tenté de rassembler toutes les informations disponibles sur les mécanismes de mort cellulaire régulée et de les représenter d'une manière structurée et lisible par un ordinateur.

La carte RCD couvre les phases d'initiation, la phase de signalisation au cours de laquelle le mode de mort cellulaire est choisi, ainsi que la phase d'exécution finale du processus.

La compréhension de l'interaction entre les modes de RCD représentés sous la forme d'une carte de signalisation complète peut aider à évaluer le statut des modes de mort cellulaire dans les maladies humaines. Cela peut s'avérer essentiel pour choisir la bonne stratégie de traitement et prédire les réponses thérapeutiques. Par exemple, la restauration de l'activité de la caspase dans le traitement du cancer par l'induction sélective de l'apoptose est l'un des objectifs communs des traitements antinéoplasiques, bien qu'elle soit entravée par l'activation des voies d'inhibition de l'apoptose. C'est pourquoi il a été suggéré de développer des agents alternatifs qui activeraient d'autres voies de RCD. Selon un scénario spéculatif, l'induction de la nécroptose, de la ferroptose ou de la mort dépendante de l'autophagie pourrait permettre de tuer les cellules cancéreuses et de restaurer ainsi leur sensibilité aux médicaments cytotoxiques (Galluzzi et al., 2012; Hoffmann, 2008; Vandenabeele et al., 2010; Xie et al., 2016). La carte RCD aiderait dans cette hypothèse à sélectionner les voies alternatives pour chaque cas particulier.

Au-delà de cette possibilité, nous démontrons ici que cette ressource peut être appliquée à d'autres pathologies. Comme les aberrations dans le contrôle de la mort cellulaire sont impliquées dans de nombreuses maladies, la carte pourrait être utilisée pour explorer des pathologies telles que les maladies infectieuses et les processus dégénératifs. Nous avons démontré dans ce travail que la carte de mort cellulaire programmée est utile pour comparer la maladie d'Alzheimer, un état de mort cellulaire excessive, et le cancer du poumon non à petites cellules, une pathologie liée à la suppression des voies de mort cellulaire.

De façon générale, l'une des applications évidentes et immédiates des cartes RCD est l'identification de signatures moléculaires des mécanismes de mort cellulaire pour une pathologie donnée. En effet, comme nous l'avons montré dans nos exemples d'application, l'intégration et l'analyse des données dans le contexte de la carte donnent un aperçu non seulement des processus RCD dérégulés dans une maladie ou un échantillon étudié, mais indiquent également les acteurs les plus importants, qui peuvent être de futures cibles

thérapeutiques. La deuxième méthode d'application de la carte RCD consiste à modéliser les commutations entre les modes de mort cellulaire. En fonction de l'approche de modélisation, la carte complète ou simplifiée du réseau RCD peut être utilisée.

### 3. Analyse de réseaux moléculaires impliqués dans la covid-19

Ce travail fera l'objet d'une publication écrite : le manuscrit est en cours d'écriture.

#### 1. Introduction

##### a. Mécanismes généraux impliqués dans le covid-19

Le coronavirus 2 du syndrome respiratoire aigu sévère (SARS) (SARS-CoV-2) l'agent infectieux de la covid-19 (Thomas et al., 2020), a représenté un défi sanitaire mondial majeur. Le SARS-CoV-2 est un virus à ARN simple brin et cible principalement les cellules épithéliales, provoquant généralement des maladies des voies respiratoires pouvant entraîner une pneumonie grave et un syndrome de détresse respiratoire aiguë (SDRA) (Cui et al., 2019). Il a été démontré que la covid-19 se propageait particulièrement facilement en raison de trois facteurs majeurs : (i) une grande partie des personnes infectées par le SARS-CoV-2 présentent des symptômes cliniques légers, voire inexistant, de covid-19 (Z. Gao et al., 2021) ; (ii) le SARS-CoV-2 semble être contagieux avant l'apparition des symptômes (Y. Gao et al., 2020); (iii) le SARS-CoV-2 semble se transmettre principalement par des aérosols provenant de l'air expiré, qui peuvent persister dans des zones mal ventilées pendant plusieurs heures, contribuant ainsi à ce que l'on appelle des événements de super propagation (Meselson, 2020). Les manifestations de la covid-19 varient d'une forme asymptomatique à une forme exceptionnellement grave, avec la possibilité de présenter de multiples symptômes systémiques, y compris des symptômes courants tels que fièvre, essoufflement, toux sèche persistante, frissons, douleurs musculaires, céphalées, perte de goût ou d'odorat, dysfonctionnement rénal et symptômes gastro-intestinaux (Guan et al., 2020). Les différences liées à l'âge et au sexe dans la présentation et la gravité de la covid-19 sont bien documentées (Li et al., 2020; O'Driscoll et al., 2021). Des atteintes de plusieurs organes ont également été signalées chez les patients atteints de covid-19 (Chen et al., 2020). En outre, les patients âgés atteints du covid-19 et présentant d'autres comorbidités peuvent présenter des manifestations respiratoires graves (Hu et al., 2021).

Il est prouvé que les facteurs viraux et les facteurs liés à l'hôte jouent un rôle essentiel dans la covid-19, les facteurs liés à l'hôte ayant un impact étroit sur la gravité de la maladie (Overmyer et al., 2021; X. Zhang et al., 2020). Les ensembles de données multiomiques d'analyse à haut débit, peuvent fournir une compréhension globale de l'hétérogénéité et de la diversité des types de données et des relations internes complexes dans la covid-19. L'identification de ces caractéristiques à l'aide d'un profil d'intégration multiomique couvrant des stades et des degrés de gravité spécifiques de la covid-19 pourrait permettre de mieux comprendre les caractéristiques des variants spécifiques du virus et les réponses individuelles, les manifestations de la maladie et les possibilités thérapeutiques, ce qui faciliterait la prédiction précise de l'évolution des patients et ainsi la prévention des complications. En agrégeant les profils d'intégration multiomique de covid-19, nous pouvons identifier des cibles préventives et thérapeutiques ainsi que des outils de diagnostic précoce, et prédire plus efficacement les résultats critiques ou graves, mieux sélectionner les médicaments et approfondir notre compréhension des conditions comorbides et des complications à long terme associées à la covid-19.

#### b. Contexte de ce projet

La maladie à coronavirus 2019 (covid-19) a été détectée pour la première fois en Chine en décembre 2019, puis déclarée pandémie mondiale par l'Organisation mondiale de la santé (OMS) en mars 2020 [1-3]. Très rapidement, de grands consortiums se sont formés afin de comprendre et de traiter cette infection sévère.

En mai 2020, un effort collaboratif de la communauté des cartes moléculaires maladies spécifiques a été officiellement lancé à l'échelle internationale par la communauté « Disease Map » (<https://doi.org/10.17881/covid19-disease-map> (Ostaszewski et al., 2021)). Il s'agissait d'un effort pour construire un référentiel de connaissances complet et normalisé des mécanismes d'interaction entre le virus SARS-CoV-2 et l'hôte, guidé par les contributions d'experts du domaine et basé sur des travaux publiés. Ces connaissances, disponibles dans la vaste littérature existante et dans le nombre croissant de nouvelles publications sur le SARS-CoV-2, nécessitent une organisation rigoureuse et efficace dans des formats lisibles à la fois par l'homme et par la machine.

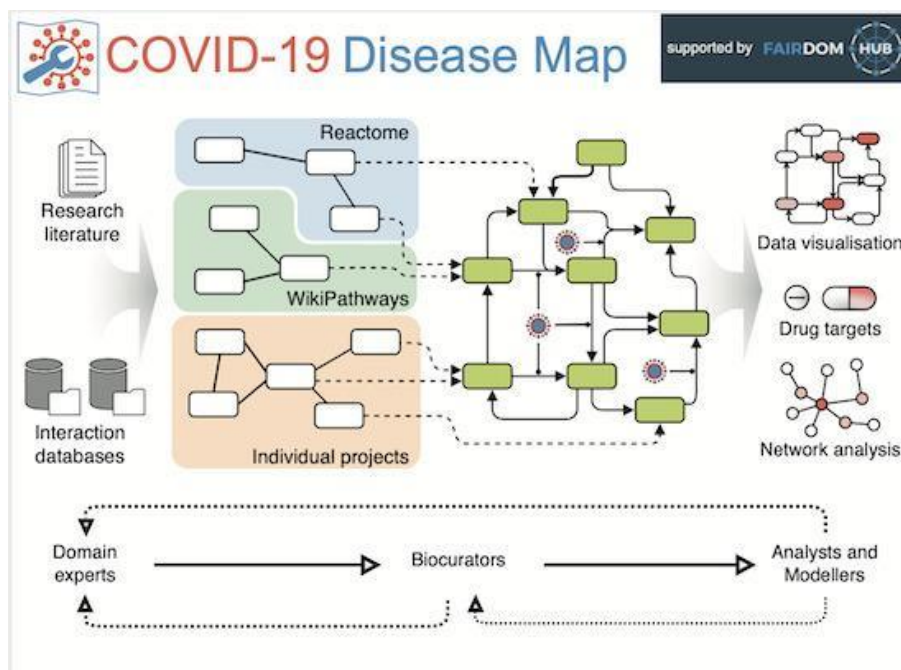


Figure 27. Covid-19 Disease Map.

La carte de la covid-19 est une collection à grande échelle de modèles informatiques et de diagrammes des mécanismes moléculaires impliqués dans l'infection par le SARS-CoV-2. La carte facilite l'exploration informatique des voies affectées par le virus. (Ostaszewski et al., 2021)

Il s'agit d'un effort communautaire à grande échelle visant à créer un référentiel de mécanismes moléculaires COVID-19 en libre accès (Ostaszewski et al., 2020). La carte de la COVID-19 (C19DMap) est une représentation graphique et interactive des mécanismes moléculaires pertinents pour les maladies, qui relie de nombreuses sources de connaissances. Il s'agit notamment d'une ressource informatique pour les analyses basées sur les graphes et la modélisation des maladies.

J'ai été sollicité par ce groupe dans le cadre de mes précédents travaux sur la mort cellulaire pour, en particulier, participer à l'écriture du chapitre sur l'autophagie impliquée dans la covid-19 (Ostaszewski et al., 2021). Par la suite, une proposition a émergé pour créer des cartes moléculaires du système immunitaire inné spécifiques à la covid-19 en se basant sur les cartes précédemment publiées (Kondratova et al., 2019) dans le système immunitaire inné. En effet, de nombreuses équipes ont étudié l'implication du système immunitaire dans la covid-19 mais à ce jour aucune carte moléculaire spécifique de la covid-19 n'a été publiée. L'objectif était donc de mettre à jour des cartes « spécifiques cancer » en les adaptant pour qu'elles deviennent « spécifiques à la covid-19 » et de rationaliser la méthodologie permettant d'y arriver.

## Description des cartes existantes

Ce projet se base sur trois cartes du système immunitaire innée spécifique au cancer (Kondratova et al., 2019). Elles sont spécifiquement conçues chacune pour un type cellulaire : les *Natural Killer* (NK), les cellules dendritiques et les macrophages (Figure 28. Cartes spécifiques à un type de cellule pour les macrophages, cellules dendritiques et *Natural Killer*.. Ces cartes ont déjà été publiées et sont intégrées dans *ACSN* (*Atlas of Cancer Signalling Network*) au même titre que la carte sur la mort cellulaire programmée. La carte spécifique aux macrophages contient 588 items et 7 modules représentant la polarisation pro-tumorale et anti-tumorale des cellules myéloïdes. La carte des cellules dendritiques contient 491 items et 8 modules. La carte NK contient 567 items et 6 modules.

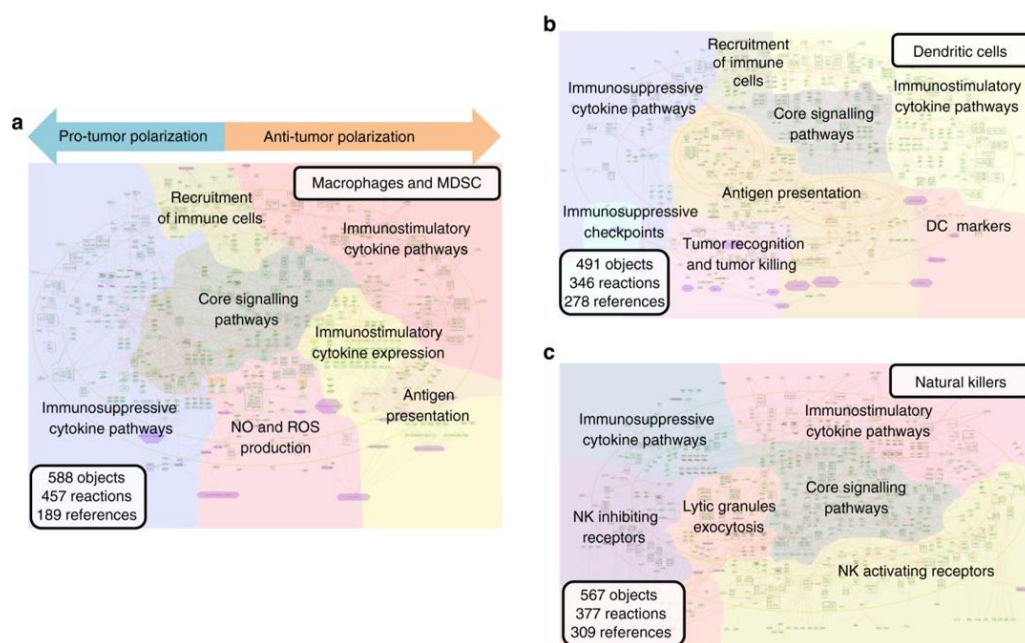


Figure 28. Cartes spécifiques à un type de cellule pour les macrophages, cellules dendritiques et *Natural Killer*.

Les réseaux spécifiques à un type de cellule sont visualisés au niveau supérieur, l'arrière-plan coloré indique les limites des modules fonctionnels des cartes. a Les cartes des macrophages et des cellules suppressives d'origine myéloïde (MDSC). b La carte des cellules dendritiques. c La carte des cellules tueuses naturelles. (Kondratova et al., 2019)

## Ailani

Afin de compléter ces cartes avec les connaissances publiées, nous nous sommes enquis d'outils de recherche de termes pour étoffer les cartes existantes des toutes dernières expériences. Nous avons pu utiliser Ailani (Artificial Intelligence LANguage Interface), un

outil de recherche sémantique basé sur du *text-mining* (Figure 29. Fonctionnement de Ailani.

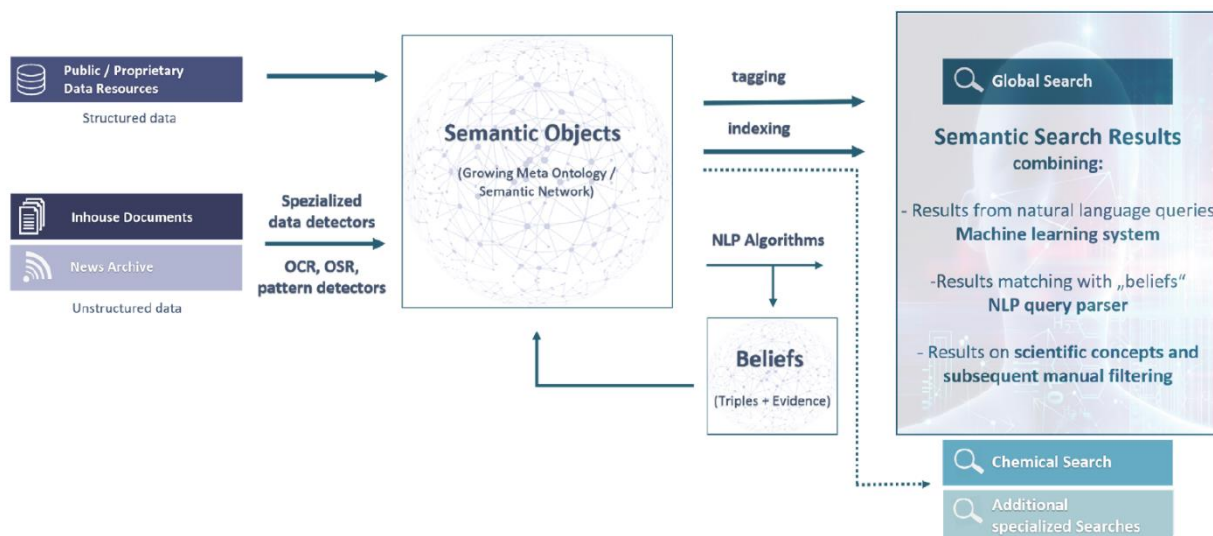


Figure 29. Fonctionnement de Ailani

Les résultats de *text-mining* sont accessibles sous la forme d'un tableau rapportant les interactions moléculaires avec des références aux articles et aux phrases à partir desquelles ces interactions ont été dérivées. Le contenu d'Ailani est accessible via une interface permettant aux utilisateurs de formuler des requêtes, telles que "Quels sont les facteurs de risque de la COVID-19 ?" ou "Quels sont les interacteurs de ACE2 ?", facilitant ainsi l'extraction de connaissances. Un tableau récapitulant toutes les interactions décrites dans la covid-19 a ainsi été construit.

### *Résultats préliminaires*

Nous avons utilisé la carte des cellules *Natural Killer* comme preuve de concept. L'objectif était de vérifier la pertinence, de mettre à jour les trois cartes précédemment décrites et d'en tirer une méthodologie.

La première question était de savoir dans quelle mesure la carte pouvait être réutilisée, c'est-à-dire à quel point les réactions dépeintes sur la carte initiale (spécifique au cancer) étaient relativement génériques (Figure 30. Analyse de la composition de la carte NK. Sur les 4511 réactions (en considérant également les réactions intra complexe), 436 étaient référencées dans la base de données Ailani. La majorité de ces interactions étaient associées

à plus de trente pathologies (389/436, 89,2 %) ; c'est-à-dire que sur l'échantillon de réactions étudié, on peut conclure que la majorité des interactions sont génériques.

Nous nous sommes ensuite demandé combien de gènes présents sur la carte de *Natural Killer* étaient associés à une réaction dans la covid-19 dans la base Ailani. L'idée étant de savoir quelle proportion de la carte existante était potentiellement réutilisable. 117 items (117/348, 33,6 %) de la carte sont présents dans Ailani et donc potentiellement associés à des réactions spécifiques à la covid-19.

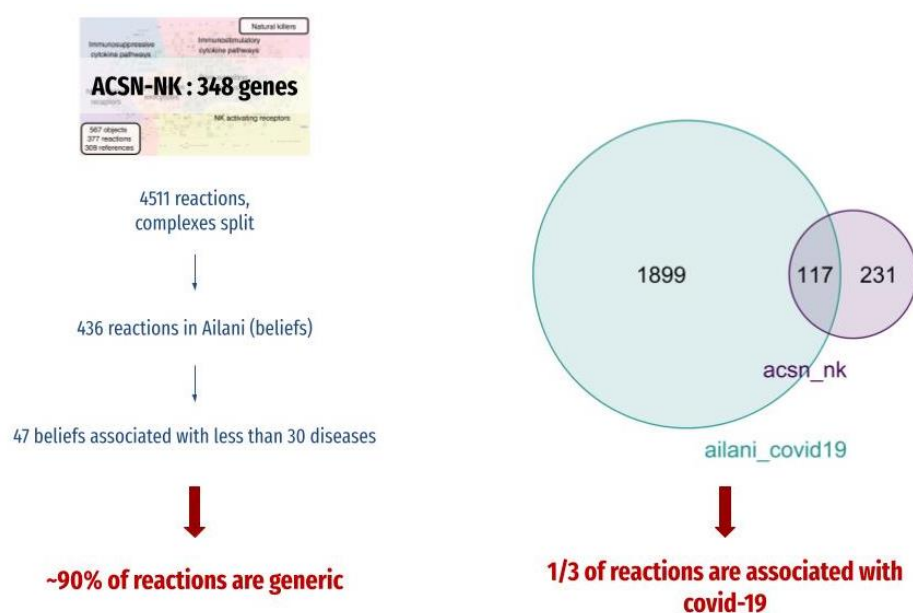


Figure 30. Analyse de la composition de la carte NK

L'analyse était décomposée en deux étapes : (à gauche) l'étude de la proportion de réactions partagées par plusieurs pathologies donc génériques et (à droite) l'étude du nombre de réactions présentes dans la carte et associées à la covid-19 dans la base de donnée Ailani.

Nous avons donc regardé plus en détail les protéines en interaction avec ces protéines déjà sur la carte et associées à la covid-19 sur Ailani. Pour ce faire, nous avons interrogé la base Ailani en limitant les protéines aux interactions liées à la covid-19. Finalement, nous avons sélectionné uniquement les interactions présentes dans Omnipath, c'est-à-dire celles qui avaient le plus de chance de correspondre à une interaction réelle (Türei et al., 2021). Nous avons ainsi obtenu une liste de 33 gènes à ajouter.

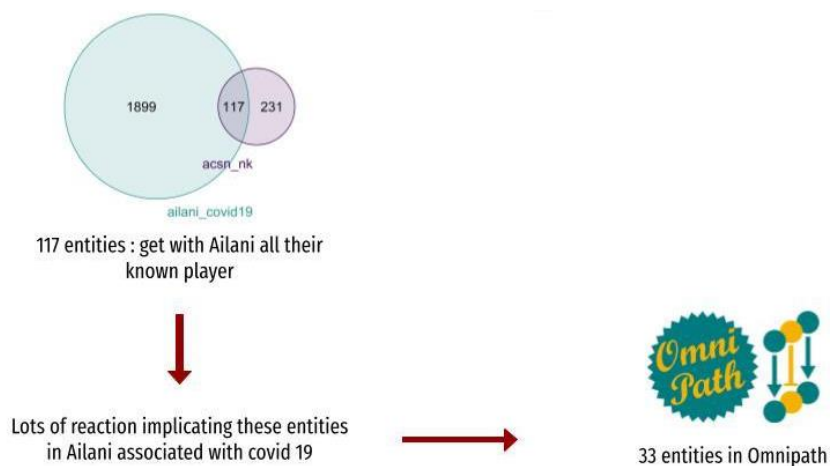


Figure : extend map using Ailani and Omnipath

Bien que limité, ce nombre correspond en pratique à une quantité supplémentaire significative de travail de recherche bibliographique. En effet, dans les cartes ACSN répertoriant des interactions biomoléculaires spécifiques, chacune des interactions nécessite une référence bibliographique en annotation et l'ajout manuel sur la carte (pour les trois cartes). A ce jour, le format MIRIAM n'est pas utilisé dans les cartes de ACSN, le système NOTE libre est utilisé. Ce sera probablement un point à adapter dans le futur. MIRIAM pour *Minimal Information Required In the Annotation of biochemical Models* correspond à un ensemble de recommandations d'écriture des annotations dans un format lisible par un logiciel et non ambigu (Le Novère et al., 2009).

Dans ce contexte, et parce que de nombreuses voies de signalisation dans les deux maladies se chevauchent, nous avons choisi d'utiliser les cartes précédemment construites dans le cadre d'études cancer en l'état pour étudier les interactions communes à la covid-19. Plus spécifiquement, l'objectif consiste à mettre en évidence des différences d'activation de voies de signalisation cellulaire- spécifiques en fonction du degré de sévérité (modéré ou sévère) de l'atteinte ou comparativement aux personnes contrôles.

## 2. Ressources et méthodes

### a. Les cartes issues de ACSN

Les trois cartes cellules - spécifiques ont été décrites ci-dessus. Les modules (*gene set*) utilisées sont les 50 modules du « hallmark gene sets » de la *Human Molecular Signatures*

*Database (MSigDB)*(Liberzon et al., 2015). L'objectif est d'avoir une approche très généraliste est donc d'analyser les voies de signalisations classiques.

b. Données utilisées

Les données de *scRNA* sont issues de deux groupes de données librement accessibles.

*Chua et al., 2020*

Le premier travail utilisé (Chua et al., 2020), a étudié la réponse immunitaire de patients atteints de la covid-19 par séquençage de l'ARN unicellulaire (*scRNA-seq*) d'échantillons nasopharyngés et bronchiques afin d'identifier des marqueurs moléculaires de la sévérité de la maladie.

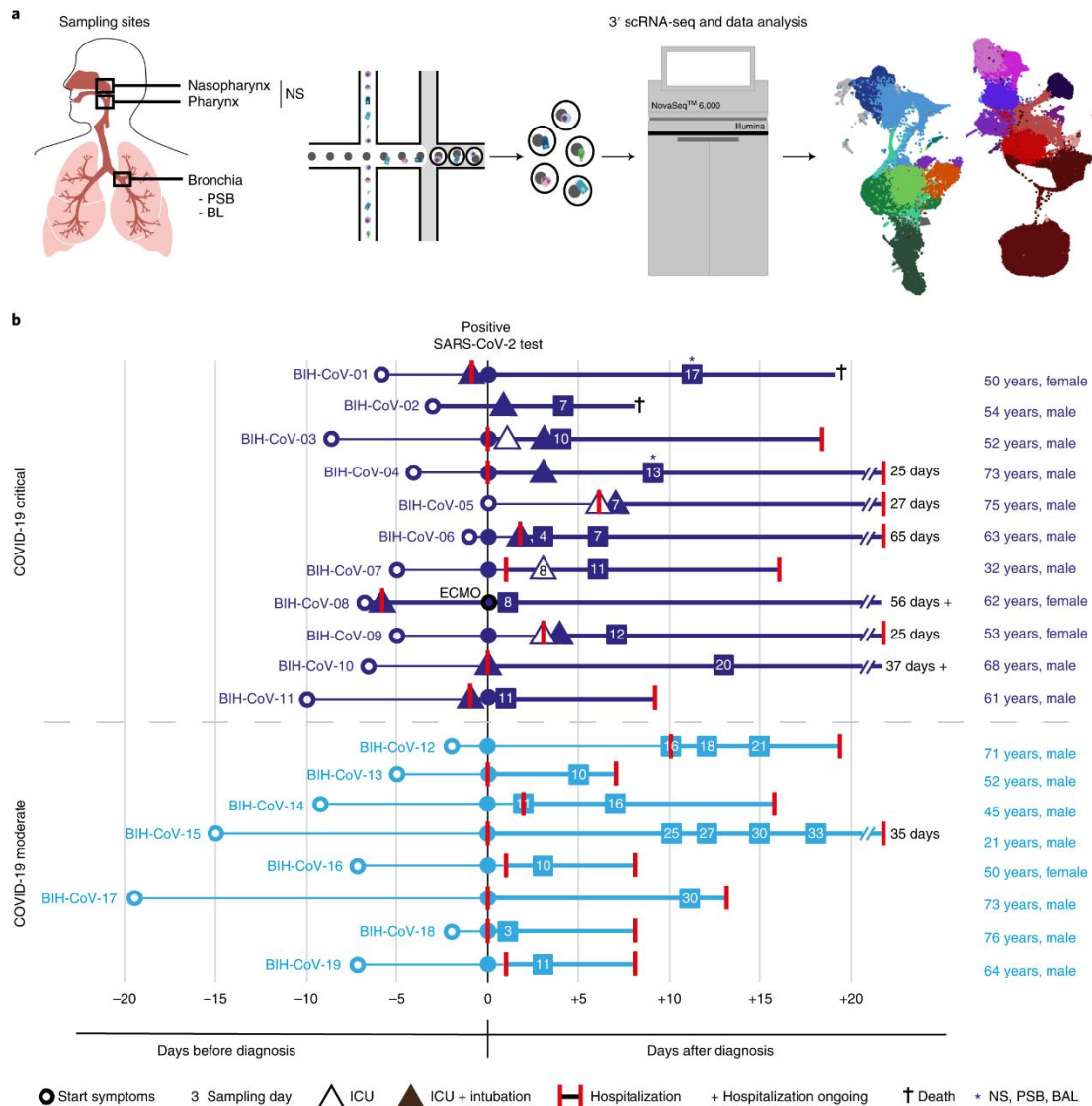


Figure 31. Représentation schématique du flux de travail expérimental extraite de (Chua et al., 2020)

a, Les sites d'échantillonnage (à gauche) et la préparation de la bibliothèque 3' scRNA-seq à l'aide de 10X Genomics (au milieu), suivie de l'analyse des données révélant l'identité du type de cellule (à droite), sont représentés. b, Vue d'ensemble de la cohorte de patients. Sont indiqués l'âge, le sexe et la gravité de la covid-19 ainsi que le début des symptômes, la durée d'hospitalisation et les points temporels d'échantillonnage en jours après le début des symptômes, tous les patients étant alignés temporellement sur le jour du test SARS-CoV-2 positif. L'admission à l'unité de soins intensifs est également décrite, le cas échéant. Un patient a nécessité une ECMO. Le jour d'échantillonnage par rapport à l'apparition des symptômes est indiqué par un nombre dans un carré ou un triangle.

Ces auteurs ont réalisé du scRNA-seq sur des écouvillons nasopharyngés ou des écouvillons nasopharyngés/pharyngés regroupés (NS), des brosses d'échantillons protégés bronchiques (PSB) et des lavages bronchiques (BL) (Figure 31. Représentation schématique du flux de travail expérimental extraite de (Chua et al., 2020)). Ils ont

analysé une cohorte de patients de la Charité (Berlin) et de l'hôpital universitaire de Leipzig, comprenant 19 patients atteints de la covid-19 (15 hommes et 4 femmes ; âge compris entre 21 et 76 ans ; âge médian : 61 ans). Cinq donneurs séronégatifs pour le SARS-CoV-2 et ne présentant aucun signe de maladie ont été inclus comme témoins (deux hommes et trois femmes ; tranche d'âge : 24-41 ans ; âge médian : 34 ans). Sur la base des recommandations de l'Organisation mondiale de la santé (OMS), ils ont classé 8 patients dans la catégorie « modérée » et 11 patients dans la catégorie « critique », dont 2 sont décédés plusieurs jours plus tard (patients BIH-CoV-01 et BIH-CoV-02). Au total, 160 528 cellules ont été obtenues à partir d'un total de 36 échantillons prélevés sur 24 individus.

*Gao et al., 2021*

De janvier à juin 2020, cette équipe a recueilli des échantillons de lavage nasal d'adultes diagnostiqués avec une infection à SARS-CoV-2 par PCR (n = 8, 7 donneurs) ainsi que des échantillons d'adultes diagnostiqués avec le virus influenza A par PCR (n = 14, 8 donneurs)(K. M. Gao et al., 2021). Ils ont également utilisé des échantillons de lavage nasal de volontaires adultes sains sans preuve d'infection virale respiratoire comme comparateur (n = 6, 6 donneurs). Ils ont analysé un total de 29 406 cellules dans tous les échantillons, en particulier 5669 cellules provenant de donneurs atteints de la covid-19, 5629 cellules provenant de donneurs sains et 18 108 cellules provenant de donneurs IAV positifs.

### c. Analyse de l'expression différentielle d'expression de gènes (DGE)

L'analyse de l'expression différentielle d'expression de gènes consiste à prendre les données normalisées du nombre de lectures d'ARN entre différents échantillons et à effectuer une analyse statistique pour mettre en évidence des changements quantitatifs dans les niveaux d'expression entre les groupes expérimentaux (Figure 32. Analyse de l'expression différentielle des gènes (DGE)).

L'objectif des tests d'expression différentielle est de déterminer quels sont les gènes qui sont exprimés à des niveaux différents d'une condition à l'autre. Ces gènes peuvent

apporter un éclairage biologique sur les processus affectés par la ou les pathologies étudiées.

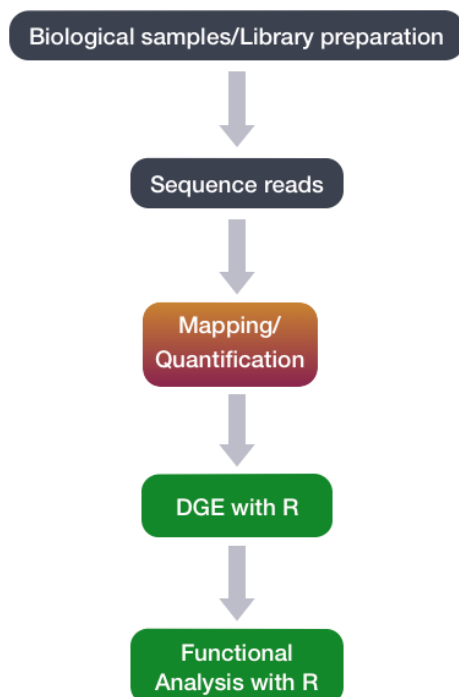


Figure 32. Analyse de l'expression différentielle des gènes (DGE)  
Adapté de (*Differential Gene Expression (DGE) Analysis*,  
[https://hbctraining.github.io/Training-modules/planning\\_successful\\_rnaseq/lessons/sample\\_level\\_QC.html](https://hbctraining.github.io/Training-modules/planning_successful_rnaseq/lessons/sample_level_QC.html))

Les comptages utilisés pour l'analyse de l'expression différentielle représentent le nombre de lectures de séquences provenant d'un gène particulier. Plus le nombre est élevé, plus il y a de lectures associées à ce gène et plus on suppose que le niveau d'expression de ce gène est élevé dans l'échantillon. Tout d'abord, ces comptages doivent être normalisés pour tenir compte des différences de taille des *library* (ensemble des transcrits présents) et de composition de l'ARN entre les échantillons.

### *Normalisation*

Les principaux facteurs souvent pris en compte lors de la normalisation sont les suivants

- Profondeur de séquençage : La prise en compte de la profondeur de séquençage est nécessaire pour comparer l'expression des gènes entre les échantillons (Figure 33. Impact de la profondeur de séquençage sur l'analyse DGE.). Dans l'exemple ci-

dessous, l'expression de chaque gène semble avoir doublé dans l'échantillon A par rapport à l'échantillon B, mais cela est dû au fait que l'échantillon A a une profondeur de séquençage deux fois plus importante.

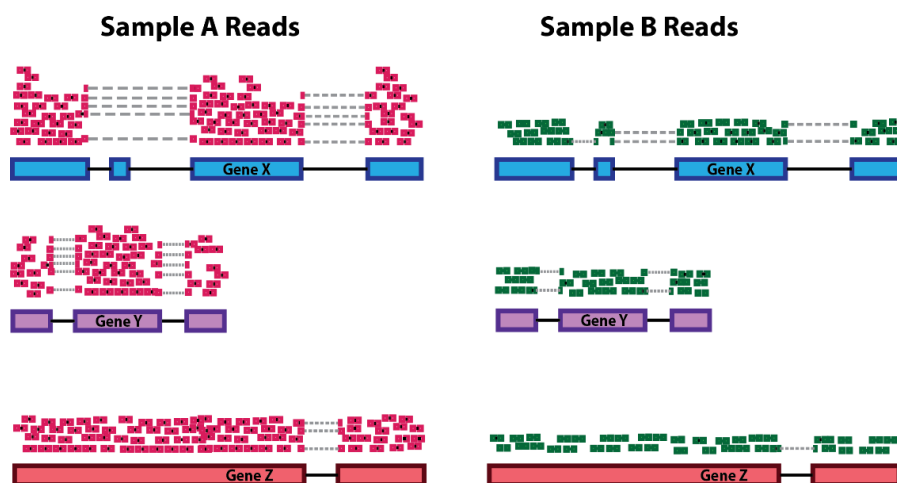


Figure 33. Impact de la profondeur de séquençage sur l'analyse DGE.

Adapté de (*Differential Gene Expression (DGE) Analysis*, [https://hbctraining.github.io/Training-modules/planning\\_successful\\_rnaseq/lessons/sample\\_level\\_QC.html](https://hbctraining.github.io/Training-modules/planning_successful_rnaseq/lessons/sample_level_QC.html))

- Longueur du gène : La prise en compte de la longueur des gènes est nécessaire pour comparer l'expression de différents gènes au sein d'un même échantillon (Figure 34. Impact de la longueur du gène sur l'analyse de DGE.. Dans l'exemple, le gène X et le gène Y ont des niveaux d'expression similaires, mais le nombre de lectures associées au gène X serait beaucoup plus élevé que le nombre de lectures associées au gène Y, car le gène X est plus long.

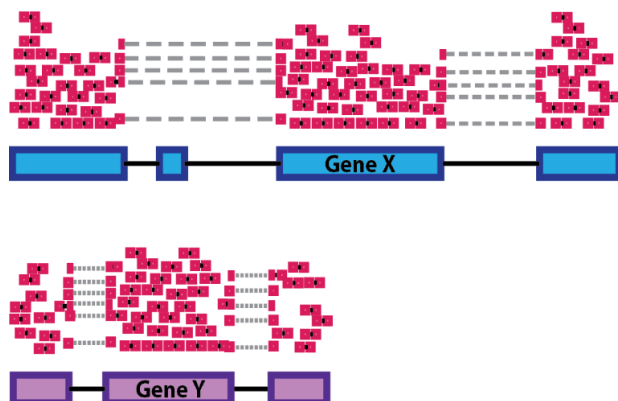


Figure 34. Impact de la longueur du gène sur l'analyse de DGE.

- Composition de l'ARN : quelques gènes fortement exprimés de manière différentielle entre les échantillons, des différences dans le nombre de gènes exprimés entre les échantillons ou la présence de contamination peuvent fausser certains types de méthodes de normalisation (Figure 35. Impact de la composition de l'ARN sur l'analyse de la DGE.). La prise en compte de la composition de l'ARN est recommandée pour une comparaison précise de l'expression entre les échantillons, et est particulièrement importante lors des analyses d'expression différentielle

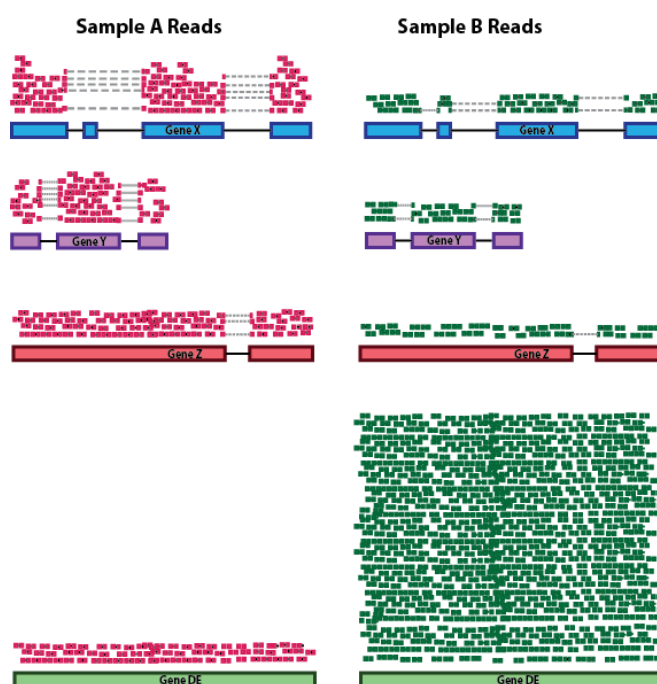


Figure 35. Impact de la composition de l'ARN sur l'analyse de la DGE.

Il existe plusieurs méthodes de normalisation. Celle qui est particulièrement recommandée pour les analyses d'expression différentielle est la *trimmed mean of M values* (TMM) d'EdgeR (Robinson et al., 2010) qui utilise une moyenne pondérée des rapports d'expression logarithmique entre les échantillons, la profondeur de séquençage, et la composition de l'ARN.

De plus, les gènes dont le nombre de *lecture* est très faible dans tous les échantillons doivent être éliminés avant l'analyse en aval. Cela se justifie à la fois pour des raisons biologiques et statistiques. D'un point de vue biologique, un gène doit être exprimé à un niveau minimal avant d'être traduit en protéine ou d'être considéré comme biologiquement important. D'un point de vue statistique, il est très peu probable que les gènes dont le nombre de *lecture* est constamment faible soient évalués comme significativement discriminants. Ces gènes peuvent donc être retirés de l'analyse sans perte d'information.

En règle générale, il faut qu'un gène ait au moins 10 lectures dans plusieurs échantillons. Dans la pratique, le filtrage est basé sur des valeurs de comptage par million (CPM) afin d'éviter de favoriser les gènes exprimés dans les grandes *library* par rapport à ceux exprimés dans les petites.

Enfin, il est nécessaire de ne pas retenir les échantillons ayant trop peu de séquences.

#### d. Analyse scRNAseq

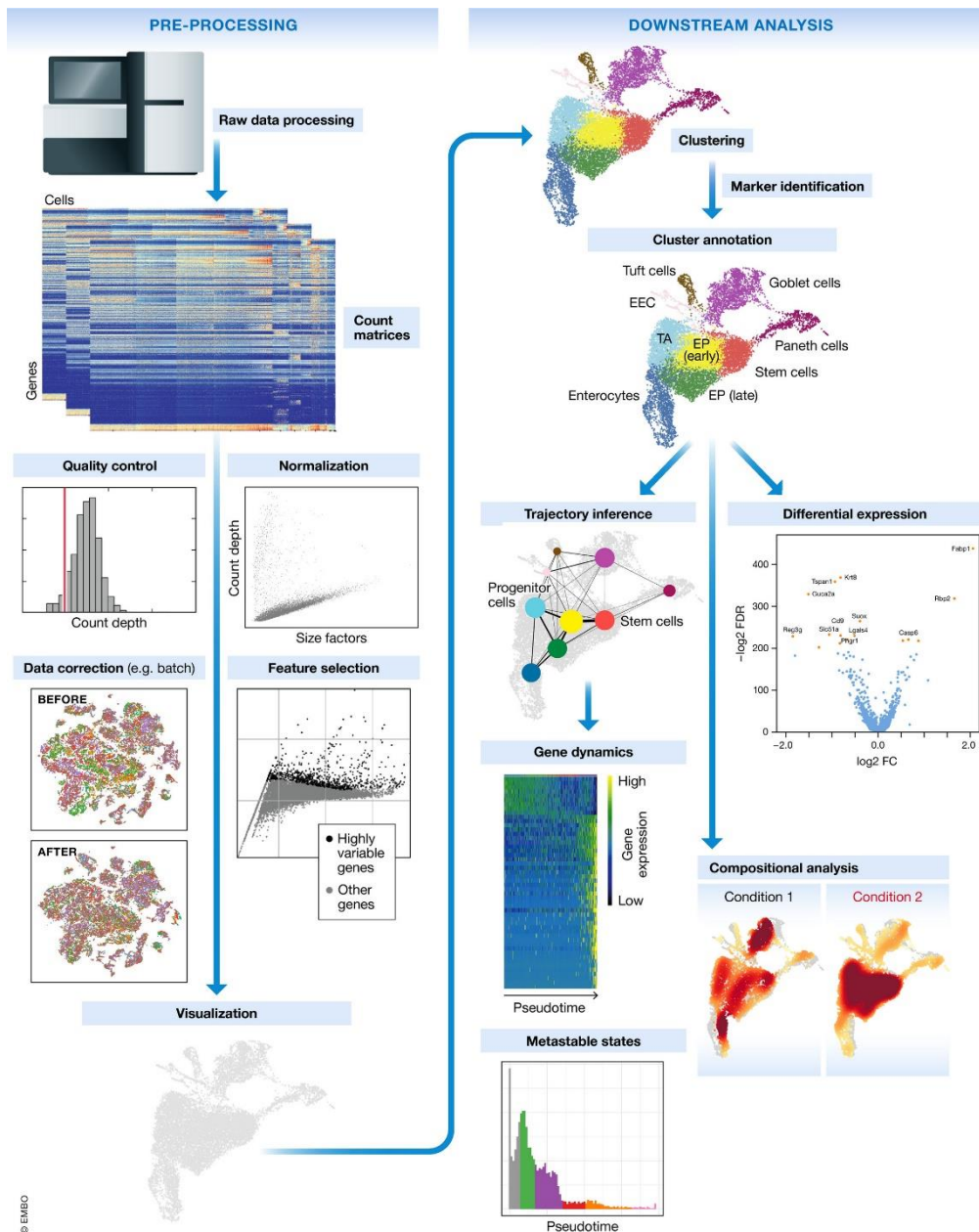


Figure 36. Schéma d'un flux de travail typique d'une analyse RNA-seq de cellule unique extraite de (Luecken & Theis, 2019).

Les données brutes de séquençage sont traitées et alignées pour donner des matrices de comptage qui représentent le début du flux de travail. Les données de comptage (*count matrix*) sont ensuite traitées (*pre processing*) et analysées. (Luecken & Theis, 2019).

Les données brutes générées par les machines de séquençage sont traitées pour obtenir des matrices de comptage moléculaire ou, alternativement, des matrices de lecture (Figure 36. Schéma d'un flux de travail typique d'une analyse RNA-seq de cellule unique extraite de (Luecken & Theis, 2019).).

Avant d'analyser les données d'expression génique d'une seule cellule, il faut s'assurer que toutes les données de codes-barres cellulaires correspondent à des cellules viables. Le contrôle qualité cellulaire est généralement effectué sur la base de trois variables :

- Le nombre de lecture (*count depth*),
- Le nombre de gènes
- La fraction de comptes provenant de gènes mitochondriaux

Par exemple, les codes-barres présentant une faible profondeur de comptage, peu de gènes détectés et une fraction élevée de comptages mitochondriaux indiquent des cellules dont l'ARNm cytoplasmique a fui à travers une membrane rompue, et donc, seul l'ARNm situé dans les mitochondries est encore conservé. À l'inverse, les cellules présentant des comptages étonnamment élevés et un grand nombre de gènes détectés peuvent représenter des doublets (plusieurs cellules capturées ensemble). C'est pourquoi des seuils de profondeur de comptage élevés sont couramment utilisés pour filtrer les doublets potentiels.

#### e. Pseudobulk

Les cellules individuelles d'un échantillon ne sont pas indépendantes les unes des autres, puisqu'elles sont isolées du même animal/échantillon provenant du même environnement. Si les cellules sont traitées comme des échantillons distincts, la variation au sein d'une population n'est pas vraiment étudiée, mais c'est plutôt la variation au sein d'un individu. En général, il est utile de déduire quels gènes peuvent être importants pour une condition au niveau de la population (et non au niveau individuel). Pour ce faire, une des meilleures pratiques actuelles consiste à utiliser une approche pseudo-bulk, ce qui implique les étapes suivantes :

- Séparation des cellules par type cellulaire d'intérêt
- Extraction des comptes bruts après filtrage qualité
- Agrégation des comptes et des métadonnées au niveau de l'échantillon.
- Réalisation de l'analyse

L'ensemble de ces étapes est basé sur les package Seurat et edgeR de R. Le script final utilisé est disponible ici :

[https://github.com/JiMouse/rRoma\\_covid19/blob/main/SeuratPseudobulk.R](https://github.com/JiMouse/rRoma_covid19/blob/main/SeuratPseudobulk.R)

### 3. Résultats préliminaires

- a. Les cellules immunitaires innées des voies respiratoires semblent avoir un profil d'expression génique différent dans les cellules covid-19

Nous avons commencé par visualiser l'expression moyenne des gènes des pseudobulk des cellules dendritiques, macrophages et cellules tueuses naturelles (*natural killer, NK*) sur les trois cartes précédemment citées sur Navicell (Figure 37. Représentation de l'expression génique moyenne des cellules dendritiques, macrophages et natural killer.). Les données utilisées sont celles de (Chua et al., 2020).

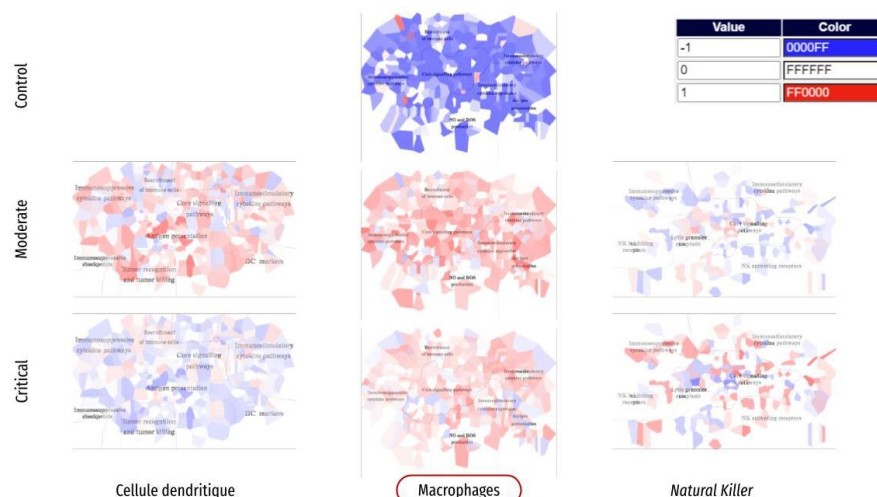


Figure 37. Représentation de l'expression génique moyenne des cellules dendritiques, macrophages et natural killer.

Il n'y avait pas de cellules *natural killer* ni de cellules dendritique à analyser chez les contrôles.

La différence d'expression au niveau des macrophages est assez remarquable entre les contrôles et les patients. Certaines différences sont à noter entre les patients modérés et critiques pour les cellules dendritiques et les *natural killers*.

Nous avons donc étudié les voies de signalisation différemment exprimées au niveau des macrophages en comparant les contrôles, les critiques et modérés en utilisant rRoma. Plus particulièrement, nous avons analysé les macrophages non-résidents qui ont été

associés à la forte réponse inflammatoire pulmonaire observées chez les cas sévère (Booz et al., 2020; Lian et al., 2022; Meidaninikjeh et al., 2021). La plupart des macrophages sont en effet dérivés des monocytes sanguins circulants et recrutés lors d'une infection (Davies et al., 2013). Plusieurs macrophages, tissus-résidents, sont dérivés de tissus embryonnaires distincts. Ces deux grandes catégories de macrophages ont des fonctions différentes et en particulier, les macrophages tissulaires ont des fonctions spécifiques au tissu où ils résident.

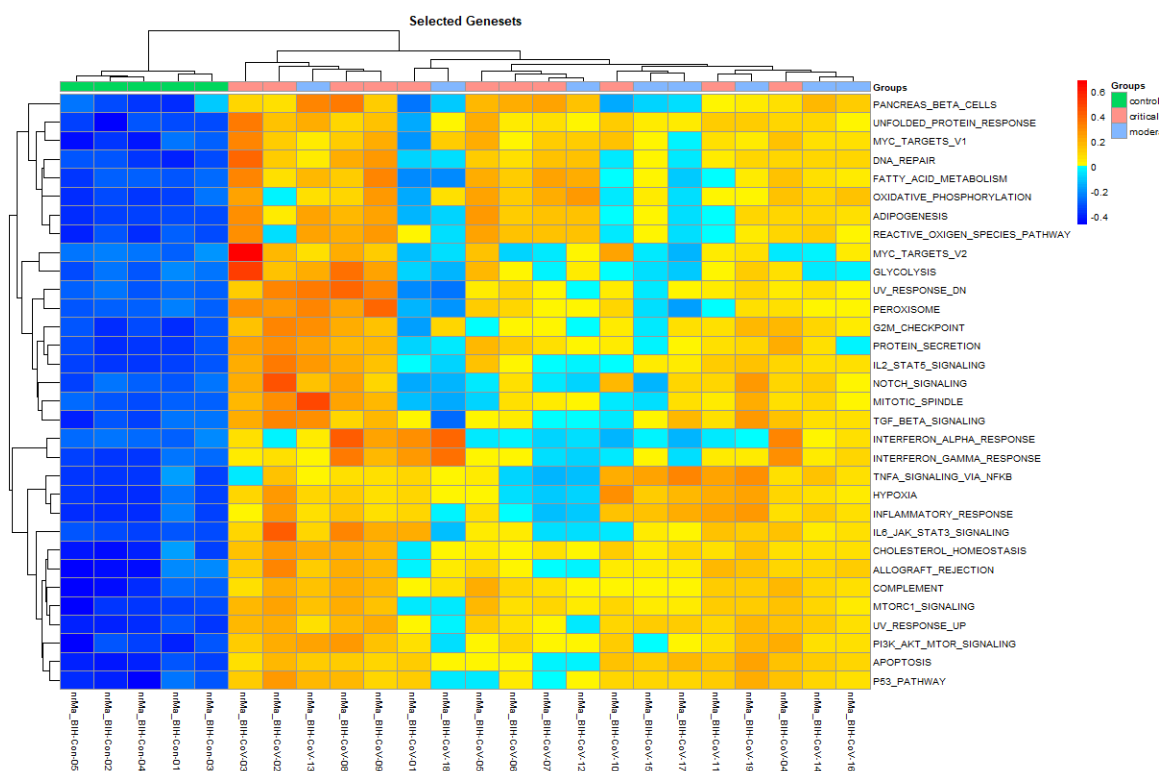


Figure 38. Analyse ROMA au niveau des cellules macrophages non-résidents.

Chaque colonne correspond à un individu. Les groupes sont indiqués dans la première ligne (contrôle, critique ou modéré). Les couleurs bleues sont en faveur d'une sous activation d'une voie de signalisation et le rouge d'une sur activation.

De façon très surprenante, un cluster net des contrôles se dégage et la totalité des voies de signalisations semble sous exprimée (Figure 38. Analyse ROMA au niveau des cellules macrophages non-résidents.. Ce motif semblant improbable, une vérification de la qualité des données originales a été faite et le manque de normalisation est clairement apparu.

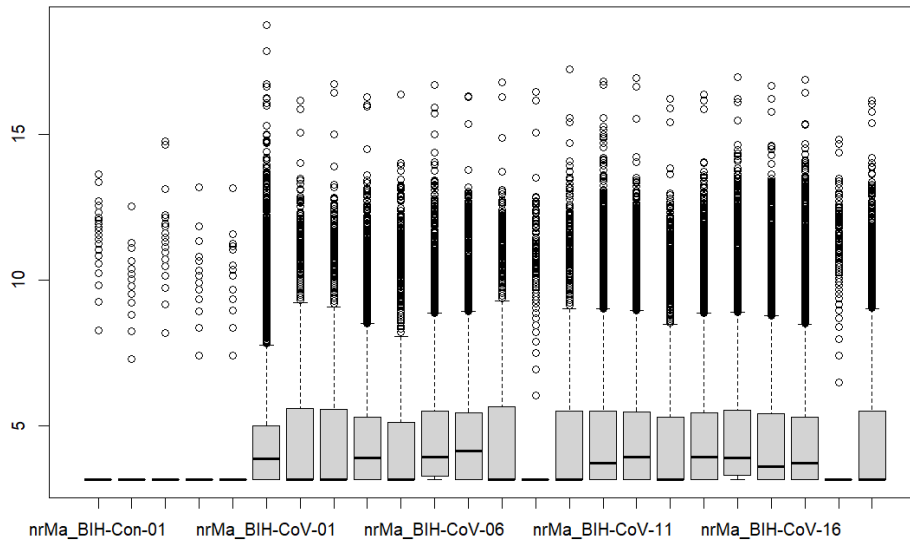


Figure 39. Boîte à moustache des données d'expression après le processus initial de traitement des données pour les cellules Macrophages non-résidentes.

L'absence de normalisation de l'échantillon étudié est apparente : les médianes des échantillons n'étant pas du tout alignées.

#### b. Une méthode plus robuste

L'étude du script a mis en évidence un problème de qualité des échantillons sélectionnés. Un nouveau script a donc été écrit en appliquant à la fois des filtres de qualité scRNA au niveau de l'objet Seurat puis des filtres de qualité « bulk RNA » via edgeR pour chaque sous type cellulaire. Pour vérification, l'export du *boxplot* a été fait de manière systématique (Figure 40. Boîte à moustache des données d'expression après le processus définitif de traitement des données.).

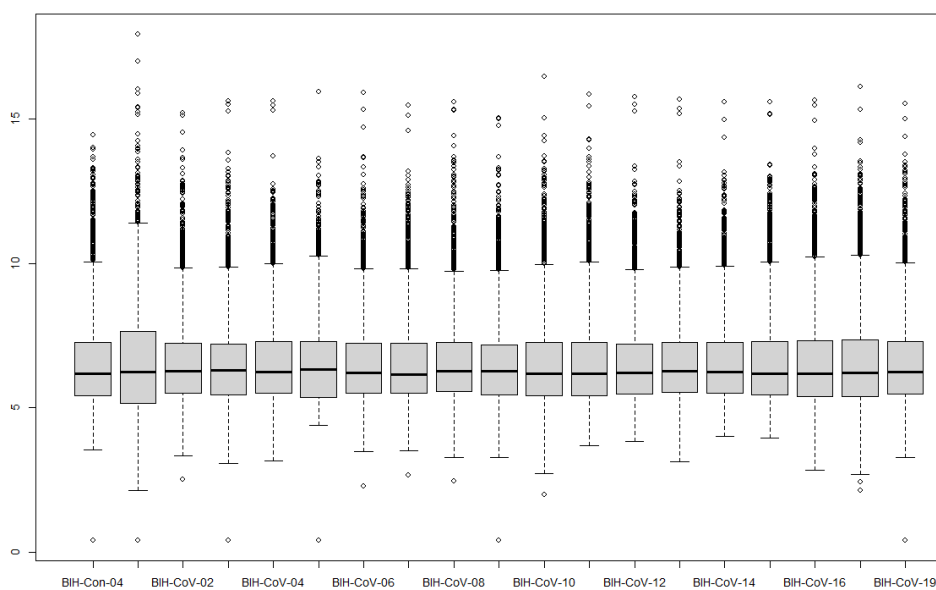


Figure 40. Boîte à moustache des données d'expression après le processus définitif de traitement des données.

Les médianes des échantillons sont maintenant alignées : ces données ont été utilisées pour les analyses suivantes.

#### 4. Des disparités d'expression entre les différents types cellulaires

Une analyse par rRoma (décrit dans l'introduction, page 27) a été effectuée dans les 17 types de façon indépendante en comparant les groupes de sévérité disponibles pour chaque échantillon (contrôle, modéré, sévère). Plusieurs sous-types ont mis en évidence des motifs intéressants. Nous nous sommes concentrés sur les modules *overdispersed non shifted* et les modules *shifted*. L'objectif était de mettre en évidence des voies de signalisation différentiellement activées entre les échantillons contrôles, covid-19 modérée et covid-19 sévère.

Il n'a pas été mis en évidence de différences de signalisation dans la majorité des types cellulaires. Des différences sont apparues uniquement dans trois types cellulaires : les macrophages (non-résidents), les neutrophiles et les cellules épithéliales squameuses. Nous allons présenter ceci en détail.

##### a. Les macrophages

###### *Contexte.*

Les macrophages sont des cellules sentinelles clés de la défense immunitaire au niveau du système respiratoire. Ils produisent en particulier des cytokines et des chimiokines qui sont des composants cruciaux de l'immunité innée et des médiateurs de l'immunopathologie (Allard et al., 2018). Des études récentes (Liao et al., 2020; Z. Xu et al., 2020) sur l'immunité des patients atteints de covid-19 ont montré que les cellules endommagées par l'infection par le SARS-CoV-2 déclenchaient une inflammation dans les poumons, largement médiée par les macrophages et les granulocytes pro-inflammatoires. Outre les lésions locales, les macrophages pro-inflammatoires libèrent des cytokines/chémo-attractants et déclenchent des réponses des cellules immunitaires adaptatives. Dans certains cas, cela a conduit à des syndromes de libération de cytokines, notamment le syndrome d'activation macrophagique (SAM), suivi d'une défaillance respiratoire, voire d'une défaillance multi-viscérale.

L'activation des macrophages leur confère une fonction et une plasticité hétérogènes qui dépendent du microenvironnement et de la durée de la stimulation, ce qui induit un spectre de phénotypes associés à différentes réponses inflammatoires. On peut ainsi distinguer plusieurs phénotypes d'activation (M1 ou M2) (Yunna et al., 2020). En général, les cellules M1 détruisent les pathogènes en produisant une grande quantité de cytokines pro-inflammatoires. En revanche, les cellules M2 présentent des propriétés anti-inflammatoires et une plus grande activité de phagocytose contre les agents pathogènes.

### *Résultats et discussion*

Concernant les macrophages, le cluster net observé Figure 38. Analyse ROMA au niveau des cellules macrophages non-résidents. n'est plus visible ici. En particulier, plusieurs échantillons contrôles ont été éliminés par les étapes de contrôles de qualité (Figure 41. Modules *overdispersed not shifted* dans les cellules macrophages non-résidentes.). La présence d'un unique contrôle rend difficile de conclure qu'il y a une différence entre les contrôles et les patients atteints de covid-19. Il semble néanmoins qu'il existe des différences d'expression entre les modérés (à gauche de la figure) et sévères (à droite) avec un contrôle ayant un motif d'expression à l'extrême des patients atteints de covid modéré. On note en particulier une augmentation de l'expression des voies TNFA\_SIGNALING\_VIA\_NFKB et IL2\_STAT5\_SIGNALING en faveur d'une activation plus importante par TNF $\alpha$  et qui a été associé à une sécrétion d'interférons (Akamatsu et al., 2021). Il existe également en particulier une augmentation assez homogène des modules APOPTOSIS et ANGIOGENESIS dans les formes sévères.

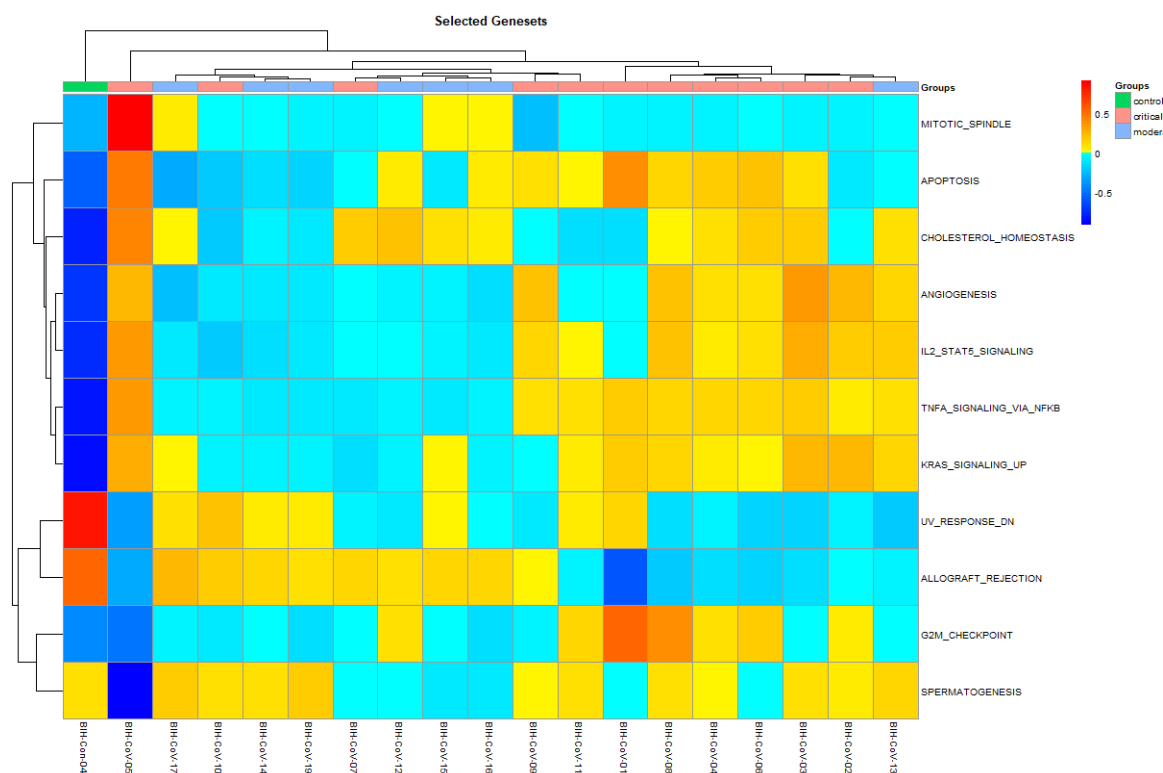


Figure 41. Modules *overdispersed not shifted* dans les cellules macrophages non-résidentes.

Il n’y avait pas de différence marquée au niveau des macrophages tissulaires ce qui est conforme aux données de la littérature selon lesquelles les macrophages infiltrant les poumons et d’autres organes dans les cas graves provenaient principalement de monocytes circulants périphériques plutôt que de types de macrophages résidant dans les tissus (Liao et al., 2020). Un autre point intéressant, montré par (Lian et al., 2022) sur des cellules iPSC, est que l’infection par le SARS-CoV-2 induisait une réaction inflammatoire dans les macrophages M1, entraînant l’expression de facteurs inflammatoires similaires à ceux observés chez les patients atteints de covid-19 sévères. En revanche, les macrophages M2 ont présenté une réaction inflammatoire minimale après l’infection par le SRAS-CoV-2. Pris ensemble, ces résultats suggèrent que les macrophages non-résidents, qui sont ici étudiés et qui présente une réaction inflammatoire chez les patients sévères, ont un phénotype majoritairement M1. Cela est cohérent avec le phénotype très inflammatoire de ces patients. Par ailleurs, une augmentation de l’apoptose a également été démontrée dans l’étude sus citée.

ROMA permet également l’export des gènes contribuant le plus aux différences d’activités observées entre les groupes. Les dix premiers gènes sont : SPP1, CCL20,

MT2A, CXCL1, SERPINB2, TNFAIP6, GPNMB. Ces gènes pourraient donc être impliqués dans la gravité de l'atteinte de la covid-19. Cette hypothèse est soutenue par le fait que plusieurs d'entre eux ont été décrits dans la covid-19. En particulier, CXCL1 et CCL20 sont des chimiokines impliquées dans l'inflammation et l'angiogenèse (Booz et al., 2020). SPP1 serait un médiateur commun de la réaction inflammatoire de la covid19 et de la polyarthrite rhumatoïde (MacDonald et al., 2021). TNFAIP6 est un marqueur classique de sévérité de la covid-19 (Nassir et al., 2021). GPNMB promeut la fibrose pulmonaire. En résumé, cette approche relativement basique semble identifier des gènes intéressants.

HALLMARK\_ANGIOGENESIS correspond aux gènes régulés lors de la formation des vaisseaux sanguins (angiogenèse). Ce résultat n'a pas été souligné par les études scRNA antérieures mais est cohérent avec la littérature et en particulier (Werlein et al., 2023) qui montre que l'atteinte cardiaque dans la COVID-19 est un processus inflammatoire angio-centrique piloté par les macrophages. L'activation de cette voie de signalisation pourrait donc constituer un critère de sévérité.

## b. Neutrophiles

### *Contexte*

Les neutrophiles sont les premières cellules à réagir à une infection et passent rapidement des vaisseaux sanguins vers les tissus. Il s'agit du leucocyte le plus abondant dans le sang représentant 40 à 60 % des cellules immunitaires circulantes chez les adultes en bonne santé (Rosales, 2018). Les neutrophiles éliminent les agents pathogènes via leur production de radicaux libres oxygénés (ROS) (Galani & Andreakos, 2015) ou de leurs pièges extracellulaires (*neutrophil extracellular trap*, NET) (Brinkmann et al., 2004).

La physiopathologie de la forme grave de la covid-19 est marquée par une altération de l'abondance, du phénotype et de la fonctionnalité des neutrophiles. Lors de l'infection par le SARS-CoV-2, un nombre élevé de neutrophiles a été observé dans l'épithélium nasopharyngé en particulier (Reusch et al., 2021).

## Résultats et discussion

L'analyse ROMA met en évidence des différences assez nettes entre les formes modérées et sévères de la covid-19. En particulier, l'analyse de la carte de chaleur et des données de ROMA montre que les modules HYPOXIA, INFLAMMATORY\_RESPONSE, IL2\_STAT5\_SIGNALING sont suractivés dans les neutrophiles de patients sévères (Figure 42. Carte de chaleur de l'analyse ROMA des cellules Neutrophiles.). Le sens de la PCA sur la représentation graphique n'est pas cohérent avec les données d'expression brutes. C'est une des limites de l'outil qui a déjà été décrit dans d'autres utilisations antérieures de cet outil. Les analyses ci-après prennent en considération le signe corrigé sur les expressions moyennes des groupes.

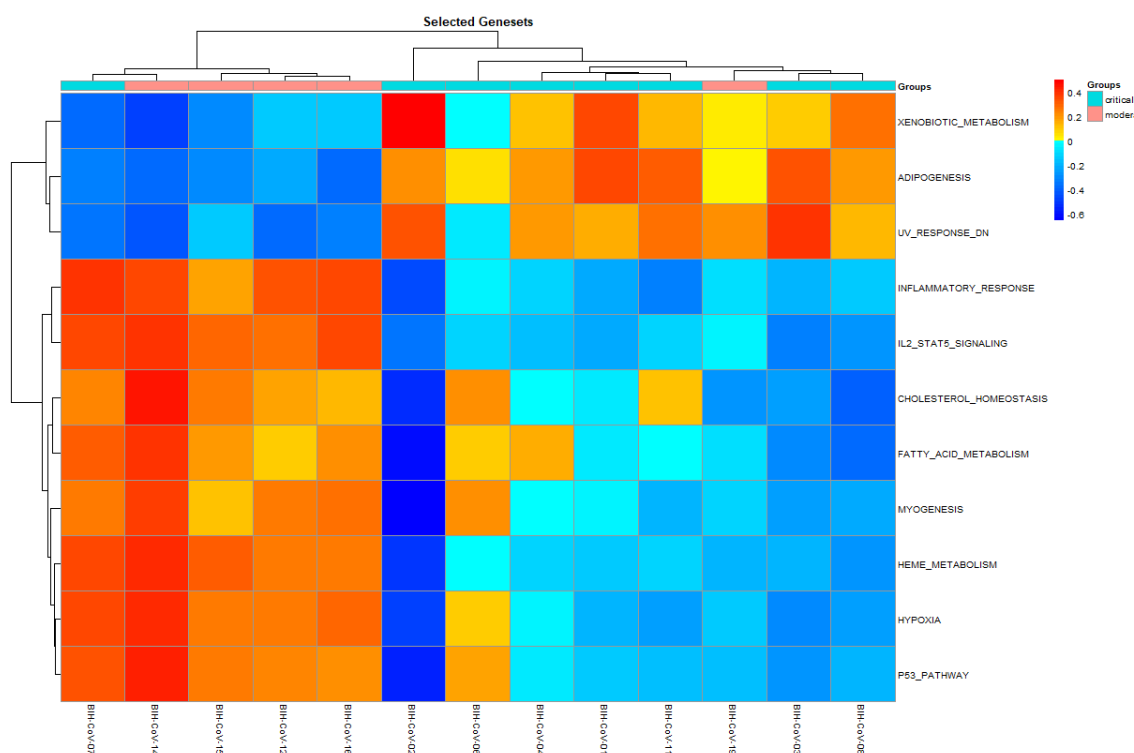


Figure 42. Carte de chaleur de l'analyse ROMA des cellules Neutrophiles.

Représentation des modules *Overdispersed not shifted*. Il n'y avait pas de cellules analysables chez les contrôles.

Il a déjà été rapporté que l'hypoxie peut contribuer à l'activation des neutrophiles, créant ainsi un cercle vicieux par une hyperréactivité des neutrophiles dans les tissus pulmonaires endommagés (Karmouty-Quintana et al., 2020). L'activation des voies inflammatoires est également assez cohérente avec la forte composante inflammatoire décrite chez les patients sévères.

*PLIN2* est présent dans les modules HYPOXIA et IL2\_STAT5\_SIGNALING. Il est dans ces deux modules le gène expliquant le plus l'*overdispersion*. Cette protéine entoure les gouttelettes lipidiques avec les phospholipides et participe au stockage des lipides neutres dans les gouttelettes lipidiques. Cette protéine n'a jamais été associée à la covid-19 ni à la dérégulation des neutrophiles. Il est intéressant de noter que l'accumulation de gouttelettes lipidiques induit la transformation des macrophages en cellules spumeuses c'est-à-dire « gonflée » de gouttelettes de graisse. La lyse des cellules spumeuses entraîne la formation de plaques d'athérosclérose qui se rompent et bloquent le vaisseau thrombotique (Son et al., 2016). Il est remarquable aussi que les gouttelettes lipidiques peuvent stocker les précurseurs des cytokines, principalement les acides arachidoniques, dans les neutrophiles. Les gouttelettes lipidiques modulent la réponse immunitaire des neutrophiles, car une lipolyse altérée des gouttelettes lipidiques s'accompagne d'une libération réduite d'acide arachidonique (Dichlberger et al., 2016). *PLIN2* est donc probablement un bon marqueur d'activation des neutrophiles (Paloschi et al., 2020) et *de facto* une piste intéressante comme biomarqueurs ou comme cible thérapeutique.

### c. Cellules épithéliales squameuses

Les cellules squameuses tapissent les bronches se ramifiant de la trachée jusqu'aux poumons. Elles sont un des types de cellules épithéliales qui sont présentes au niveau du tissu pulmonaire.

Le profil d'expression mis en évidence par ROMA est assez frappant puisqu'il met en évidence un gradient contrôle > modéré > sévère (Figure 43. Carte de chaleur de l'analyse ROMA des cellules squameuses montrant les modules *Overdispersed not shifted*). Il est intéressant de noter que les cellules immunitaires ne sont pas les seules à avoir des profils altérés et que les différences entre les différents stades de sévérité semblent relativement graduelles.

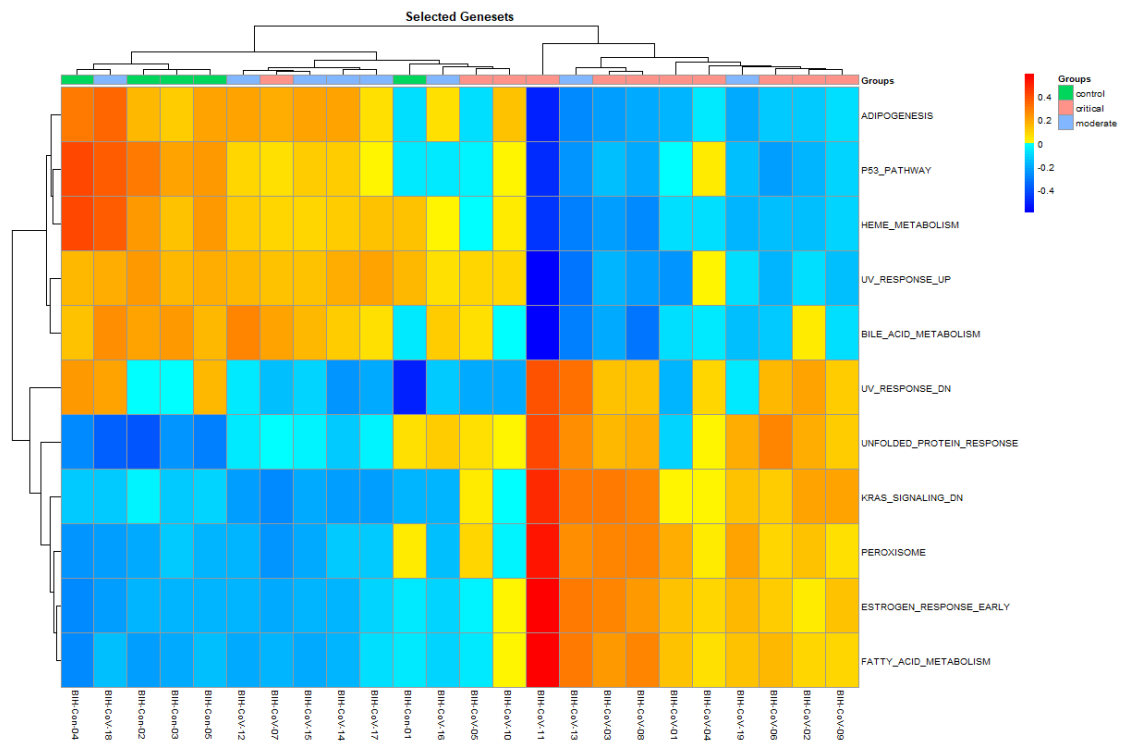


Figure 43. Carte de chaleur de l'analyse ROMA des cellules squameuses montrant les modules *Overdispersed not shifted*.

Enfin, nous avons visualisé ces données sur la carte de la mort cellulaire régulée décrite dans le chapitre précédent (page 40). L'analyse met en évidence une nette activation des caspases dans les neutrophiles des patients critiques (Figure 44. Carte de chaleur de l'analyse ROMA des cellules squameuses montrant les modules *Overdispersed not shifted* en analysant les modules de la carte de la mort cellulaire régulée (Ravel et al., 2020)).

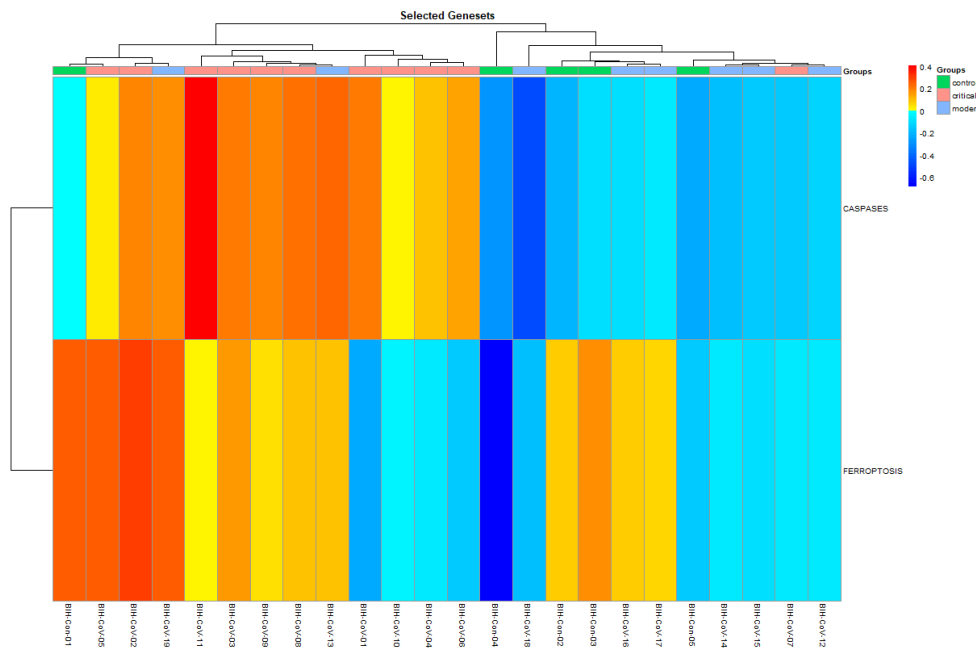


Figure 44. Carte de chaleur de l'analyse ROMA des cellules squameuses montrant les modules *Overdispersed not shifted* en analysant les modules de la carte de la mort cellulaire régulée (Ravel et al., 2020)

#### d. Des mécanismes différents de la grippe ?

Les agents précédemment utilisés pour traiter le SARS-CoV1, le MERS ou même le virus de la grippe sont des candidats potentiels pour traiter la covid-19 (Sanders et al., 2020). Comparer les mécanismes de la covid-19 avec des maladies des voies respiratoires comme la grippe est donc essentiel avant d'utiliser une approche de *drug repositionning*. Plusieurs traitement ont d'ailleurs ainsi été étudiés comme le Umifenovir (Sanders et al., 2020)

L'incidence accrue de la morbidité et de la mortalité associées au SARS-CoV-2 par rapport au virus de la grippe A semblent due à des réponses immunitaires spécifiques de l'hôte ainsi qu'au type et à l'étendue des lésions tissulaires causées par le virus (Flerlage et al., 2021). L'hyper-inflammation provoquée par le virus SARS-Cov2 contribue aux manifestations graves de la maladie par l'augmentation des réponses de cytokines et chimiokines pro-inflammatoires créant une « tempête de cytokines ».

Nous avons donc appliqué de nouveau notre approche par l'algorithme ROMA pour vérifier si des différences entre la covid-19 et la grippe pouvait être mis en évidence. Les

données proviennent d'une publication différente décrite plus haut (K. M. Gao et al., 2021).

Il existe par exemple une différence assez marquée au niveau des macrophages M1 avec une forte expression de `TNF1_SIGNALING_VIA_NFKB` et `INFLAMMATORY_RESPONSE` correspondant à une augmentation de l'expression de `NFKBIA` et `IL1B` (Figure 45. Analyse ROMA des macrophages M1, modules *shifted*). Assez peu de différences ont été mises en évidence parmi les autres types cellulaires, ce qui est probablement en lien avec le faible nombre d'individus étudiés retenus après le contrôle qualité. Il sera essentiel de reproduire cette analyse sur un autre groupe de données.

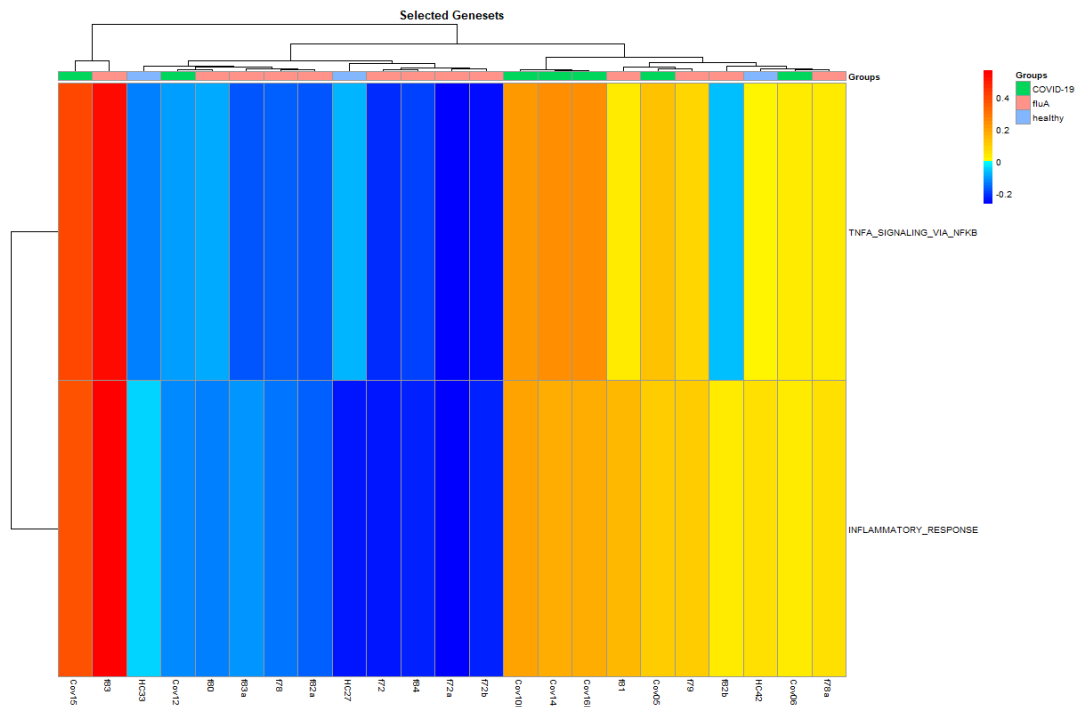


Figure 45. Analyse ROMA des macrophages M1, modules *shifted*.

## 5. Discussion

Nous avons pu utiliser l'approche ROMA dans deux groupes de données différents et mis en évidence des différences d'activation au niveau de plusieurs sous-types cellulaires et en particulier les macrophages et les neutrophiles. Les résultats ne se limitent pas aux cellules immunitaires puisque les cellules squameuses révèlent des différences intéressantes entre les patients et contrôles. Finalement, nous avons pu réappliquer cette approche pour

comparer deux pathologies proches mais avec des différences phénotypiques marquées : la grippe et la covid-19.

La reproductibilité est une des questions majeures de l'analyse de scRNA : il est nécessaire qu'un résultat soit au moins confirmé par deux études. Une des limites majeures de ce type d'approche est la gestion de l'évolutivité d'une pathologie dans le temps et la gestion des comorbidités. Cet aspect n'a pas encore été traité dans ce travail mais devra être approfondi en vue d'une publication. De même, l'extrapolation de ces résultats se doit prudente puisque très dépendante des échantillons utilisés. En plus de la discordance entre les études, les signatures moléculaires sont très dépendantes des échantillons utilisés (nasopharyngés, lavement bronchiques, sang périphérique).

Ce travail souligne l'importance d'un contrôle qualité exemplaire afin de pouvoir garantir une analyse correcte en aval. De plus, les résultats obtenus vont grandement dépendre de la qualité de l'annotation des cellules et donc de l'algorithme utilisé.

#### Perspective : création d'un réseau d'inférence avec Pypath

L'algorithme ROMA permet d'extraire les gènes qui expliquent le plus les différences d'activation entre les modules. Il semble pertinent de construire des réseaux d'inférence à partir des gènes ayant le poids le plus élevé. Cela pourrait reposer par exemple sur une pipeline telle que celle proposée par Marco Ruscone : [https://github.com/marcorusc/Pypath\\_pipeline](https://github.com/marcorusc/Pypath_pipeline) utilisant la base de données Omnipath qui englobe de nombreuses bases de réseaux construites et validées manuellement par des experts. L'idée est de construire un réseau automatisé en reliant les protéines grâce à des bases de données existantes.

## 4. Conclusions et perspectives

Cette thèse est construite autour de deux axes complémentaires : d'une part une approche pragmatique d'amélioration de la pratique en génétique médicale et d'autre part une proposition d'analyse de voies de signalisation à partir de rRoma à la fois sur des données *bulk* et en cellule unique.

### Le premier axe

GeneTree est un outil simple de prise en main qui permet un gain de temps dans la pratique de la génétique médicale et un archivage numérique rapide des données. Il permet en une saisie la création de l'arbre généalogique, de scores de risque nécessaires à la prise en charge du patient et à l'élaboration du compte rendu.

Cet outil permet également d'impliquer le patient dans sa prise en charge et de raccourcir la prise de notes du praticien. L'intégration du patient au processus de soin (« Patient-centered care ») se développe en particulier dans le cadre des maladies rares où, vu la rareté de certaines pathologies, le patient – et ses parents – connaissent parfois mieux que le praticien la pathologie sur certains aspects (NEJM Catalyst, 2017). Le principal objectif et avantage des soins centrés sur le patient est d'améliorer la prise en charge de santé d'un individu donné (et non plus la population générale). Non seulement les patients en bénéficient, mais l'idée est qu'ensuite le système de soins en profite également, grâce à : (1) L'amélioration des taux de satisfaction des patients et de leurs familles, (2) l'amélioration du moral et de la productivité des cliniciens et du personnel auxiliaire. (3) l'amélioration de l'affectation des ressources. (4) la réduction des dépenses tout au long du continuum de soins.

La bio-informatique a révolutionné la pratique de la génétique médicale. Comme exposé précédemment, le séquençage de première génération était fondé sur l'étude de séquences connues. Le séquençage haut-débit, lui, l'est sur la génération de séquences aléatoires : les outils bio-informatiques sont ensuite utilisés pour aligner les séquences sur des régions génomiques connues. Le séquençage se décompose ainsi en deux étapes : la technique moléculaire puis la bio-informatique. Si l'on considère l'étude du séquençage de l'exome

ou du génome, l'analyse des séquences obtenues va consister en l'interprétation des variations de la séquence étudiée par rapport à un génome de référence synthétique. Cette interprétation va également se fonder sur des outils bio-informatiques : bases de données patients ou contrôles, prédiction de pathogénicité, etc. Sur cette étape de nombreux outils sont publiés, et on pourra par exemple citer la publication de Alphafold (Jumper et al., 2021). Il est probable que le paysage de la génétique soit transformé par les apports de l'*artificial intelligence* aussi bien concernant la priorisation des variations, que la découverte de nouveau mécanisme de pathogénicité de variation, ou des prédictions de pathogénicité (Qi et al., 2021). D'autre part, certains outils permettent également d'extraire d'autres données au-delà des variations de séquences ponctuelles. On pourra citer par exemple la détection des variations de nombre de copie (CNV) ou des disomies (Quinodoz et al., 2021).

#### Le second axe

Deux chapitres de cette thèse utilisent la biologie des systèmes pour étudier deux maladies communes : la mort cellulaire dans le cancer et la covid19. Mon travail a abouti à la création d'une carte représentant les processus impliqués dans la mort cellulaire ainsi que la démonstration de son application dans deux cas concrets : la comorbidité inverse entre la maladie d'Alzheimer et le cancer du poumon non à petites cellules et la description de sous-types de cancer de l'ovaire. J'ai pu également souligner plusieurs limites de ce travail comme le fait de maintenir une telle ressource à jour et la complexité de vérifier l'exhaustivité des informations recueillies. L'étude de la covid-19 au niveau cellulaire, a permis de développer une méthode d'analyse basée sur l'algorithme ROMA en utilisant des données de cellules uniques (*scRNA seq*). Après avoir mis en évidence les difficultés techniques d'une telle analyse, j'ai pu montrer des différences d'expressions de certaines voies de signalisation sur trois types cellulaires : les macrophages, les neutrophiles et les cellules épithéliales squameuses. Cette analyse a mis en exergue plusieurs gènes candidats intéressants dont *SPP1*, *CCL20*, *MT2A*, *CXCL1*, *SERPINB2*, *TNFAIP6*, *GPNMB*. En particulier, ce travail a permis de suggérer l'implication potentielle de l'angiogenèse au centre de la réaction inflammatoire des macrophages. De plus, cette approche semble également pertinente pour comparer plusieurs pathologies comme le confirme l'étude de

la covid et du virus de la grippe avec la mise en évidence d'une signature inflammatoire dans les macrophages.

La biologie des systèmes permet une intégration puissante de larges groupes de données et il me semble que, par exemple, son utilisation pourrait être développée pour la description de nouveaux gènes clés. Une approche qui serait intéressante à développer serait de déterminer des signatures moléculaires spécifiques à des variations pathogènes de certains gènes afin de faciliter l'interprétations des variations de signification incertaine : une signature superposable à une variation pathogène serait en faveur de sa pathogénicité. De la même façon, à l'instar des GWAS, la biologie des systèmes peut également permettre d'identifier des gènes candidats à des pathologies monogéniques. Des approches superposables ont été proposées dans certains gènes impliqués dans les phénomènes de méthylation (Courraud et al., 2021).

La numérisation et l'intégration des données cliniques (phénotypes, histoire familiale) et génétique (génomique, transcriptome, épigénomique, protéome) grâce à des outils bio-informatiques est au centre de la médecine dite de précision. Ces outils vont ainsi faciliter chaque étape du parcours de soin et de recherche. En particulier, l'usage de la biologie des systèmes sera probablement un des piliers de l'analyse de toutes les données obtenues avec sa formidable capacité d'intégrer plusieurs types de données à haut débit. Enfin ce travail de thèse souligne également l'intérêt de réinterpréter des données déjà générées et de la nécessité de poursuivre l'effort de mise à disposition de nombreux groupes de données.

## Références

- Akamatsu, M. A., Castro, J. T. de, Takano, C. Y., & Ho, P. L. (2021). Off balance: Interferons in COVID-19 lung infections. *EBioMedicine*, *73*. <https://doi.org/10.1016/j.ebiom.2021.103642>
- Allard, B., Panariti, A., & Martin, J. G. (2018). Alveolar Macrophages in the Resolution of Inflammation, Tissue Repair, and Tolerance to Infection. *Frontiers in Immunology*, *9*, 1777. <https://doi.org/10.3389/fimmu.2018.01777>
- Arntzen, M. Ø., & Thiede, B. (2012). ApoptoProteomics, an integrated database for analysis of proteomics data obtained from apoptotic cells. *Mol. Cell Proteom*, *11*.
- Asgari, S., & Pousaz, L. A. (2021). Human genetic variants identified that affect COVID susceptibility and severity. *Nature*, *600*(7889), 390–391. <https://doi.org/10.1038/d41586-021-01773-7>
- Ashkenazi, A., & Dixit, V. M. (1998). Death receptors: Signalling and modulation. *Science*, *281*, 1305–1308.
- Avril, T., Vauléon, E., & Chevet, E. (2017). Endoplasmic reticulum stress signalling and chemotherapy resistance in solid cancers. *Oncogenesis*, *6*, 373.
- Balci, H., Siper, M. C., Saleh, N., Safarli, I., Roy, L., Kilicarslan, M., Ozaydin, R., Mazein, A., Auffray, C., Babur, Ö., Demir, E., & Dogrusoz, U. (2021). Newt: A comprehensive web-based tool for viewing, constructing and analyzing biological maps. *Bioinformatics*, *37*(10), 1475–1477. <https://doi.org/10.1093/bioinformatics/btaa850>
- Balmaña, J., Díez, O., Rubio, I. T., Cardoso, F., & ESMO Guidelines Working Group. (2011). BRCA in breast cancer: ESMO Clinical Practice Guidelines. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, *22 Suppl 6*, vi31-34. <https://doi.org/10.1093/annonc/mdr373>
- Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., ... Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, *462*(7269), Article 7269. <https://doi.org/10.1038/nature08460>
- Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D. W., Dao, F., Dhir, R., DiSaia, P., Gabra, H., & Glenn, P. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*, *474*, 609–615.
- Birch, J. M., Alston, R. D., McNally, R. J., Evans, D. G., Kelsey, A. M., Harris, M., Eden, O. B., & Varley, J. M. (2001). Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene*, *20*(34), 4621–4628. <https://doi.org/10.1038/sj.onc.1204621>
- Bloemenkamp, K. W., Rosendaal, F. R., Helmerhorst, F. M., Büller, H. R., & Vandenbroucke, J. P. (1995). Enhancement by factor V Leiden mutation of risk of deep-vein thrombosis associated with oral contraceptives containing a third-generation

progestagen. *Lancet (London, England)*, 346(8990), 1593–1596.  
[https://doi.org/10.1016/s0140-6736\(95\)91929-5](https://doi.org/10.1016/s0140-6736(95)91929-5)

Bonnet, E., Calzone, L., Rovera, D., Stoll, G., Barillot, E., & Zinovyev, A. (2013). BiNoM 2.0, a Cytoscape plugin for accessing and analyzing pathways using standard systems biology formats. *BMC Syst. Biol*, 7, 18.

Booz, G. W., Altara, R., Eid, A. H., Wehbe, Z., Fares, S., Zaraket, H., Habeichi, N. J., & Zouein, F. A. (2020). Macrophage responses associated with COVID-19: A pharmacological perspective. *European Journal of Pharmacology*, 887, 173547.  
<https://doi.org/10.1016/j.ejphar.2020.173547>

Bowness, P. (2015). HLA-B27. *Annual Review of Immunology*, 33, 29–48.  
<https://doi.org/10.1146/annurev-immunol-032414-112110>

Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D. S., Weinrauch, Y., & Zychlinsky, A. (2004). Neutrophil extracellular traps kill bacteria. *Science (New York, N.Y.)*, 303(5663), 1532–1535.  
<https://doi.org/10.1126/science.1092385>

Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., Mittag, F., Schubert, M., Glont, M., Golebiewski, M., van Iersel, M., Keating, S., Rall, M., Wybrow, M., Hermjakob, H., Hucka, M., Kell, D. B., Müller, W., Mendes, P., ... Le Novère, N. (2013). Path2Models: Large-scale generation of computational models from biochemical pathway maps. *BMC Systems Biology*, 7, 116. <https://doi.org/10.1186/1752-0509-7-116>

Caputo, S. M., Golmard, L., Léone, M., Damiola, F., Guillaud-Bataille, M., Revillion, F., Rouleau, E., Derive, N., Buisson, A., Basset, N., Schwartz, M., Vilquin, P., Garrec, C., Privat, M., Gay-Bellile, M., Abadie, C., Abidallah, K., Airaud, F., Allary, A.-S., ... Stoppa-Lyonnet, D. (2021). Classification of 101 BRCA1 and BRCA2 variants of uncertain significance by cosegregation study: A powerful approach. *The American Journal of Human Genetics*, 108(10), 1907–1923.  
<https://doi.org/10.1016/j.ajhg.2021.09.003>

Carver, T., Cunningham, A. P., Babb de Villiers, C., Lee, A., Hartley, S., Tischkowitz, M., Walter, F. M., Easton, D. F., & Antoniou, A. C. (2018). pedigreejs: A web-based graphical pedigree editor. *Bioinformatics*, 34(6), 1069–1071.  
<https://doi.org/10.1093/bioinformatics/btx705>

Carver, T., Hartley, S., Lee, A., Cunningham, A. P., Archer, S., Babb de Villiers, C., Roberts, J., Ruston, R., Walter, F. M., Tischkowitz, M., Easton, D. F., & Antoniou, A. C. (2021). CanRisk Tool-A Web Interface for the Prediction of Breast and Ovarian Cancer Risk and the Likelihood of Carrying Genetic Pathogenic Variants. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 30(3), 469–473.  
<https://doi.org/10.1158/1055-9965.EPI-20-1319>

Carvunis, A.-R., Gomez, E., Thierry-Mieg, N., Trilling, L., & Vidal, M. (2009). Biologie systémique—Des concepts d’hier aux découvertes de demain. *médecine/sciences*, *25*(6–7), Article 6–7. <https://doi.org/10.1051/medsci/2009256-7578>

Chaabane, W., User, S. D., El-Gazzah, M., Jaksik, R., Sajjadi, E., Rzeszowska-Wolny, J., & Los, M. J. (2013). Autophagy, apoptosis, mitoptosis and necrosis: Interdependence between those pathways and effects on cancer. *Arch. Immunol. Ther. Exp*, *61*, 43–58.

Chen, Y.-M., Zheng, Y., Yu, Y., Wang, Y., Huang, Q., Qian, F., Sun, L., Song, Z.-G., Chen, Z., Feng, J., An, Y., Yang, J., Su, Z., Sun, S., Dai, F., Chen, Q., Lu, Q., Li, P., Ling, Y., ... Zhang, Y.-Z. (2020). Blood molecular markers associated with COVID-19 immunopathology and multi-organ damage. *The EMBO Journal*, *39*(24), e105896. <https://doi.org/10.15252/embj.2020105896>

Choi, K., Ryu, S.-W., Song, S., Choi, H., Kang, S. W., & Choi, C. (2010). Caspase-dependent generation of reactive oxygen species in human astrocytoma cells contributes to resistance to TRAIL-mediated apoptosis. *Cell Death Differ*, *17*, 833–845.

Chua, R. L., Lukassen, S., Trump, S., Hennig, B. P., Wendisch, D., Pott, F., Debnath, O., Thürmann, L., Kurth, F., Völker, M. T., Kazmierski, J., Timmermann, B., Twardziok, S., Schneider, S., Machleidt, F., Müller-Redetzky, H., Maier, M., Krannich, A., Schmidt, S., ... Eils, R. (2020). COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nature Biotechnology*, *38*(8), Article 8. <https://doi.org/10.1038/s41587-020-0602-4>

Collins, F. S. (1992). Positional cloning: Let’s not call it reverse anymore. *Nature Genetics*, *1*(1), 3–6. <https://doi.org/10.1038/ng0492-3>

Collins, F. S., & Fink, L. (1995). The Human Genome Project. *Alcohol Health and Research World*, *19*(3), 190–195.

Courraud, J., Chater-Diehl, E., Durand, B., Vincent, M., Del Mar Muniz Moreno, M., Boujelbene, I., Drouot, N., Genschik, L., Schaefer, E., Nizon, M., Gerard, B., Abramowicz, M., Cogné, B., Bronicki, L., Burglen, L., Barth, M., Charles, P., Colin, E., Coubes, C., ... Piton, A. (2021). Integrative approach to interpret DYRK1A variants, leading to a frequent neurodevelopmental disorder. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *23*(11), 2150–2159. <https://doi.org/10.1038/s41436-021-01263-1>

Critchley, J. A., & Unal, B. (2003). Health effects associated with smokeless tobacco: A systematic review. *Thorax*, *58*(5), 435–443. <https://doi.org/10.1136/thorax.58.5.435>

Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., & Kamdar, M. R. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Res*, *42*, 472–477.

Cui, J., Li, F., & Shi, Z.-L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews. Microbiology*, *17*(3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>

Czabotar, P. E., Lessene, G., Strasser, A., & Adams, J. M. (2014). Control of apoptosis by the BCL-2 protein family: Implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol*, *15*, 49–63.

David, K. K., Andrabi, S. A., Dawson, T. M., & Dawson, V. L. (2009). Parthanatos, a messenger of death. *Front. Biosci*, *14*, 1116–1128.

Davies, L. C., Jenkins, S. J., Allen, J. E., & Taylor, P. R. (2013). Tissue-resident macrophages. *Nature Immunology*, *14*(10), 986–995. <https://doi.org/10.1038/ni.2705>

Dichlberger, A., Schlager, S., Kovanen, P. T., & Schneider, W. J. (2016). Lipid droplets in activated mast cells – a significant source of triglyceride-derived arachidonic acid for eicosanoid production. *European Journal of Pharmacology*, *785*, 59–69. <https://doi.org/10.1016/j.ejphar.2015.07.020>

Dickens, L. S., Powley, I. R., Hughes, M. A., & MacFarlane, M. (2012). The ‘complexities’ of life and death: Death receptor signalling platforms. *Exp. Cell Res*, *318*, 1269–1277.

*Differential gene expression (DGE) analysis*. (n.d.). Training-Modules. Retrieved 19 August 2023, from [https://hbctraining.github.io/Training-modules/planning\\_successful\\_rnaseq/lessons/sample\\_level\\_QC.html](https://hbctraining.github.io/Training-modules/planning_successful_rnaseq/lessons/sample_level_QC.html)

Doctor, K. S., Reed, J. C., Godzik, A., & Bourne, P. E. (2003). The apoptosis database. *Cell Death Differ*, *10*, 621–633.

Ellis, R. E., Yuan, J. Y., & Horvitz, H. R. (1991). Mechanisms and functions of cell death. *Annu. Rev. Cell Biol*, *7*, 663–698.

Etemadmoghadam, D., deFazio, A., Beroukhi, R., Mermel, C., George, J., Getz, G., Tothill, R., Okamoto, A., Raeder, M. B., & Harnett, P. (2009). Integrated Genome-Wide DNA Copy Number and Expression Analysis Identifies Distinct Mechanisms of Primary Chemoresistance in Ovarian Carcinomas. *Clin. Cancer Res*, *15*, 1417–1427.

Fatokun, A. A., Dawson, V. L., & Dawson, T. M. (2014). Parthanatos: Mitochondrial-linked mechanisms and therapeutic opportunities. *Br. J. Pharmacol*, *171*, 2000–2016.

FitzGerald, M. G., Marsh, D. J., Wahrer, D., Bell, D., Caron, S., Shannon, K. E., Ishioka, C., Isselbacher, K. J., Garber, J. E., Eng, C., & Haber, D. A. (1998). Germline mutations in PTEN are an infrequent cause of genetic predisposition to breast cancer. *Oncogene*, *17*(6), 727–731. <https://doi.org/10.1038/sj.onc.1201984>

Flerlage, T., Boyd, D. F., Meliopoulos, V., Thomas, P. G., & Schultz-Cherry, S. (2021). Influenza virus and SARS-CoV-2: Pathogenesis and host responses in the respiratory tract. *Nature Reviews. Microbiology*, *19*(7), 425–441. <https://doi.org/10.1038/s41579-021-00542-7>

Fujita, K. A., Ostaszewski, M., Matsuoka, Y., Ghosh, S., Glaab, E., Trefois, C., Crespo, I., Perumal, T. M., Jurkowski, W., Antony, P. M. A., Diederich, N., Buttini, M., Kodama, A., Satagopam, V. P., Eifes, S., Del Sol, A., Schneider, R., Kitano, H., & Balling, R. (2014). Integrating pathways of Parkinson’s disease in a molecular interaction map. *Molecular Neurobiology*, *49*(1), 88–102. <https://doi.org/10.1007/s12035-013-8489-4>

Funahashi, A., Morohashi, M., Matsuoka, Y., Jouraku, A., & Kitano, H. (2007). CellDesigner: A Graphical Biological Network Editor and Workbench Interfacing Simulator. In S. Choi (Ed.), *Introduction to Systems Biology* (pp. 422–434). Humana Press. [https://doi.org/10.1007/978-1-59745-531-2\\_21](https://doi.org/10.1007/978-1-59745-531-2_21)

Galani, I. E., & Andreakos, E. (2015). Neutrophils in viral infections: Current concepts and caveats. *Journal of Leukocyte Biology*, *98*(4), 557–564. <https://doi.org/10.1189/jlb.4VMR1114-555R>

Galluzzi, L., Kepp, O., Krautwald, S., Kroemer, G., & Linkermann, A. (2014). Molecular mechanisms of regulated necrosis. *Semin. Cell Dev. Biol*, *35*, 24–32.

Galluzzi, L., Kepp, O., Trojel-Hansen, C., & Kroemer, G. (2012). Mitochondrial control of cellular life, stress, and death. *Circ. Res*, *111*, 1198–1207.

Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., Alnemri, E. S., Altucci, L., Amelio, I., & Andrews, D. W. (2018). Molecular mechanisms of cell death: Recommendations of the Nomenclature Committee on Cell Death 2018. *Cell Death Differ*, *25*, 486–541.

Gao, K. M., Derr, A. G., Guo, Z., Nündel, K., Marshak-Rothstein, A., Finberg, R. W., & Wang, J. P. (2021). Human nasal wash RNA-Seq reveals distinct cell-specific innate immune responses in influenza versus SARS-CoV-2. *JCI Insight*, *6*(22). <https://doi.org/10.1172/jci.insight.152288>

Gao, Y., Shi, C., Chen, Y., Shi, P., Liu, J., Xiao, Y., Shen, Y., & Chen, E. (2020). A cluster of the Corona Virus Disease 2019 caused by incubation period transmission in Wuxi, China. *The Journal of Infection*, *80*(6), 666–670. <https://doi.org/10.1016/j.jinf.2020.03.042>

Gao, Z., Xu, Y., Sun, C., Wang, X., Guo, Y., Qiu, S., & Ma, K. (2021). A systematic review of asymptomatic infections with COVID-19. *Journal of Microbiology, Immunology, and Infection = Wei Mian Yu Gan Ran Za Zhi*, *54*(1), 12–16. <https://doi.org/10.1016/j.jmii.2020.05.001>

García-Campos, M. A., Espinal-Enríquez, J., & Hernández-Lemus, E. (2015). Pathway Analysis: State of the Art. *Frontiers in Physiology*, *6*. <https://www.frontiersin.org/articles/10.3389/fphys.2015.00383>

Gawron, P., Ostaszewski, M., Satagopam, V., Gebel, S., Mazein, A., Kuzma, M., Zorzan, S., McGee, F., Otjacques, B., & Balling, R. (2016). MINERVA—a platform for visualization and curation of molecular interaction networks. *NPJ Syst. Biol. Appl*, *2*, 16020.

Geng, J., Ito, Y., Shi, L., Amin, P., Chu, J., Ouchida, A. T., Mookhtiar, A. K., Zhao, H., Xu, D., & Shan, B. (2017). Regulation of RIPK1 activation by TAK1-mediated phosphorylation dictates apoptosis and necroptosis. *Nat Commun*, *8*, 359.

Giampietri, C., Petrongaro, S., Conti, S., Facchiano, A., Filippini, A., & Ziparo, E. (2015). Cancer microenvironment and endoplasmic reticulum stress response. *Mediat. Inflamm*.

Gilks, C. B., & Prat, J. (2009). Ovarian carcinoma pathology and genetics: Recent advances. *Hum. Pathol*, *40*, 1213–1223.

Guan, W.-J., Ni, Z.-Y., Hu, Y., Liang, W.-H., Ou, C.-Q., He, J.-X., Liu, L., Shan, H., Lei, C.-L., Hui, D. S. C., Du, B., Li, L.-J., Zeng, G., Yuen, K.-Y., Chen, R.-C., Tang, C.-L., Wang, T., Chen, P.-Y., Xiang, J., ... China Medical Treatment Expert Group for Covid-19. (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England Journal of Medicine*, *382*(18), 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. <https://doi.org/10.1158/2159-8290.CD-21-1059>

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70. [https://doi.org/10.1016/s0092-8674\(00\)81683-9](https://doi.org/10.1016/s0092-8674(00)81683-9)

Hardy, J., & Singleton, A. (2009). Genomewide association studies and human disease. *The New England Journal of Medicine*, *360*(17), 1759–1768. <https://doi.org/10.1056/NEJMra0808700>

Hemedan, A. A., Niarakis, A., Schneider, R., & Ostaszewski, M. (2022). Boolean modelling as a logic-based dynamic approach in systems medicine. *Computational and Structural Biotechnology Journal*, *20*, 3161–3172. <https://doi.org/10.1016/j.csbj.2022.06.035>

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9362–9367. <https://doi.org/10.1073/pnas.0903103106>

Hitomi, J., Christofferson, D. E., Ng, A., Yao, J., Degterev, A., Xavier, R. J., & Yuan, J. (2008). Identification of a Molecular Signalling Network that Regulates a Cellular Necrotic Cell Death Pathway. *Cell*, *135*, 1311–1323.

Hoffmann, R. (2008). A wiki for the life sciences where authorship matters. *Nat. Genet*, *40*, 1047–1051.

Hokama, M., Oka, S., Leon, J., Ninomiya, T., Honda, H., Sasaki, K., Iwaki, T., Ohara, T., Sasaki, T., & LaFerla, F. M. (2014). Altered expression of diabetes-related genes in Alzheimer's disease brains: The Hisayama study. *Cereb. Cortex*, *24*, 2476–2488.

Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews. Microbiology*, *19*(3), 141–154. <https://doi.org/10.1038/s41579-020-00459-7>

Hunter, D. J. (2005). Gene-environment interactions in human diseases. *Nature Reviews. Genetics*, *6*(4), 287–298. <https://doi.org/10.1038/nrg1578>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler,

J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

Kaczmarek, A., Vandenabeele, P., & Krysko, D. V. (2013). Necroptosis: The release of damage-associated molecular patterns and its physiological relevance. *Immunity*, *38*, 209–223.

Kajiwarra, Y., McKenzie, A., Dorr, N., Gama Sosa, M. A., Elder, G., Schmeidler, J., Dickstein, D. L., Bozdagi, O., Zhang, B., & Buxbaum, J. D. (2016). The human-specific CASP4 gene product contributes to Alzheimer-related synaptic and behavioural deficits. *Hum. Mol. Genet.*, *25*, 4315–4327.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*, *40*, 109–114.

Kanehisa, M. T. K. E. G. G. database. (2002). *Novartis Found Symp* (Vol. 247, pp. 244–252).

Kang, K., Lee, S.-R., Piao, X., & Hur, G. M. (2019). Post-translational modification of the death receptor complex as a potential therapeutic target in cancer. *Arch Pharm Res*, *42*, 76–87.

Karmouty-Quintana, H., Thandavarayan, R. A., Keller, S. P., Sahay, S., Pandit, L. M., & Akkanti, B. (2020). Emerging Mechanisms of Pulmonary Vasoconstriction in SARS-CoV-2-Induced Acute Respiratory Distress Syndrome (ARDS) and Potential Therapeutic Targets. *International Journal of Molecular Sciences*, *21*(21), 8081. <https://doi.org/10.3390/ijms21218081>

Kaushal, V., Dye, R., Pakavathkumar, P., Foveau, B., Flores, J., Hyman, B., Ghetti, B., Koller, B. H., & LeBlanc, A. C. (2015). Neuronal NLRP1 inflammasome activation of Caspase-1 coordinately regulates inflammatory interleukin-1-beta production and axonal degeneration-associated Caspase-6 activation. *Cell Death Differ*, *22*, 1676–1686.

King, M.-C., Marks, J. H., Mandell, J. B., & New York Breast Cancer Study Group. (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science (New York, N.Y.)*, *302*(5645), 643–646. <https://doi.org/10.1126/science.1088759>

Klau, J., Abou Jamra, R., Radtke, M., Oppermann, H., Lemke, J. R., Beblo, S., & Popp, B. (2022). Exome first approach to reduce diagnostic costs and time—Retrospective analysis of 111 individuals with rare neurodevelopmental disorders. *European Journal of Human Genetics: EJHG*, *30*(1), 117–125. <https://doi.org/10.1038/s41431-021-00981-z>

Köhler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., Baynam, G., Bello, S. M., Boerkoel, C. F., Boycott, K. M., Brudno, M., Buske, O. J., Chinnery, P. F., Cipriani, V., Connell, L. E., Dawkins, H. J. S., DeMare, L. E., Devereau, A. D., de Vries, B. B. A., ... Robinson, P. N. (2017). The Human Phenotype Ontology in 2017. *Nucleic Acids Research*, *45*(D1), D865–D876. <https://doi.org/10.1093/nar/gkw1039>

Kondratova, M., Czerwinska, U., Sompairac, N., Amigorena, S. D., Soumelis, V., Barillot, E., Zinovyev, A., & Kuperstein, I. (2019). A multiscale signalling network map

of innate immune response in cancer reveals cell heterogeneity signatures. *Nature Communications*, 10(1), 4808. <https://doi.org/10.1038/s41467-019-12270-x>

Kondratova, M., Sompairac, N., Barillot, E., Zinovyev, A., & Kuperstein, I. (2018). Signalling maps in cancer research: Construction and data analysis. *Database*.

Konecny, G. E., Wang, C., Hamidi, H., Winterhoff, B., Kalli, K. R., Dering, J., Ginther, C., Chen, H.-W., Dowdy, S., & Cliby, W. (2014). Prognostic and Therapeutic Relevance of Molecular Subtypes in High-Grade Serous Ovarian Cancer. *JNCI J. Natl. Cancer Inst*, 106.

Kuperstein, I., Bonnet, E., Nguyen, H.-A., Cohen, D., Viara, E., Grieco, L., Fourquet, S., Calzone, L., Russo, C., & Kondratova, M. (2015). Atlas of Cancer Signalling Network: A systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis*, 4, 160.

Kuperstein, I., Cohen, D. P., Pook, S., Viara, E., Calzone, L., Barillot, E., & Zinovyev, A. (2013). NaviCell: A web-based environment for navigation, curation and maintenance of large molecular interaction maps. *BMC Syst. Biol*, 7, 100.

Lalouel, J.-M., & Rohrwasser, A. (2007). Genetic Susceptibility to Essential Hypertension. *Hypertension*, 49(3), 597–603. <https://doi.org/10.1161/01.HYP.0000257145.20363.9c>

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., ... International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. <https://doi.org/10.1038/35057062>

Latunde-Dada, G. O. (2017). Ferroptosis: Role of lipid peroxidation, iron and ferritinophagy. *Biochim. Biophys. Acta Gen. Subj*, 1861, 1893–1900.

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., & Wimalaratne, S. M. (2009). The Systems Biology Graphical Notation. *Nat. Biotechnol*, 27, 735–741.

Lee, A., Mavaddat, N., Wilcox, A. N., Cunningham, A. P., Carver, T., Hartley, S., Villiers, C. B. de, Izquierdo, A., Simard, J., Schmidt, M. K., Walter, F. M., Chatterjee, N., Garcia-Closas, M., Tischkowitz, M., Pharoah, P., Easton, D. F., & Antoniou, A. C. (2019). BOADICEA: A comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genetics in Medicine*, 21(8), 1708–1718. <https://doi.org/10.1038/s41436-018-0406-9>

Li, G., Liu, Y., Jing, X., Wang, Y., Miao, M., Tao, L., Zhou, Z., Xie, Y., Huang, Y., Lei, J., Gong, G., Jin, P., Hao, Y., Faria, N. R., Clercq, E. D., & Zhang, M. (2020). Mortality risk of COVID-19 in elderly males with comorbidities: A multi-country study. *Aging*, 13(1), 27–60. <https://doi.org/10.18632/aging.202456>

Lian, Q., Zhang, K., Zhang, Z., Duan, F., Guo, L., Luo, W., Mok, B. W.-Y., Thakur, A., Ke, X., Motalebnejad, P., Nicolaescu, V., Chen, J., Ma, C. Y., Zhou, X., Han, S., Han, T., Zhang, W., Tan, A. Y., Zhang, T., ... Chen, H. J. (2022). Differential effects of

macrophage subtypes on SARS-CoV-2 infection in a human pluripotent stem cell-derived model. *Nature Communications*, 13(1), Article 1. <https://doi.org/10.1038/s41467-022-29731-5>

Liang, W. S., Reiman, E. M., Valla, J., Dunckley, T., Beach, T. G., Grover, A., Niedzielko, T. L., Schneider, L. E., Mastroeni, D., & Caselli, R. (2008). Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc. Natl. Acad. Sci. USA*, 105, 4441–4446.

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., Liu, L., Amit, I., Zhang, S., & Zhang, Z. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nature Medicine*, 26(6), Article 6. <https://doi.org/10.1038/s41591-020-0901-9>

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, 1(6), 417–425. <https://doi.org/10.1016/j.cels.2015.12.004>

Lim, W., Olschwang, S., Keller, J. J., Westerman, A. M., Menko, F. H., Boardman, L. A., Scott, R. J., Trimbath, J., Giardiello, F. M., Gruber, S. B., Gille, J. J. P., Offerhaus, G. J. A., de Rooij, F. W. M., Wilson, J. H. P., Spigelman, A. D., Phillips, R. K. S., & Houlston, R. S. (2004). Relative frequency and morphology of cancers in STK11 mutation carriers. *Gastroenterology*, 126(7), 1788–1794. <https://doi.org/10.1053/j.gastro.2004.03.014>

Lister, R., Pelizzola, M., Downen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), Article 7271. <https://doi.org/10.1038/nature08514>

Logue, S. E., Cleary, P., Saveljeva, S., & Samali, A. (2013). New directions in ER stress-induced cell death. *Apoptosis*, 18, 537–546.

Loveday, C., Turnbull, C., Ruark, E., Xicola, R. M. M., Ramsay, E., Hughes, D., Warren-Perry, M., Snape, K., Breast Cancer Susceptibility Collaboration (UK), Eccles, D., Evans, D. G., Gore, M., Renwick, A., Seal, S., Antoniou, A. C., & Rahman, N. (2012). Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nature Genetics*, 44(5), 475–476; author reply 476. <https://doi.org/10.1038/ng.2224>

Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: A tutorial. *Molecular Systems Biology*, 15(6), e8746. <https://doi.org/10.15252/msb.20188746>

MacDonald, L., Alivernini, S., Toluoso, B., Elmesmari, A., Somma, D., Perniola, S., Paglionico, A., Petricca, L., Bosello, S. L., Carfì, A., Sali, M., Stigliano, E., Cingolani, A., Murri, R., Arena, V., Fantoni, M., Antonelli, M., Landi, F., Franceschi, F., ... Gremese, E. (2021). COVID-19 and RA share an SPP1 myeloid pathway that drives PD-L1+ neutrophils and CD14+ monocytes. *JCI Insight*, 6(13), e147413. <https://doi.org/10.1172/jci.insight.147413>

Mack, S. C., Witt, H., Piro, R. M., Gu, L., Zuyderduyn, S., Stütz, A. M., Wang, X., Gallo, M., Garzia, L., Zayne, K., Zhang, X., Ramaswamy, V., Jäger, N., Jones, D. T. W., Sill, M., Pugh, T. J., Ryzhova, M., Wani, K. M., Shih, D. J. H., ... Taylor, M. D. (2014). Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature*, *506*(7489), 445–450. <https://doi.org/10.1038/nature13108>

Magtanong, L., Ko, P. J., & Dixon, S. J. (2016). Emerging roles for lipids in non-apoptotic cell death. *Cell Death Differ*, *23*, 1099–1109.

Mamik, M. K., & Power, C. (2017). Inflammasomes in neurological diseases: Emerging pathogenic and therapeutic concepts. *Brain*, *140*, 2273–2285.

Marshall, C. R., Bick, D., Belmont, J. W., Taylor, S. L., Ashley, E., Dimmock, D., Jobanputra, V., Kearney, H. M., Kulkarni, S., Rehm, H., & on behalf of the Medical Genome Initiative. (2020). The Medical Genome Initiative: Moving whole-genome sequencing for rare disease diagnosis to the clinic. *Genome Medicine*, *12*(1), 48. <https://doi.org/10.1186/s13073-020-00748-z>

Martignetti, L., Calzone, L., Bonnet, E., Barillot, E., & Zinovyev, A. (2016). ROMA: Representation and quantification of module activity from target expression data. *Front Genet*, *7*, 1–12.

Mazein, A., Ivanova, O., Balaur, I., Ostaszewski, M., Berzhitskaya, V., Serebriyskaya, T., Ligon, T., Hasenauer, J., De Meulder, B., Overall, R. W., Roy, L., Knowles, R. G., Wheelock, C. E., Dahlen, S.-E., Chung, K. F., Adcock, I. M., Roberts, G., Djukanovic, R., Pellet, J., ... eTRIKS Consortium. (2021). AsthmaMap: An interactive knowledge repository for mechanisms of asthma. *The Journal of Allergy and Clinical Immunology*, *147*(3), 853–856. <https://doi.org/10.1016/j.jaci.2020.11.032>

McCluggage, W. G. (2011). Morphological subtypes of ovarian carcinoma: A review with emphasis on new developments and pathogenesis. *Pathology*, *43*, 420–432.

Meidaninikjeh, S., Sabouni, N., Marzouni, H. Z., Bengar, S., Khalili, A., & Jafari, R. (2021). Monocytes and macrophages in COVID-19: Friends and foes. *Life Sciences*, *269*, 119010. <https://doi.org/10.1016/j.lfs.2020.119010>

Melson, M. (2020). Droplets and Aerosols in the Transmission of SARS-CoV-2. *The New England Journal of Medicine*, *382*(21), 2063. <https://doi.org/10.1056/NEJMc2009324>

Monraz Gomez, L. C., Kondratova, M., Ravel, J.-M., Barillot, E., Zinovyev, A., & Kuperstein, I. (2019). Application of Atlas of Cancer Signalling Network in preclinical studies. *Briefings in Bioinformatics*, *20*(2), 701–716. <https://doi.org/10.1093/bib/bby031>

Musicco, M., Adorni, F., Santo, S., Prinelli, F., Pettenati, C., Caltagirone, C., Palmer, K., & Russo, A. (2013). Inverse occurrence of cancer and Alzheimer disease: A population-based incidence study. *Neurology*, *81*, 322–328.

Najm, M., Cornet, M., Albergante, L., Zinovyev, A., Sermet-Gaudelus, I., Stoven, V., Calzone, L., & Martignetti, L. (2023). *Representation and quantification Of Module Activity from omics data with rROMA* (p. 2022.10.24.513448). bioRxiv. <https://doi.org/10.1101/2022.10.24.513448>

Nassir, N., Tambi, R., Bankapur, A., Al Heialy, S., Karuvantevida, N., Khansaheb, H. H., Zehra, B., Begum, G., Hameid, R. A., Ahmed, A., Deesi, Z., Alkhajeh, A., Uddin, K. M. F., Akter, H., Safizadeh Shabestari, S. A., Almidani, O., Islam, A., Gaudet, M., Kandasamy, R. K., ... Uddin, M. (2021). Single-cell transcriptome identifies FCGR3B upregulated subtype of alveolar macrophages in patients with critical COVID-19. *IScience*, *24*(9), 103030. <https://doi.org/10.1016/j.isci.2021.103030>

NEJM Catalyst. (2017). What Is Patient-Centered Care? *Catalyst Carryover*, *3*(1). <https://doi.org/10.1056/CAT.17.0559>

Nguyen, T.-M., Shafi, A., Nguyen, T., & Draghici, S. (2019). Identifying significantly impacted pathways: A comprehensive review and assessment. *Genome Biology*, *20*(1), 203. <https://doi.org/10.1186/s13059-019-1790-4>

Nicholls, R. D. (2000). The impact of genomic imprinting for neurobehavioral and developmental disorders. *The Journal of Clinical Investigation*, *105*(4), 413–418. <https://doi.org/10.1172/JCI9460>

O'Driscoll, M., Ribeiro Dos Santos, G., Wang, L., Cummings, D. A. T., Azman, A. S., Paireau, J., Fontanet, A., Cauchemez, S., & Salje, H. (2021). Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*, *590*(7844), 140–145. <https://doi.org/10.1038/s41586-020-2918-0>

Ojala, J., Alafuzoff, I., Herukka, S. K., Groen, T., Tanila, H., & Pirttilä, T. (2009). Expression of interleukin-18 is increased in the brains of Alzheimer's disease patients. *Neurobiol. Aging*, *30*, 198–209.

Ollila, H. M., Ravel, J.-M., Han, F., Faraco, J., Lin, L., Zheng, X., Plazzi, G., Dauvilliers, Y., Pizza, F., Hong, S.-C., Jennum, P., Knudsen, S., Kornum, B. R., Dong, X. S., Yan, H., Hong, H., Coquillard, C., Mahlios, J., Jolanki, O., ... Mignot, E. (2015). HLA-DPB1 and HLA class I confer risk of and protection from narcolepsy. *American Journal of Human Genetics*, *96*(1), 136–146. <https://doi.org/10.1016/j.ajhg.2014.12.010>

Ostaszewski, M., Gebel, S., Kuperstein, I., Mazein, A., Zinovyev, A., Dogrusoz, U., Hasenauer, J., Fleming, R. M. T., Le Novère, N., Gawron, P., Ligon, T., Niarakis, A., Nickerson, D., Weindl, D., Balling, R., Barillot, E., Auffray, C., & Schneider, R. (2019). Community-driven roadmap for integrated disease maps. *Briefings in Bioinformatics*, *20*(2), 659–670. <https://doi.org/10.1093/bib/bby024>

Ostaszewski, M., Mazein, A., Gillespie, M. E., Kuperstein, I., Niarakis, A., Hermjakob, H., Pico, A. R., Willighagen, E. L., Evelo, C. T., Hasenauer, J., Schreiber, F., Dräger, A., Demir, E., Wolkenhauer, O., Furlong, L. I., Barillot, E., Dopazo, J., Orta-Resendiz, A., Messina, F., ... Schneider, R. (2020). COVID-19 Disease Map, building a computational repository of SARS-CoV-2 virus-host interaction mechanisms. *Scientific Data*, *7*(1), Article 1. <https://doi.org/10.1038/s41597-020-0477-8>

Ostaszewski, M., Niarakis, A., Mazein, A., Kuperstein, I., Phair, R., Orta-Resendiz, A., Singh, V., Aghamiri, S. S., Acencio, M. L., Glaab, E., Ruepp, A., Fobo, G., Montrone, C., Brauner, B., Frishman, G., Monraz Gómez, L. C., Somers, J., Hoch, M., Kumar Gupta, S., ... the, C.-19 D. M. C. (2021). COVID19 Disease Map, a computational knowledge

repository of virus–host interaction mechanisms. *Molecular Systems Biology*, 17(10), e10387. <https://doi.org/10.15252/msb.202110387>

Ou, S. M., Lee, Y. J., Hu, Y. W., Liu, C. J., Chen, T. J., Fuh, J. L., & Wang, S. J. (2012). Does Alzheimer's disease protect against cancers? A nationwide population-based study. *Neuroepidemiology*, 40, 42–49.

Ouyang, L., Shi, Z., Zhao, S., Wang, F. T., Zhou, T. T., Liu, B., & Bao, J. K. (2012). Programmed cell death pathways in cancer: A review of apoptosis, autophagy and programmed necrosis. *Cell Prolif*, 45, 487–498.

Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., Meyer, J. G., Quan, Q., Muehlbauer, L. K., Trujillo, E. A., He, Y., Chopra, A., Chieng, H. C., Tiwari, A., Judson, M. A., Paulson, B., Brademan, D. R., Zhu, Y., Serrano, L. R., ... Jaitovich, A. (2021). Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Systems*, 12(1), 23-40.e7. <https://doi.org/10.1016/j.cels.2020.10.003>

Paloschi, M. V., Lopes, J. A., Boeno, C. N., Silva, M. D. S., Evangelista, J. R., Pontes, A. S., da Silva Setúbal, S., Rego, C. M. A., Néry, N. M., Ferreira E Ferreira, A. A., Pires, W. L., Filipin, K. P., Ferreira, G. E. M., Bozza, P. T., & Zuliani, J. P. (2020). Cytosolic phospholipase A2- $\alpha$  participates in lipid body formation and PGE2 release in human neutrophils stimulated with an L-amino acid oxidase from *Calloselasma rhodostoma* venom. *Scientific Reports*, 10(1), 10976. <https://doi.org/10.1038/s41598-020-67345-3>

Parton, A., McGilligan, V., Chemaly, M., O'Kane, M., & Watterson, S. (2019). New models of atherosclerosis and multi-drug therapeutic interventions. *Bioinformatics (Oxford, England)*, 35(14), 2449–2457. <https://doi.org/10.1093/bioinformatics/bty980>

Patel, M. N., Carroll, R. G., Galván-Peña, S., Mills, E. L., Olden, R., Triantafilou, M., Wolf, A. I., Bryant, C. E., Triantafilou, K., & Masters, S. L. (2017). Inflammasome Priming in Sterile Inflammatory Disease. *Trends Mol. Med*, 23, 165–180.

Pentimalli, F., Grelli, S., Daniele, N., Melino, G., & Amelio, I. (2018). Cell death pathologies: Targeting death pathways and the immune system for cancer therapy. *Genes Immun*.

Perfetto, L., Briganti, L., Calderone, A., Perpetuini, A. C., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., & Pavlidou, T. (2016). SIGNOR: A database of causal relationships between biological entities. *Nucleic Acids Res*, 44, 548.

Pharoah, P. D., Guilford, P., Caldas, C., & International Gastric Cancer Linkage Consortium. (2001). Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology*, 121(6), 1348–1353. <https://doi.org/10.1053/gast.2001.29611>

Pompl, P. N., Yemul, S., Xiang, Z., Ho, L., Haroutunian, V., Purohit, D., Mohs, R., & Pasinetti, G. M. (2003). Caspase gene expression in the brain as a function of the clinical progression of alzheimer disease. *Arch Neurol*, 60, 369–376.

Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *American Journal of Human Genetics*, 69(1), 124–137. <https://doi.org/10.1086/321272>

Qi, H., Zhang, H., Zhao, Y., Chen, C., Long, J. J., Chung, W. K., Guan, Y., & Shen, Y. (2021). MVP predicts the pathogenicity of missense variants by deep learning. *Nature Communications*, *12*(1), Article 1. <https://doi.org/10.1038/s41467-020-20847-0>

Quinodoz, M., Peter, V. G., Bedoni, N., Royer Bertrand, B., Cisarova, K., Salmaninejad, A., Sepahi, N., Rodrigues, R., Piran, M., Mojarrad, M., Pashar, A., Ghanbari Asad, A., Sousa, A. B., Coutinho Santos, L., Superti-Furga, A., & Rivolta, C. (2021). AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. *Nature Communications*, *12*(1), 518. <https://doi.org/10.1038/s41467-020-20584-4>

Raskin, J., Cummings, J., Hardy, J., Schuh, K., & Dean, R. A. (2015). Neurobiology of Alzheimer's Disease: Integrated Molecular, Physiological, Anatomical, Biomarker, and Cognitive Dimensions. *Curr. Alzheimer Res*, *12*, 712–722.

Ravel, J.-M., Benkirane, M., Calmels, N., Marelli, C., Ory-Magne, F., Ewencyk, C., Halleb, Y., Tison, F., Lecocq, C., Pische, G., Casenave, P., Chaussonot, A., Frismand, S., Tyvaert, L., Larrieu, L., Pointaux, M., Drouot, N., Bossenmeyer-Pourié, C., Oussalah, A., ... Renaud, M. (2021). Expanding the clinical spectrum of STIP1 homology and U-box containing protein 1-associated ataxia. *Journal of Neurology*, *268*(5), 1927–1937. <https://doi.org/10.1007/s00415-020-10348-x>

Ravel, J.-M., Dreumont, N., Mosca, P., Smith, D. E. C., Mendes, M. I., Wiedemann, A., Coelho, D., Schmitt, E., Rivière, J.-B., Tran Mau-Them, F., Thevenon, J., Kuentz, P., Polivka, M., Fuchs, S. A., Kok, G., Thauvin-Robinet, C., Guéant, J.-L., Salomons, G. S., Faivre, L., & Feillet, F. (2021). A bi-allelic loss-of-function SARS1 variant in children with neurodevelopmental delay, deafness, cardiomyopathy, and decompensation during fever. *Human Mutation*, *42*(12), 1576–1583. <https://doi.org/10.1002/humu.24285>

Ravel, J.-M., Maud, M., Solène, F., Céline, B., & Mathilde, R. (2022). GCH1 heterozygous pathogenic variation in a woman presenting with severe spastic tetraplegia. *Mov Disord Clin Pract (Soumis)*.

Ravel, J.-M., Michaud, M., Frismand, S., Puisieux, S., Banneau, G., Benoist, J.-F., Lambert, L., Bonnet, C., & Renaud, M. (2023). Heterozygous pathogenic variation in GCH1 associated with treatable severe spastic tetraplegia. *Parkinsonism & Related Disorders*, *109*, 105310. <https://doi.org/10.1016/j.parkreldis.2023.105310>

Ravel, J.-M., & Mignot, E. J. M. (2019). [Narcolepsy: From the discovery of a wake promoting peptide to autoimmune T cell biology and molecular mimicry with flu epitopes]. *Biologie Aujourd'hui*, *213*(3–4), 87–108. <https://doi.org/10.1051/jbio/2019026>

Ravel, J.-M., Monraz Gomez, L. C., Sompairac, N., Calzone, L., Zhivotovsky, B., Kroemer, G., Barillot, E., Zinovyev, A., & Kuperstein, I. (2020). Comprehensive Map of the Regulated Cell Death Signaling Network: A Powerful Analytical Tool for Studying Diseases. *Cancers*, *12*(4), Article 4. <https://doi.org/10.3390/cancers12040990>

Relton, C. L., & Smith, G. D. (2010). Epigenetic Epidemiology of Common Complex Disease: Prospects for Prediction, Prevention, and Treatment. *PLOS Medicine*, *7*(10), e1000356. <https://doi.org/10.1371/journal.pmed.1000356>

Reusch, N., De Domenico, E., Bonaguro, L., Schulte-Schrepping, J., Baßler, K., Schultze, J. L., & Aschenbrenner, A. C. (2021). Neutrophils in COVID-19. *Frontiers in Immunology*, *12*. <https://www.frontiersin.org/articles/10.3389/fimmu.2021.652470>

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>

Rosales, C. (2018). Neutrophil: A Cell with Many Roles in Inflammation or Several Cell Types? *Frontiers in Physiology*, *9*, 113. <https://doi.org/10.3389/fphys.2018.00113>

Salaroglio, I. C., Panada, E., Moiso, E., Buondonno, I., Provero, P., Rubinstein, M., Kopecka, J., & Riganti, C. (2017). PERK induces resistance to cell death elicited by endoplasmic reticulum stress and chemotherapy. *Mol. Cancer*, *16*, 1–13.

Sánchez-Valle, J., Tejero, H., Ibáñez, K., Portero, J. L., Krallinger, M., Al-Shahrour, F., Tabarés-Seisdedos, R., Baudot, A., & Valencia, A. (2017). A molecular hypothesis to explain direct and inverse co-morbidities between Alzheimer's Disease, Glioblastoma and Lung cancer. *Scientific Reports*, *7*(1), Article 1. <https://doi.org/10.1038/s41598-017-04400-6>

Sanders, J. M., Monogue, M. L., Jodlowski, T. Z., & Cutrell, J. B. (2020). Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, *323*(18), 1824–1836. <https://doi.org/10.1001/jama.2020.6019>

Saresella, M., Rosa, F., Piancone, F., Zoppis, M., Marventano, I., Calabrese, E., Rainone, V., Nemni, R., Mancuso, R., & Clerici, M. (2016). The NLRP3 and NLRP1 inflammasomes are activated in Alzheimer's disease. *Mol. Neurodegener*, *11*, 1–14.

Schmid-Burgk, J. L., Chauhan, D., Schmidt, T., Ebert, T. S., Reinhardt, J., Endl, E., & Hornung, V. (2016). A genome-wide CRISPR (clustered regularly interspaced short palindromic repeats) screen identifies NEK7 as an essential component of NLRP3 inflammasome activation. *J. Biol. Chem*, *291*, 103–109.

Shi, J., Gao, W., & Shao, F. (2017). Pyroptosis: Gasdermin-Mediated Programmed Necrotic Cell Death. *Trends Biochem. Sci*, *42*, 245–254.

Shi, J., Ye, J., Fei, H., Jiang, S., Wu, Z., Chen, Y., Zhang, L., & Yang, X. (2018). YWHAZ promotes ovarian cancer metastasis by modulating glycolysis. *Oncol. Rep*, *41*, 1101–1112.

Silke, J., & Vucic, D. (2014). *IAP Family of Cell Death and Signalling Regulators* (1st ed., Vol. 545). Elsevier Inc.

Son, S.-H., Goo, Y.-H., Choi, M., Saha, P. K., Oka, K., Chan, L. C. B., & Paul, A. (2016). Enhanced atheroprotection and lesion remodelling by targeting the foam cell and increasing plasma cholesterol acceptors. *Cardiovascular Research*, *109*(2), 294–304. <https://doi.org/10.1093/cvr/cvv241>

Soon, W. W., Hariharan, M., & Snyder, M. P. (2013). High-throughput sequencing for biology and medicine. *Molecular Systems Biology*, *9*(1), 640. <https://doi.org/10.1038/msb.2012.61>

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., & Montine, T. J. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement*, *7*, 280–292.

Stevens, J. B., Abdallah, B. Y., Liu, G., Horne, S. D., Bremer, S. W., Ye, K. J., Huang, J. Y., Kurkinen, M., Ye, C. J., & Heng, H. H. (2013). Heterogeneity of cell death. *Cytogenet Genome Res*, *139*, 164–173.

Stoppa-Lyonnet, D., & Jeanpierre, M. (2004). BRCA1: De l'identification du gène à l'estimation des risques tumoraux. *médecine/sciences*, *20*(3), Article 3. <https://doi.org/10.1051/medsci/2004203262>

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., & Lander, E. S. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, *102*, 15545–15550.

Tabarés-Seisdedos, R., & Baudot, A. (2016). Editorial: Direct and Inverse Comorbidities between Complex Disorders. *Front Physiol*, *7*, 117.

Tang, Q., Jiang, X., Li, H., Lin, Z., Zhou, X., Luo, X., Liu, L., & Chen, G. (2011). Expression and prognostic value of WISP-1 in patients with endometrial endometrioid adenocarcinoma. *J. Obstet. Gynaecol. Res*, *37*, 606–612.

Tarca, A. L., Bhatti, G., & Romero, R. (2013). A Comparison of Gene Set Analysis Methods in Terms of Sensitivity, Prioritization and Specificity. *PLOS ONE*, *8*(11), e79217. <https://doi.org/10.1371/journal.pone.0079217>

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., & Church, G. M. (1999). Systematic determination of genetic network architecture. *Nature Genetics*, *22*(3), Article 3. <https://doi.org/10.1038/10343>

Thomas, T., Stefanoni, D., Dzieciatkowska, M., Issaian, A., Nemkov, T., Hill, R. C., Francis, R. O., Hudson, K. E., Buehler, P. W., Zimring, J. C., Hod, E. A., Hansen, K. C., Spitalnik, S. L., & D'Alessandro, A. (2020). Evidence of Structural Protein Damage and Membrane Lipid Remodeling in Red Blood Cells from COVID-19 Patients. *Journal of Proteome Research*, *19*(11), 4455–4469. <https://doi.org/10.1021/acs.jproteome.0c00606>

Tomczak, K., Czerwińska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. In *Contemp. Oncol* (Vol. 19, pp. 68–77).

Tsuchihashi, N. A., Hayashi, K., Dan, K., Goto, F., Nomura, Y., Fujioka, M., Kanzaki, S., Komune, S., & Ogawa, K. (2015). Autophagy through 4EBP1 and AMPK regulates oxidative stress-induced premature senescence in auditory cells. *Oncotarget*, *6*, 3644–3655.

Tufo, G., Jones, A. W. E., Wang, Z., Hamelin, J., Tajeddine, N., Esposti, D. D., Martel, C., Boursier, C., Gallerne, C., & Migdal, C. (2014). The protein disulfide isomerases PDIA4 and PDIA6 mediate resistance to cisplatin-induced cell death in lung adenocarcinoma. *Cell Death Differ*, *21*, 685–695.

Türei, D., Földvári-Nagy, L., Fazekas, D., Módos, D., Kubisch, J., Kadlecsek, T., Demeter, A., Lenti, K., Csermely, P., & Vellai, T. (2015). Autophagy Regulatory Network—A systems-level bioinformatics resource for studying the mechanism and regulation of autophagy. *Autophagy*, *11*, 155–165.

Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Klein, M., Ivanova, O., Ölbei, M., Gábor, A., Theis, F., Módos, D., Korcsmáros, T., & Saez-Rodriguez, J. (2021). Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Molecular Systems Biology*, *17*(3), e9923. <https://doi.org/10.15252/msb.20209923>

Turnbull, C., & Hodgson, S. (2005). Genetic predisposition to cancer. *Clinical Medicine*, *5*(5), 491–498. <https://doi.org/10.7861/clinmedicine.5-5-491>

Vaissière, T., Sawan, C., & Herceg, Z. (2008). Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutation Research/Reviews in Mutation Research*, *659*(1), 40–48. <https://doi.org/10.1016/j.mrrev.2008.02.004>

Välk, K., Vooder, T., Kolde, R., Reintam, M.-A., Petzold, C., Vilo, J., & Metspalu, A. (2010). Gene expression profiles of non-small cell lung cancer: Survival prediction and new biomarkers. *Oncology*, *79*, 283–292.

Vandenabeele, P., Galluzzi, L., Vanden Berghe, T., & Kroemer, G. (2010). Molecular mechanisms of necroptosis: An ordered cellular explosion. *Nat. Rev. Mol. Cell Biol*, *11*, 700–714.

Visscher, P. M. (2008). Sizing up human height variation. *Nature Genetics*, *40*(5), 489–490. <https://doi.org/10.1038/ng0508-489>

Wang, B., Li, J., Ye, Z., Li, Z., & Wu, X. (2014). N-myc downstream regulated gene 1 acts as a tumour suppressor in ovarian cancer. *Oncol. Rep*, *31*, 2279–2285.

Wanichthanarak, K., Cvijovic, M., Molt, A., & Petranovic, D. (2013). yApoptosis: Yeast apoptosis database. *Database (Oxford)*.

Werlein, C., Ackermann, M., Stark, H., Shah, H. R., Tzankov, A., Haslbauer, J. D., von Stillfried, S., Bülow, R. D., El-Armouche, A., Kuenzel, S., Robertus, J. L., Reichardt, M., Haverich, A., Höfer, A., Neubert, L., Plucinski, E., Braubach, P., Verleden, S., Salditt, T., ... Jonigk, D. (2023). Inflammation and vascular remodeling in COVID-19 hearts. *Angiogenesis*, *26*(2), 233–248. <https://doi.org/10.1007/s10456-022-09860-7>

Witt, A., & Vucic, D. (2017). Diverse ubiquitin linkages regulate RIP kinases-mediated inflammatory and cell death signalling. *Cell Death Differ*, *24*, 1160–1171.

Xia, Y., Shen, S., & Verma, I. M. N. F.-. (2014). B, an Active Player in Human Cancers. *Cancer Immunol. Res*, *2*, 823–830.

Xie, Y., Hou, W., Song, X., Yu, Y., Huang, J., Sun, X., Kang, R., & Tang, D. (2016). Ferroptosis: Process and function. *Cell Death Differ*.

Xu, Y., Hiyoshi, A., Smith, K. A., Piehl, F., Olsson, T., Fall, K., & Montgomery, S. (2021). Association of Infectious Mononucleosis in Childhood and Adolescence With Risk for a Subsequent Multiple Sclerosis Diagnosis Among Siblings. *JAMA Network Open*, *4*(10), e2124932. <https://doi.org/10.1001/jamanetworkopen.2021.24932>

Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., Tai, Y., Bai, C., Gao, T., Song, J., Xia, P., Dong, J., Zhao, J., & Wang, F.-S. (2020). Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *The Lancet. Respiratory Medicine*, *8*(4), 420–422. [https://doi.org/10.1016/S2213-2600\(20\)30076-X](https://doi.org/10.1016/S2213-2600(20)30076-X)

Yan, J., Risacher, S. L., Shen, L., & Saykin, A. J. (2018). Network approaches to systems biology analysis of complex disease: Integrative methods for multi-omics data. *Briefings in Bioinformatics*, *19*(6), 1370–1381. <https://doi.org/10.1093/bib/bbx066>

Yatim, N., Cullen, S., & Albert, M. L. (2017). Dying cells actively regulate adaptive immune responses. *Nat. Rev. Immunol.*, *17*, 262–275.

Yunna, C., Mengru, H., Lei, W., & Weidong, C. (2020). Macrophage M1/M2 polarization. *European Journal of Pharmacology*, *877*, 173090. <https://doi.org/10.1016/j.ejphar.2020.173090>

Zerrouk, N., Aghakhani, S., Singh, V., Augé, F., & Niarakis, A. (2022). A Mechanistic Cellular Atlas of the Rheumatic Joint. *Frontiers in Systems Biology*, *2*. <https://www.frontiersin.org/articles/10.3389/fsysb.2022.925791>

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezchnikov, A. A., Zhang, C., Xie, T., Tran, L., Dobrin, R., Fluder, E., Clurman, B., Melquist, S., Narayanan, M., Suver, C., Shah, H., Mahajan, M., Gillis, T., Mysore, J., ... Emilsson, V. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, *153*(3), 707–720. <https://doi.org/10.1016/j.cell.2013.03.030>

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M. C., Zimmerman, L. J., Shaddox, K. F., Kim, S., Davies, S. R., Wang, S., Wang, P., Kinsinger, C. R., Rivers, R. C., Rodriguez, H., Townsend, R. R., Ellis, M. J. C., Carr, S. A., ... NCI CPTAC. (2014). Proteogenomic characterization of human colon and rectal cancer. *Nature*, *513*(7518), 382–387. <https://doi.org/10.1038/nature13438>

Zhang, D., Chen, P., Zheng, C.-H., & Xia, J. (2016). Identification of ovarian cancer subtype-specific network modules and candidate drivers through an integrative genomics approach. *Oncotarget*, *7*, 4298–4309.

Zhang, Q., Bastard, P., Liu, Z., Le Pen, J., Moncada-Velez, M., Chen, J., Ogishi, M., Sabli, I. K. D., Hodeib, S., Korol, C., Rosain, J., Bilguvar, K., Ye, J., Bolze, A., Bigio, B., Yang, R., Arias, A. A., Zhou, Q., Zhang, Y., ... Casanova, J.-L. (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science*, *370*(6515), eabd4570. <https://doi.org/10.1126/science.abd4570>

Zhang, X., Tan, Y., Ling, Y., Lu, G., Liu, F., Yi, Z., Jia, X., Wu, M., Shi, B., Xu, S., Chen, J., Wang, W., Chen, B., Jiang, L., Yu, S., Lu, J., Wang, J., Xu, M., Yuan, Z., ... Lu, H. (2020). Viral and host factors related to the clinical outcome of COVID-19. *Nature*, *583*(7816), 437–440. <https://doi.org/10.1038/s41586-020-2355-0>



## Annexes

### 1. Mode d'emploi de l'outils GeneTree

#### a. Présentation générale

L'application permet de créer des « familles » et est articulée autour de trois volets : le tableau, l'arbre et l'histoire clinique. Une « famille » peut être chargée à partir d'un fichier existant ou créée et modifiée avec le tableau et l'interface graphique. Les données sont sauvegardées en cache local et sont supprimées lorsque l'onglet du navigateur est fermé.

Une première barre de navigation (Figure 46) permet de charger une « famille » existante (en sélectionnant un fichier ou en le faisant glisser, 1), d'enregistrer la « famille » en cours localement (2) ou de réinitialiser la « famille » (3).

Il existe également deux modes d'utilisation optionnels permettant de restreindre les pathologies (7) :

- Le mode « onco » est destiné aux consultations d'oncogénétique. Il permet de pré-remplir la liste des pathologies avec les cancers fréquents et de gagner ainsi du temps dans la majorité des cas.
- Le mode « HPO » limite la liste des pathologies aux symptômes et maladies des bases HPO et ORPHAdata. Cette approche permet de reporter les symptômes et maladies de façon constante. L'ontologie des phénotypes humains (HPO) fournit un vocabulaire normalisé des anomalies phénotypiques rencontrées dans les maladies humaines (Köhler et al., 2017). Chaque terme de l'HPO décrit une anomalie phénotypique.

Le tableur (8) est également utilisable comme un tableur Excel : il est possible de copier et coller des lignes. Il est possible également de sélectionner une ou plusieurs lignes et de les glisser à une autre position. Cela peut être utile lorsque l'arbre généré a des branches croisées.

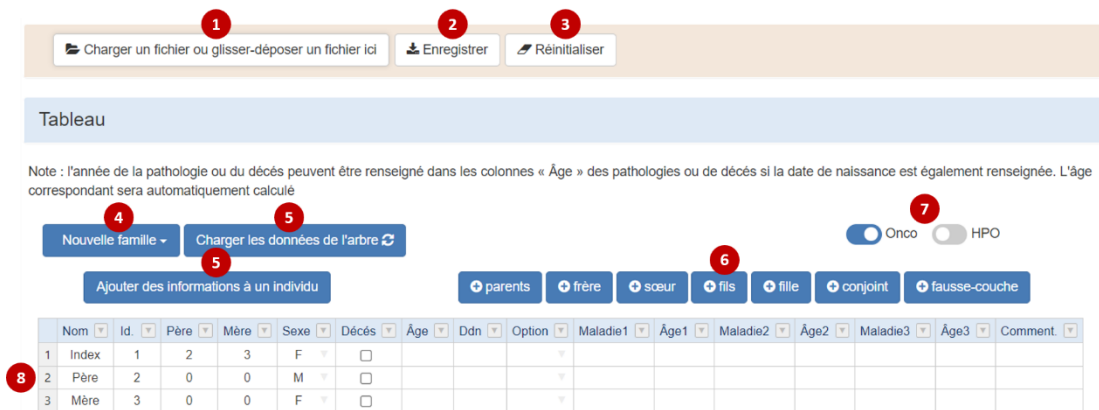


Figure 46. Barre de navigation et section « tableau » de l'interface médecin.

L'interface en ligne optimisée pour le navigateur Google Chrome permet une utilisation simple sans nécessiter d'installation préalable. Une vidéo de présentation est disponible ici : <https://www.youtube.com/watch?v=FneR3XdMjhc&feature=youtu.be>

## b. Création d'une famille

### *Chargement d'un fichier*

Une famille déjà créée peut être chargée grâce au bouton dédié (1). De nombreux formats de fichiers sont pris en charge : les fichiers créés par l'application (.json), les fichiers CanRisk et Boadicea (v2 et v4) ainsi que les fichiers .ped. Les fichiers GEDCOM sont pris en charge partiellement (import de la structure de la famille avec les noms des membres)

### *Création d'une nouvelle famille*

La structure de base d'une famille est composée de trois membres : le cas index et ses parents. Il est possible de créer une famille plus élaborée en quelques clics en cliquant sur le menu déroulant « Nouvelle Famille » (4, Figure 47. Menu déroulant 'Nouvelle Famille)

:

- Soit à partir d'une structure standard (9)
  - Le cas index et ses parents.
  - Le cas index, ses parents, et ses grands-parents.

- Le cas index, ses parents, ses grands-parents, et les arrière-grands-parents.
- Soit à partir d'une structure personnalisée (10) avec le nombre de membres adéquats (frères, soeurs, oncles et tantes de chaque branches). La famille est créée en cliquant sur le bouton « Créer un nouvel arbre » (11).

Note : il est également possible, à tout moment, de sélectionner un individu sur le tableau et de lui ajouter des membres via ce menu en cliquant sur le bouton « Ajouter à l'individu » (12).

Figure 47. Menu déroulant 'Nouvelle Famille'

### c. Modification d'une famille au niveau du tableau

#### *Ajout d'un individu*

Une fois un individu sélectionné, il est possible de cliquer sur les boutons afin d'ajouter un apparenté (6) : frère, sœur, fils, fille, parents, nouveau conjoint, fausse-couche.

#### *Personnalisation d'un individu*

Plusieurs paramètres peuvent être remplis pour chaque individu : genre, décès, âge, année de naissance, pathologie avec âge de début, etc. L'année d'une pathologie ou du décès peut être renseignée dans les colonnes « Âge » des pathologies ou de décès. Si la date de naissance est également précisée, l'âge correspondant sera automatiquement

calculé. Un champs « commentaire » est également existant et son contenu ajouté dans le texte généré.

Un menu supplémentaire existe pour remplir d'autres informations et en particulier les données nécessaires au calcul du CanRisk. Il est ouvert en cliquant sur « Ajouter des informations à un individu » (5). Par défaut, le cas index est sélectionné.

#### d. Prise en main de l'interface graphique de l'arbre généalogique

L'arbre est chargé depuis le tableau grâce au bouton dédié (13, Figure 48. Interface Arbre généalogique.). Il est possible de défaire ou refaire la dernière action, de réinitialiser l'arbre, mettre l'arbre en plein écran, et également de centrer l'arbre (14).

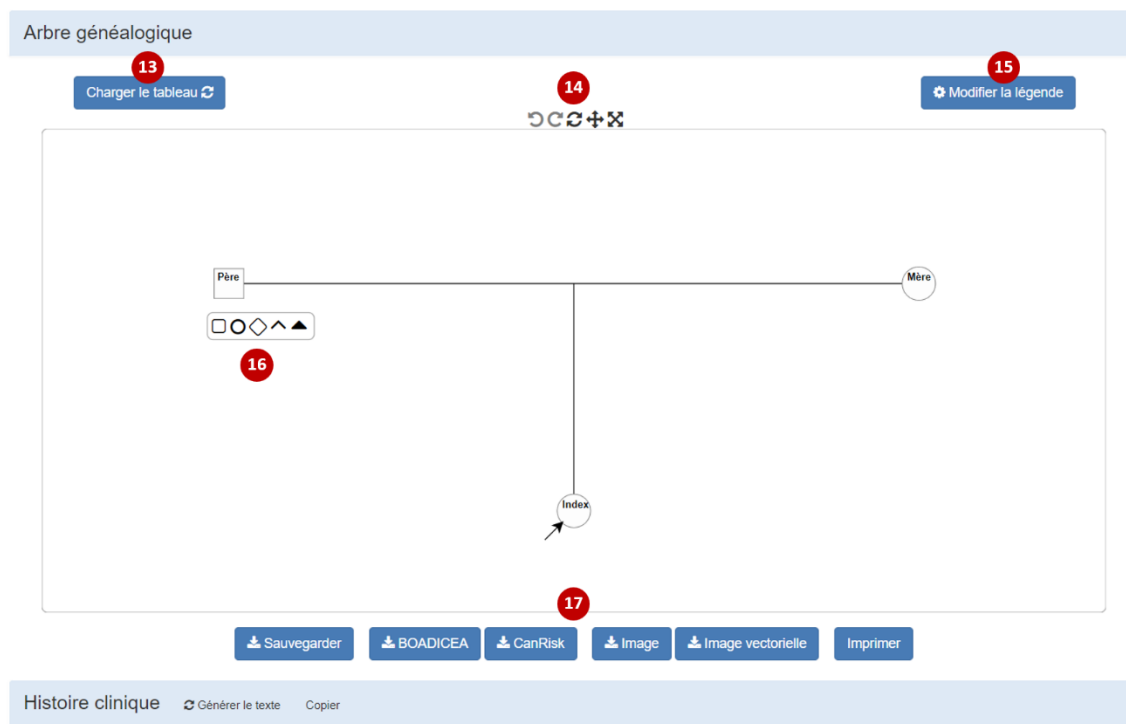


Figure 48. Interface Arbre généalogique.

Une fois que vous avez commencé à construire votre arbre généalogique, placez le pointeur de la souris sur un individu, pour faire apparaître les « widgets » utilisés pour l'édition (Figure 49. « Widgets » autour d'un individu). Vous pouvez utiliser ces « widgets » pour ajouter des partenaires, des parents, des frères et sœurs, des enfants à votre arbre. La roulette permet, elle, d'ouvrir un menu afin de modifier l'individu : âge, sexe, pathologies, etc.

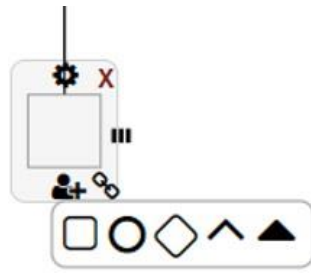


Figure 49. « Widgets » autour d'un individu

Les couleurs des pathologies sont paramétrables grâce au bouton à droite (15). Chaque pathologie correspond à une couleur (en HEX ou en toute lettre) ou à un motif. Un bouton spécifique permet de transformer les couleurs en motifs.

e. Export des différents fichiers

De nombreux fichiers sont exportables, soit au niveau du tableau, soit au niveau de l'interface graphique. Les deux interfaces ne sont pas synchronisées par défaut pour optimiser la performance de l'outil. Il suffit de cliquer sur « Charger les données de l'arbre » (5) pour charger les données de l'interface graphique au niveau du tableau et symétriquement de cliquer sur « Charger le tableau » (13) pour charger les données du tableau au niveau de l'interface graphique.

Il est possible d'exporter le tableau aux formats :

- GeneTree (.json)
- CanRisk
- Boadicea v4
- tableur simple (.tsv)
- PedigreeJS (.json)
- PED
- GEDCOM (export partiel avec la structure et les noms)

Il est possible d'exporter l'interface graphique aux formats :

- GeneTree (.json)
- CanRisk
- Boadicea v4

- image (.jpeg)
- image vectorielle (.svg)
- Imprimer (.pdf)

#### f. Génération du texte

Le texte est généré à partir du tableau (et non de l'arbre) grâce au bouton « générer le texte ». Le texte est ensuite modifiable, sélectionnable et copiable.

#### g. Interface patient

La présence d'une interface dédiée aux patients est également un avantage majeur et original de cet outil : avant la consultation, les patients peuvent, sur sollicitation de leur médecin généticien ou de leur conseillère en génétique, rentrer les informations familiales nécessaires (membres de leur famille et leurs dates de naissance respectives, éventuelles pathologies, ascendants, descendants, etc.). Le médecin généticien ou la conseillère en génétique qui effectue la consultation a accès au fichier créé et n'a plus qu'à le relire et rectifier si besoin, ce qui permet de gagner un temps précieux.

L'interface est interactive et permet au patient de décrire sa famille section par section :

- Son histoire personnelle.
- Ses enfants.
- Ses frères et sœurs.
- Ses parents.
- Sa famille paternelle.
- Sa famille maternelle.

Les données sont ensuite téléchargeables et peuvent être envoyées au médecin correspondant.

## Résumé

Faciliter l'interprétation des données générées par séquençage haut-débit et l'accès à ces technologies est un des défis majeurs actuels. La biologie des systèmes, l'étude des interactions entre les composants d'un système biologique, permet en particulier d'exploiter les données issues de séquençage haut-débit, telles que des données transcriptomiques ou épigénétique, pour étudier des processus complexes telles que le cancer ou l'immunité.

Le premier chapitre de mon travail, développé en parallèle des deux suivants, se focalise sur l'optimisation du temps d'expertise en génétique médicale par la conception d'une application : *GeneTree*. Il s'agit d'une application Web *open-source* créée sur la base de d'une problématique concrète : la redondance des saisies en pratique clinique. Elle permet de recueillir les antécédents médicaux du patient, de dessiner automatiquement des arbres généalogiques et de générer le compte rendu clinique correspondant. L'application est articulée autour de deux interfaces : un formulaire interactif pour le patient et une interface pour le praticien. Cet outil a été particulièrement pensé pour les consultations d'oncogénétique en permettant en plus l'export du score de risque CanRisk.

La deuxième partie de mon doctorat a été axée sur la construction d'une carte des voies de la mort cellulaire régulée (*regulated cell death, RCD*) incorporé dans *ACSN*, un atlas répertoriant les mécanismes impliqués dans le cancer. Une carte a été construite sur la base de données expérimentales extraites manuellement de la littérature. Elle contient les processus biologiques impliqués dans la mort cellulaire régulée sous la forme d'une carte représentant le réseau de signalisation moléculaire. En plus de permettre une vision synthétique de la mort cellulaire régulée, cette carte permet l'interprétation de données omiques. Le premier exemple d'application de cette carte est la comparaison de données transcriptomique de tissus hippocampiques de patients atteints de la maladie d'Alzheimer avec celles d'autres patients atteints de cancer du poumon non à petites cellules. L'approche a mis en évidence des mécanismes moléculaires sous-jacents à la comorbidité inverse, déjà décrite sur le plan épidémiologique, entre ces deux pathologies. Nous avons pu ensuite utiliser des données transcriptomiques de cancer de l'ovaire. Ce deuxième

exemple a permis d'obtenir des signatures moléculaires distinctes de régulation de la mort cellulaire dans quatre sous-types de cancer ovariens préalablement décrits.

La dernière partie de mon travail s'est intéressée à l'étude du système immunitaire inné et son implication dans la sévérité de la covid-19 à travers des cartes moléculaire. La réponse immunitaire innée est en effet une composante majeure de la réponse immunitaire dans la covid-19. Les cartes de réseaux moléculaires ont été construites pour chacun de ces types cellulaires et précédemment publiées par l'équipe (Kondrovata et al., 2019). Nous avons ensuite mis en place une méthode d'analyse basée sur l'algorithme ROMA pour comparer l'activité, entre des patients atteints de la covid-19 et des contrôles, des voies de signalisations canoniques, par type cellulaire, à partir de données en cellule unique (*single-cell*, *scRNA*) déjà publiées. Cette approche a permis la mise en évidence de plusieurs voies de signalisations dérégulée au niveau des macrophages, des neutrophiles et des cellules squameuses. Cette étude a aussi permis d'identifier plusieurs gènes candidats et souligne l'intérêt de l'angiogenèse dans l'inflammation liée aux macrophages. Finalement, nous avons réappliqué cette méthode en comparant les profils de patients atteint de la covid-19 à ceux atteints du virus de la grippe et ainsi montré une suractivation de voies de signalisation inflammatoire au niveau de macrophages dans la covid.

Au final, ce travail de thèse a permis proposer une solution efficace aux redondances de saisie dans le cadre des consultations de génétique, et en particulier en oncogénétique. Il souligne également l'intérêt de l'utilisation de biologie des systèmes dans les pathologies communes en permettant d'identifier des nouveaux mécanismes et des biomarqueurs de sévérité.

Mots clés : GeneTree, patient-center care, biologie des systèmes, ROMA, scRNA.

## Abstract

Facilitating the interpretation of data generated by high-throughput sequencing and access to these technologies is one of today's major challenges. Systems biology, the study of interactions between the components of a biological system, makes it possible in particular to exploit high-throughput sequencing data, such as transcriptomic or epigenetic data, to study complex processes such as cancer or immunity.

The first chapter of my work, developed in parallel with the next two, focuses on the optimization of expertise time in medical genetics through the design of an application: GeneTree. GeneTree is an open-source Web application based on a concrete problem: redundant data entry in clinical practice. It collects a patient's medical history, automatically draws family trees and generates the corresponding clinical report. The application is built around two interfaces: an interactive form for the patient and an interface for the practitioner. This tool has been specially designed for oncogenetic consultations, with the added benefit of exporting the CanRisk risk score.

The second part of my PhD focused on building a map of regulated cell death (RCD) pathways incorporated into ACSN, an atlas listing the mechanisms involved in cancer. The map was constructed on the basis of experimental data extracted manually from the literature. It contains the biological processes involved in regulated cell death in the form of a map representing the molecular signalling network. In addition to providing a synthetic view of regulated cell death, this map enables omics data to be interpreted. The first example of application of this map is the comparison of transcriptomic data from hippocampal tissues of Alzheimer's disease patients with those of other non-small cell lung cancer patients. The approach revealed molecular mechanisms underlying the inverse comorbidity, already described epidemiologically, between these two pathologies. We were then able to use transcriptomic data from ovarian cancer. This second example provided distinct molecular signatures of cell death regulation in four previously described ovarian cancer subtypes.

The final part of my work focused on studying the innate immune system and its involvement in the severity of covid-19 through molecular mapping. The innate immune response is indeed a major component of the immune response in covid-19. Molecular network maps have been constructed for each of these cell types and previously published by the team (Kondrovata et al., 2019). We then implemented an analysis method based on the ROMA algorithm to compare the activity, between covid-19 patients and controls, of canonical signaling pathways, by cell type, from previously published single-cell (scRNA) data. This approach revealed several deregulated signaling pathways in macrophages, neutrophils and squamous cells. The study also identified several candidate genes, highlighting the importance of angiogenesis in macrophage-mediated inflammation. Finally, we reapplied this method by comparing the profiles of patients with covid-19 to those with influenza virus, showing an overactivation of macrophage inflammatory signaling pathways in covid.

All in all, this thesis has provided an effective solution to the problem of redundant data entry in genetic consultations, particularly in oncogenetics. It also highlights the value of using systems biology in common pathologies, enabling the identification of new mechanisms and biomarkers of severity.

Keywords : GeneTree, patient-center care, system biology, ROMA, scRNA

## **Apport des outils bio-informatiques à l'étude de pathologies humaines par l'analyse de réseaux moléculaires et à l'amélioration des pratiques en génétique médicale.**

Faciliter l'interprétation des données générées par séquençage haut-débit et l'accès à ces technologies est un des défis majeurs actuels.

Le premier chapitre de mon travail a permis la conception d'une application : *GeneTree*. Il s'agit d'une application Web *open-source* qui permet de recueillir les antécédents médicaux du patient, de dessiner automatiquement des arbres généalogiques et de générer le compte rendu clinique et les scores de risque (CanRisk) correspondants. L'application est articulée autour de deux interfaces : un formulaire interactif pour le patient et une interface pour le praticien.

La deuxième partie est axée sur la construction d'une carte des processus biologiques impliqués de la mort cellulaire régulée (*regulated cell death, RCD*). Le premier exemple d'application est la comparaison de données transcriptomique de patients atteints de la maladie d'Alzheimer avec celles d'autres atteints de cancer du poumon non à petites cellules. L'approche a mis en évidence des mécanismes moléculaires sous-jacents à la comorbidité inverse déjà décrite sur le plan épidémiologique. Nous avons ensuite pu utiliser des données transcriptomiques de cancer de l'ovaire. Cet exemple a permis d'obtenir des signatures moléculaires dans quatre sous-types de cancer ovariens.

La dernière partie de mon travail s'est intéressée à l'étude du système immunitaire inné et son implication dans la sévérité de la covid-19 à travers des cartes moléculaires. Nous avons mis en place une méthode d'analyse basée sur l'algorithme ROMA pour comparer l'activité de voies moléculaires à partir de données en cellule unique. Cette approche a permis la mise en évidence de plusieurs voies de signalisations dérégulée au niveau des macrophages, des neutrophiles et des cellules squameuses. Plusieurs gènes candidats ont été d'identifié impliquant, en particulier, l'angiogenèse dans l'inflammation liée aux macrophages. Finalement, nous avons réappliqué cette méthode en comparant les profils de patients atteint de la covid-19 à ceux atteints du virus de la grippe et montré une suractivation de voies de signalisation inflammatoire au niveau de macrophages dans la covid-19.

Au total, ce travail propose une solution efficace aux redondances de saisie dans le cadre des consultations de génétique, et en particulier en oncogénétique. Il souligne également l'intérêt de l'utilisation de biologie des systèmes dans les pathologies communes en permettant d'identifier des nouveaux mécanismes et des biomarqueurs de sévérité.

**Mots clés :** GeneTree, application Web, CanRisk, biologie des systèmes, ROMA, scRNA, covid-19.

## **The contribution of bioinformatics tools to the study of human pathologies through the analysis of molecular networks and to the improvement of medical genetics practices.**

Facilitating the interpretation of and access to data generated by high-throughput sequencing is one of today's major challenges.

The first chapter of my work led to the design of an application: GeneTree. This is an open-source Web application that collects a patient's medical history, automatically draws family pedigree and generates the corresponding clinical report and risk scores (CanRisk). The application is structured around two interfaces: an interactive form for the patient and an interface for the practitioner.

The second part focuses on building a map of the biological processes involved in regulated cell death (RCD). The first application is the comparison of transcriptomic data from Alzheimer's disease patients with those from non-small cell lung cancer patients. The approach revealed molecular mechanisms underlying the inverse comorbidity already described epidemiologically. We were then able to use transcriptomic data from ovarian cancer. This example enabled us to obtain molecular signatures in four ovarian cancer subtypes.

The final part of my work focused on studying the innate immune system and its involvement in covid-19 severity through molecular mapping. We set up an analysis method based on the ROMA algorithm to compare the activity of molecular pathways from single-cell data. This approach revealed several deregulated signaling pathways in macrophages, neutrophils and squamous cells. Several candidate genes were identified involving, in particular, angiogenesis in macrophage-related inflammation. Finally, we reapplied this method by comparing the profiles of patients with covid-19 to those with influenza virus, and showed overactivation of macrophage inflammatory signaling pathways in covid-19.

This work offers an effective solution to redundant data entry in genetic consultations, particularly in oncogenetics. It also highlights the value of using systems biology in common pathologies to identify new mechanisms and biomarkers of severity.

**Key words:** GeneTree, Web application, CanRisk, systems biology, ROMA, scRNA, covid-19.