



# **Identification de nouvelles protéines et déterminants génétiques de la pancréatite aiguë**

**Mémoire**

**Jérôme Bourgault**

**Maîtrise en sciences cliniques et biomédicales - avec mémoire**  
Maître ès sciences (M. Sc.)

Québec, Canada

# **Identification de nouvelles protéines et déterminants génétiques de la pancréatite aiguë**

**Mémoire**

**Jérôme Bourgault**

Sous la direction de :

Benoit Arsenault, directeur de recherche

## Résumé

La pancréatite aiguë (PA) constitue la principale cause d'admissions en centre hospitalier liée à des troubles gastrointestinaux dans le monde, et sa prévalence est en augmentation. La PA est associée à plusieurs comorbidités dont la prévalence mondiale est également en augmentation, comme le syndrome métabolique et le diabète de type 2. L'architecture génétique de la PA demeure toutefois peu caractérisée et peu ou pas d'options thérapeutiques existent présentement pour sa prévention ou son traitement.

Notre hypothèse était qu'il est possible d'identifier de nouveaux gènes associés à la PA à l'aide d'études d'association pangénomiques et d'identifier des protéines circulantes associées à la PA à l'aide de la randomisation mendélienne à l'échelle du protéome sanguine humain, et que les résultats de ces analyses pouvaient permettre d'identifier des cibles thérapeutiques potentielles pour la PA.

Nos résultats ont non seulement confirmé des associations connues entre certains gènes et la PA, mais également mis en évidence de nouveaux gènes potentiellement impliqués dans la maladie. Nous avons également pu identifier plusieurs protéines sanguines causalement impliquées dans la pathophysiologie de la PA dont plusieurs sont des cibles thérapeutiques potentielles pour la PA. En outre, nous avons identifié des médicaments ciblant ces protéines, dont l'effet est cohérent avec celui obtenu grâce aux analyses de randomisation mendélienne et qui pourraient être réutilisés dans le cadre de la prévention ou du traitement de la PA.

# Abstract

Acute pancreatitis (AP) is the leading cause of hospital admissions related to gastrointestinal disorders worldwide, and its prevalence is increasing. AP is associated with several comorbidities whose global prevalence is also increasing, such as metabolic syndrome and type 2 diabetes. Despite this, its genetic architecture remains poorly characterized and few or no therapeutic options currently exist for its prevention or his treatment.

Our hypothesis was that it is possible to identify novel genes associated with AP using genome-wide association studies and to identify circulating proteins associated with AP using blood proteome-wide Mendelian randomization, and that the results of these analyzes could help identify potential therapeutic targets for AP.

Our results not only confirmed known associations between known genes and AP, but also highlighted new genes potentially involved in the disease. We were also able to identify several blood proteins causally involved in the pathophysiology of AP, several of which are potential therapeutic targets for AP. In addition, we identified drugs target those proteins, whose effect is consistent with that obtained through Mendelian randomization analyzes and which could be reused in the context of the prevention or treatment of AP.

# Table des matières

Résumé .....	ii
Abstract.....	iii
Table des matières .....	iv
Liste des figures.....	vi
Liste des abréviations .....	vii
Remerciements.....	ix
Avant-propos .....	x
Introduction .....	1
1. La pancréatite.....	1
1.1. Rôles du pancréas .....	1
1.2. Complications de la pancréatite .....	5
1.3. Incidence et taux de mortalité .....	6
1.4. Étiologie et pathophysiologie.....	7
1.5. Facteurs de risque .....	7
1.5.1. L'âge .....	8
1.5.2. Les antécédents familiaux.....	8
1.5.3. Les calculs biliaires .....	9
1.5.4. L'obésité.....	9
1.5.5. Le diabète de type 2.....	9
1.5.6. La prise de médicaments .....	10
1.5.7. Le tabagisme.....	10
1.5.8. La consommation d'alcool.....	10
2. L'héritabilité génétique.....	12
2.1. Généralités.....	12
2.2. Héritabilité génétique de la pancréatite aiguë .....	13
2.3. Corrélations génétiques.....	14
3. La randomisation mendélienne .....	16
3.1. Principe et utilisation de la randomisation mendélienne.....	17
3.2. Randomisation mendélienne inverse .....	19
3.3. Limites de la randomisation mendélienne .....	20
3.4. Colocalisation génétique .....	20
3.5. Locus de caractère quantitatif .....	23

3.6. Le génome thérapeutique .....	25
4. Problématique, hypothèses et objectifs .....	27
5. Résumé des méthodes.....	30
5.1. Cohortes.....	30
5.2. Études d'association pangénomique et méta-analyse .....	31
5.3. Héritabilité et corrélation génétique.....	32
5.4. Analyses de randomisation mendélienne à l'échelle du protéome.....	33
5.5. Loci de traits quantitatifs d'expression .....	35
5.6. Identification de cibles thérapeutiques .....	36
Chapitre 1. Proteome-Wide Mendelian Randomization Identifies Causal Links Between Blood Proteins and Acute Pancreatitis .....	37
1.1. Résumé .....	38
1.2. Abstract .....	39
1.3. Introduction.....	40
1.4. Methods.....	41
1.5. Results .....	46
1.6. Discussion .....	50
1.7. Conclusion.....	53
1.8. References .....	55
1.9. Figure Legend .....	59
1.10. Figures .....	61
1.11. Supplementary Figures .....	66
Discussion .....	70
Conclusion.....	76
Bibliographie.....	77

# Liste des figures

Figure 1. Anatomie du pancréas .....	2
Figure 2. Résumé des fonctions exocrines du pancréas .....	4
Figure 3. Pyramide hiérarchique des méthodes d'inférence causales utilisées en épidémiologie. ....	17
Figure 4. Représentation schématique d'une étude de randomisation mendélienne .....	18
Figure 5. Différents scénarios de la colocalisation génétique .....	22
Figure 6. Types de QTL et exemple de mécanisme d'action .....	24

## Chapitre 1

Graphical abstract 1. Overview of the study design.....	61
Figure 1. Manhattan plot of AP GWAS meta-analysis .....	61
Figure 2. Cross-trait analysis and genetic correlation of acute pancreatitis with other traits.....	62
Figure 3. PW-MR studies of acute pancreatitis.....	63
Figure 4. HyPrColoc plot of UGT1A1 variants.....	64
Figure 5. Functional annotation of the genetic architecture of AP .....	65

## Liste des abréviations

Abbréviation	Mot ou expression	
	Anglais	Français
ABCG5/G8	ATP binding cassette subfamily G member 5/8	Le membre 5/8 de la sous-famille G de la cassette de liaison à l'ATP
ADN	Desoxyribonucleic acid	Acide désoxyribonucléique
AMPK	AMP-activated protein kinase	Protéine kinase AMP-dépendante
AMY2A/2B	Alpha amylase 2A/2B	Amylase alpha 2A/2B
ANGPTL3	Angiopoietin-like protein 3	Protéine de type angiopoïétine 3
ARN	Ribonucleic acid	Acide ribonucléique
CFTR	Cystic fibrosis transmembrane conductance regulator	Modulateur régulateur transmembranaire de la fibrose kystique
CTRB2	Chymotrypsinogen B2	Chymotrypsinogène B2
CTRC	Chymotrypsin C	Chymotrypsine C
DT2	Type 2 Diabetes Mellitus	Diabète de type 2
ECR	Randomized controlled trial	Essai contrôlé randomisé
eQTL	Expression quantitative trait loci	Loci de caractère quantitatif d'expression
FDR	False discovery rate	Taux de fausses découvertes
FGF21	Fibroblast growth factor 21	Facteur de croissance des fibroblastes 21
GLUT2	Glucose transporter 2	Transporteur de glucose de classe 2
GREML	Genomic-relatedness-based restricted maximum-likelihood	Probabilité maximale restreinte basée sur la parenté génomique
GWAS	Genome-wide association study	Étude d'association pangénomique
HTG	Hypertriglyceridemia	Hypertriglycéridémie
ICD	International Classification of Diseases	Classification internationale des maladies
iCPAGdb	interactive Cross-Phenotype Analyse of GWAS database	base de données interactive de l'analyse de phénotypes croisées à partir de GWAS
IMC	Body mass index	Indice de masse corporelle
IV	Instrumental variable	Variable instrumentale
IVW	Inverse-variance weighting	Pondération par la variance inverse
kb	Kilobase	Kilobase
LD	Linkage disequilibrium	Déséquilibre de liaison
LDL	Low-density lipoprotein	Lipoprotéine de faible densité
LDSC	Linkage disequilibrium score regression	Régression du score de déséquilibre de liaison
LPL	Lipoprotein lipase	Lipoprotéine lipase
Mb	Megabase	Mégabase
MLN	Motilin	Motiline
MR	Mendelian randomization	Randomisation mendélienne
mTOR	Mechanistic target of rapamycin	Cible mécanistique de la rapamycine
NAD(P)H	Nicotinamide adenine dinucleotide phosphate	Nicotinamide adénine dinucléotide phosphate
NQO1	NAD(P)H quinone dehydrogenase 1	NAD(P)H quinone déshydrogénase 1
PA	Acute pancreatitis	Pancréatite aiguë

PC	Chronic pancreatitis	Pancréatite chronique
PCA	Principal components analysis	Analyse par composantes principales
pQTL	Protein quantitative trait loci	Loci de caractère quantitatif de protéine
PRS	Polygenic risk score	Score de risque polygénique
PRSS1/2	Serine protease 1/2	Protéase à sérine 1/2
PWMR	Protome-wide Mendelian randomization	Randomisation mendélienne à l'échelle du protéome
QTL	Quantitative trait locus	Locus de caractère quantitatif
ROS	Reactive oxygen species	Espèces réactives de l'oxygène
RSPO3	R-Spondin 3	R-Spondine 3
SCUBE1	CUB domain and EGF-like domain-containing 1	La protéine contenant un peptide signal, domaine CUB et domaine de type EGF 1
SNP	Single nucleotide polymorphism	Polymorphisme d'un seul nucléotide
SPINK1	Serine peptidase inhibitor Kazal type 1	Inhibiteur de la peptidase à sérine Kazal de type 1
TG	Triglycerides	Triglycérides
TPM	Transcript per million	Transcrits par million
TRPV6	Transient receptor potential cation channel subfamily V member 6	Membre 6 de la sous-famille V des canaux cationiques potentiels des récepteurs transitoires
TWIST2	Twist Family BHLH Transcription Factor 2	Facteur de transription BHLH de la famille Twist 2
UGT1A1	UDP glucuronosyltransferase 1A1	UDP glucuronosyltransférase 1A1
VLDL	Very-low-density lipoprotein	Lipoprotéine de très faible densité
WES	Whole exome sequencing	Séquençage de l'exome entier
WGS	Whole genome sequencing	Séquençage du génome entier

## Remerciements

Merci Benoit, tout d'abord, d'avoir accepté de m'inclure dans ton équipe en tant que stagiaire, alors que je n'en étais qu'au premier trimestre du baccalauréat en bio-informatique. Sans ta confiance et ton désir de voir s'épanouir la relève, je n'en serais pas où j'en suis aujourd'hui. Merci ensuite de m'avoir soutenu tout au long de ce projet et de m'avoir donné ta confiance dans la réalisation de cette maîtrise et de tous les projets stimulants qui sont survenus en cours de route. Grâce à toi, non seulement j'ai pu m'épanouir dans la réalisation de ces projets, mais je me sens également beaucoup mieux outillé pour affronter les épreuves à venir.

À toute l'équipe du laboratoire, passée et présente, Audrey-Anne, Raphaëlle, Jakie, William, Trish, Sylvain, Arnaud, Éloi, Émilie, Nooshin, Audrey, Annabelle, merci pour les 5 à 7 et les conversations. J'aimerais également remercier spécialement Hasanga d'avoir été présent pour moi lorsque j'en avais besoin, ainsi que Christian, pour son aide incommensurable.

Merci à Sébastien Thériault pour ses commentaires toujours pertinents.

Merci Julie de faire partie de ma vie et de m'avoir soutenu durant tous ces soirs que j'ai passés à étudier et travailler d'arrache-pied afin qu'un jour mes efforts soient récompensés.

Enfin, je tiens à remercier l'ensemble des co-auteurs et organismes subventionnaires ainsi que tous les participants inclus dans cette étude, sans qui la réalisation et la publication de ce travail n'auraient pas été possibles.

## **Avant-propos**

L'article présenté dans le chapitre 1, intitulé « Proteome-Wide Mendelian Randomization Identifies Causal Links Between Blood Proteins and Acute Pancreatitis » a été publié dans l'édition du mois de mai 2023 dans le journal « Gastroenterology », le journal de l'association américaine de gastroentérologie. Étant donnée leur grande taille, les tableaux supplémentaires inclus dans la version publiée de cet article ont été retirés de la version intégrée à ce mémoire. Je suis le premier auteur de cet article. J'ai effectué la collecte et l'analyse des données et ai participé à la conceptualisation et à la rédaction de cet article. Les coauteurs sont : Erik Abner, Hasanga D. Manikpurage, Natàlia Pujol-Gualdo, Triin Laisk, la Estonian Biobank Research Team; Émilie Gobeil, Eloi Gagnon, Arnaud Girard, Patricia L. Mitchell, Sébastien Thériault, Tõnu Esko, Patrick Mathieu et Benoit J. Arsenault.

# Introduction

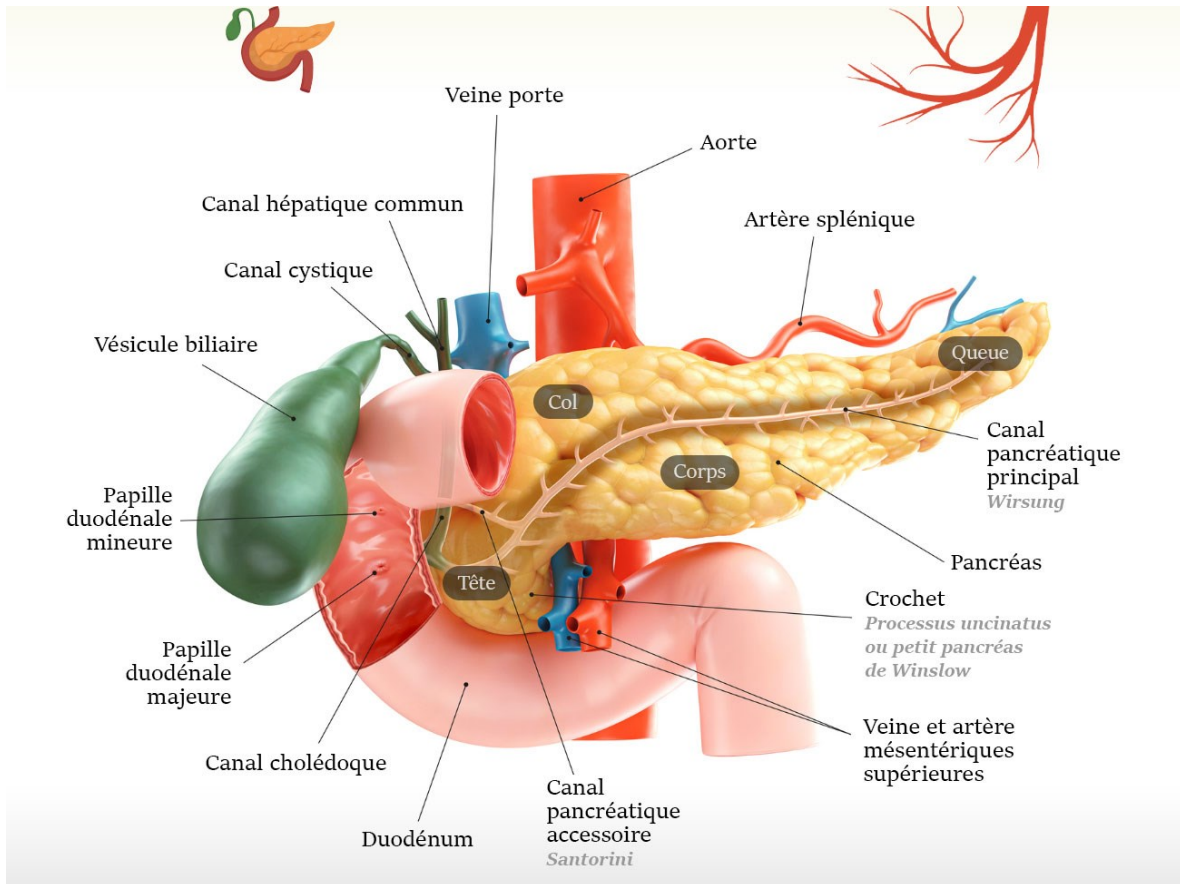
## 1. La pancréatite

### 1.1. Rôles du pancréas

Le pancréas est à la fois un organe vital et une glande du système digestif du corps humain. Il est situé dans l'abdomen, au niveau du rétropéritoine, derrière l'estomac et devant l'aorte, entre la rate et le duodénum. Le pancréas constitue, après le foie, la deuxième glande la plus volumineuse chez l'humain. À la différence du foie, le pancréas comporte deux parties anatomiquement et fonctionnellement distinctes : la partie endocrine et la partie exocrine, et peut-être divisé en trois sections distinctes : la tête, le corps et la queue. Le pancréas est parcouru dans sa longueur par le canal pancréatique principal et, dans la tête, par le canal pancréatique accessoire (présent chez environ 90% de la population). Le canal pancréatique principal débute dans la queue et rejoint, dans la partie inférieure de la tête, la partie inférieure du canal cholédoque pour former l'ampoule hépatopancréatique, laquelle s'ouvre dans la lumière du duodénum. Deux artères principales assurent l'apport sanguin au pancréas : le tronc cœliaque et l'artère mésentérique supérieure. Le drainage veineux du pancréas est assuré par la veine splénique et la veine mésentérique supérieure, laquelle se draine dans la veine porte. Le pancréas est innervé par les systèmes nerveux parasympathique, via le nerf vague, et sympathique, qui innervent ses parties endocrine et exocrine.

Le pancréas est intimement lié au foie, lequel est impliqué dans le métabolisme énergétique et la production de bile, entre autres, et des dysfonctionnements de l'un peuvent conduire à des dysfonctionnements de l'autre (1). La vésicule biliaire est un organe creux de petite taille, en forme de poire, accolé au foie sur la scissure médiane hépatique et faisant partie des voies biliaires. Son rôle principal consiste à stocker la bile, un suc digestif épais jaune verdâtre produit par le foie. La vésicule biliaire n'est pas nécessaire au bon fonctionnement de l'organisme, le foie pouvant produire de la bile en présence ou en l'absence de celle-ci. En situation postprandiale, la vésicule biliaire peut se contracter pour déverser la bile qu'elle contient dans le canal cystique, qui rejoint ensuite le canal hépatique commun pour former le canal cholédoque. La bile passe ensuite par l'ampoule hépatopancréatique pour être acheminée jusqu'au duodénum, où elle se mélange aux sucs pancréatiques et participe à l'absorption des lipides.

Étant donné le rôle majeur du pancréas dans le métabolisme énergétique, un dysfonctionnement de ses fonctions endocrines ou exocrines peut avoir de graves conséquences sur la santé de l'individu.



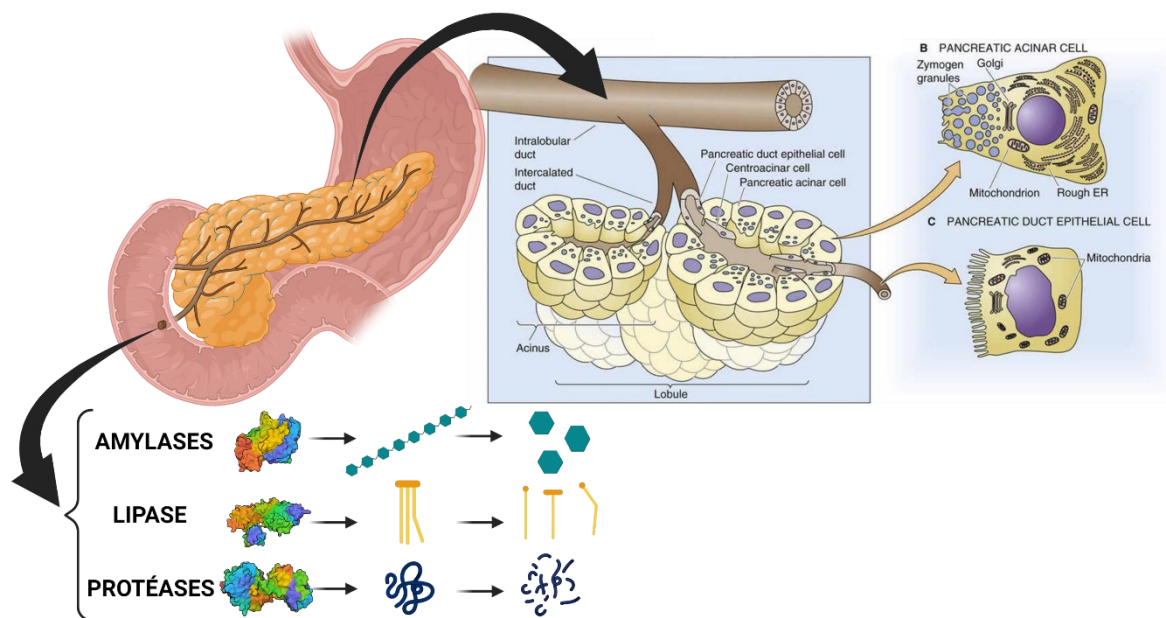
**Figure 1. Anatomie du pancréas.** Image adaptée de Index Santé - Le répertoire santé du Québec (<https://www.indexsante.ca/chroniques/308/le-pancreas.php>)

La partie endocrine du pancréas représente moins de 5% du pancréas en taille et en masse (2). Contrairement à la partie exocrine, elle produit et sécrète des hormones directement dans la circulation sanguine plutôt que via un canal. La partie endocrine est assurée par des regroupements de cellules appelés « îlots pancréatiques » ou « îlots de Langerhans ». Les îlots pancréatiques sont responsables de la production de glucagon, d'insuline, d'amyline, de somatostatine, de ghréline et de polypeptides pancréatiques et sont composés de cellules alpha, bêta, delta, epsilon et PP, ou upsilon (3,4). Les cellules alpha composent entre 25% et 35% de chaque îlot et sont responsables de la production de glucagon. Le glucagon est relâché à de faibles concentrations de glucose sanguin des cellules alpha

pancréatiques vers le foie, où il favorise la glycogénolyse, la gluconéogenèse et la lipolyse. Les cellules bêta représentent 50% à 75% de chaque îlot et sont responsables de la production d'insuline. La sécrétion d'insuline est régulée en réponse aux changements de concentration sanguine de glucose, acides aminés et acides gras libres. Lorsque la concentration de glucose sanguin augmente, celui-ci est importé dans les cellules bêta pancréatiques via le transporteur GLUT2, où il est phosphorylé par la glucokinase, ce qui augmente la ratio ATP/ADP dans la cellule. Cette augmentation d'ATP entraîne la fermeture de canaux potassiques et l'entrée de  $Ca^{2+}$  dans la cellule et la dépolarisation subséquente de la membrane cellulaire, laquelle stimule la sécrétion d'insuline par exocytose. L'insuline est ensuite relâchée des cellules bêta pancréatiques vers le foie, où elle favorise la glycogénolyse et la conversion de glucose en triglycérides (TG), les muscles, où elle favorise la glycogénolyse, et le tissu adipeux, où elle stimule la conversion de glucose en TG. Les cellules bêta pancréatiques sont également responsables de la production d'amyline, une hormone cosécétrée avec l'insuline et relâchée vers les cellules alpha pancréatiques, où elle inhibe la production de glucagon. Cette hormone a également pour effet de ralentir la vidange gastrique et l'absorption de glucose dans l'intestin grêle, de réduire la sécrétion d'enzymes digestives et de bile, ainsi que d'accroître la satiété. Les cellules delta composent environ 10% des îlots pancréatiques et sont responsables de la production de somatostatine. Le glucagon stimule la sécrétion de cette hormone qui est alors relâchée dans les cellules bêta pancréatiques, la glande pituitaire antérieure et les glandes gastriques où elle inhibe la production de plusieurs hormones, dont l'hormone de croissance, l'insuline, le glucagon, la gastrine, le peptide intestinal vasoactif et l'hormone thyroïdostimulante. Les cellules epsilon et upsilon représentent moins de 1% de chaque îlot et sont responsables de la production de ghréline et du polypeptide pancréatique, respectivement. La ghréline inhibe la sécrétion d'insuline dans les cellules bêta pancréatiques et stimule l'appétit et la production de l'hormone de croissance dans les cellules somatotrophes de la glande pituitaire antérieure. Le polypeptide pancréatique réduit la libération de glucose dans le foie et aurait pour effet d'accroître le nombre de sites de liaison à l'insuline dans le foie (5), de stimuler la sécrétion d'acide gastrique, de réduire la sécrétion de sucs pancréatiques stimulée par la cholécystokinine et pourrait stimuler, avec la motiline, la migration du complexe moteur migrant (6), un processus de contraction récurrent et cyclique qui aurait pour fonction d'éliminer les résidus de la digestion jusqu'au côlon.

La partie exocrine du pancréas représente près de 95% de la masse du pancréas (2) et joue un rôle essentiel dans le processus de digestion. Elle est responsable de la sécrétion d'enzymes digestives,

d'ions et d'eau dans le duodénum. L'eau et les ions sodium, potassium, chlorure et bicarbonate sont les principaux éléments inorganiques sécrétés par les cellules épithéliales des conduits pancréatiques et ont pour rôle d'assurer le transport des enzymes pancréatiques dans la lumière intestinale. Les sécrétions pancréatiques sont caractérisées par un pH alcalin, étant donnée la forte concentration de bicarbonate de sodium présent dans celles-ci, ce qui leur permet de neutraliser le pH acide du chyme gastrique, le liquide composé de nourriture partiellement digérée déversé de l'estomac dans le duodénum, et d'assurer une fonction optimale des enzymes pancréatiques. La partie exocrine du pancréas est également constituée de cellules acineuses, ou acini, regroupées en glandes acineuses. Ces cellules spécialisées ont pour fonction de synthétiser, stocker et sécréter les enzymes digestives et contiennent une forte densité de réticulums endoplasmiques rugueux pour assurer la synthèse protéique ainsi que des granules de zymogènes qui stockent et sécrètent les enzymes produites (Figure 2).



**Figure 2. Résumé des fonctions exocrines du pancréas.** Adapté de Boron et Boulpaep. *Medical Physiology, 3rd Edition* (2017).

Chaque glande acineuse est drainée par un petit conduit relié à un conduit plus large qui rejoint à son tour le canal pancréatique principal. Les principales enzymes digestives produites par les acini sont les trypsinogène, chymotrypsinogène, procarboxypeptidase, amylase, colipase et lipase. La sécrétion de ces enzymes dans la partie exocrine du pancréas est médiée d'une part par le nerf vagal, qui active

le transport des enzymes des cellules acineuses jusqu'aux conduits pancréatiques, et d'autre part par la libération de sécrétine de la muqueuse du duodénum à la suite du passage du chyme gastrique, laquelle stimule en retour la sécrétion de bicarbonate et d'eau. La présence de protéines et de lipides dans le duodénum stimule également la sécrétion de cholécystokinine dans le duodénum, laquelle stimule à son tour la libération des enzymes pancréatiques ainsi que la sécrétion de bile par la vésicule biliaire. Afin d'éviter l'autodigestion du pancréas, les enzymes protéolytiques sont libérées dans leur forme inactive. Après avoir atteint le duodénum, le trypsinogène est converti en trypsine, sa forme active, par l'entéropeptidase, une hydrolase sécrétée par la muqueuse duodénale. La trypsine active en retour le chymotrypsinogène en chymotrypsine et la procarboxypeptidase en carboxypeptidase. Les trypsine et chymotrypsine clivent des liens internes dans les chaînes peptidiques, alors que la carboxypeptidase clive les chaînes peptidiques à leur extrémité N-terminale ou C-terminale. La lipase pancréatique est la principale enzyme responsable de l'hydrolyse des TG. La colipase a pour rôle de prévenir l'inactivation de la lipase par les sels biliaires. L'amylase pancréatique a pour fonction d'hydrolyser les liens glycosidiques des polysaccharides et oligosaccharides comme l'amidon et le glycogène.

## 1.2. Complications de la pancréatite

La pancréatite est une inflammation du pancréas. Il existe trois grands types de pancréatites : la pancréatite chronique (PC), la pancréatite aiguë (PA) et la pancréatite héréditaire. La PC est une inflammation progressive du pancréas qui affecte les fonctions endocrine et exocrine du pancréas. Les individus dont la fonction exocrine est affectée peuvent présenter une insuffisance pancréatique, lorsqu'au moins 90% du pancréas est endommagé, une stéatorrhée ou une perte de poids. Les individus dont la fonction endocrine est affectée présentent généralement un diabète pancréatoprive (diabète de type 3c) (7). Les individus atteints de PC peuvent être asymptomatiques durant une longue période, puis ressentir une douleur persistante qui peut nécessiter une hospitalisation. Contrairement à la PC, la PA consiste en une inflammation aiguë du pancréas normalement suivie d'un retour des fonctions et morphologie normales du pancréas entre les attaques. Il existe deux manifestations de la PA : la pancréatite œdémateuse interstitielle et la pancréatite nécrosante. La pancréatite œdémateuse interstitielle est caractérisée par un gonflement du pancréas causé par un œdème inflammatoire et est la version la plus fréquemment rencontrée de PA. Dans la majorité des cas, celle-ci se résout d'elle-même (8). La pancréatite nécrosante est définie par la présence de nécrose dans le pancréas. Forme

plus rare de PA, elle est présente chez 5% à 10% des cas de PA et est associée à une évolution prolongée et plus sévère de la maladie (8). La sévérité de la PA varie d'un individu à l'autre et peut être légère, modérée ou sévère. Dans les cas légers, l'inflammation n'affecte que le pancréas et se résorbe en quelques heures ou jours, tandis que dans les cas les plus sévères, elle peut entraîner la défaillance du pancréas et d'autres organes. La récurrence des crises de PA peut entraîner le développement d'une PC. La pancréatite héréditaire est causée par une mutation autosomale dominante du gène du trypsinogène cationique (*PRSS1*) et est caractérisée par des crises récurrentes de PA qui mènent au développement d'une PC.

### 1.3. Incidence et taux de mortalité

La PA est une maladie relativement commune associée à de hauts taux de comorbidités et de mortalité. La pancréatite est la cause la plus fréquente d'admissions en centre hospitalier due à des troubles gastrointestinaux. Aux États-Unis, elle est responsable de près de 275 000 hospitalisations par année et engendre des coûts annuels estimés à 2,5 milliards de dollars (9,10). Au Canada, ces coûts sont estimés à 200 millions annuellement (11). Son incidence est en augmentation dans le monde, particulièrement, mais pas uniquement, dans les pays occidentalisés. Au Canada, le nombre de nouveaux cas de pancréatite a augmenté de 75% entre 1990 et 2016 (11), ce qui pourrait être dû à l'augmentation de la prévalence de facteurs de risque comme l'obésité et le diabète de type 2. Mondialement, la prévalence de la PA est estimée à 34 cas pour 100 000 individus (12) et son pourcentage de variation annuelle moyenne global est passé de 3,07% en 1961 à 3,84% en 2016 (13). Près de 20% des premiers épisodes de pancréatites deviendront récurrents et près de 35% des cas récurrents de PA évolueront en PC (12). L'incidence de la PA varie également en fonction du statut socio-économique et de la période de l'année (13,14). Son incidence est similaire entre les femmes et les hommes lorsque l'on tient compte de la consommation d'alcool et augmente avec l'âge (15,16). Cependant, elle est de plus en plus fréquente chez les personnes de moins de 20 ans (16). Le taux de mortalité brut global associé à la PA est de 2% et est en diminution depuis les dernières années (9,10,16,17). Cependant, le taux de mortalité brut peut atteindre 30% dans les cas les plus sévères (9). Dans les cas légers, la PA n'est pas accompagnée de complications systémiques ou de défaillance d'organes. Plus la sévérité de la PA augmente, plus le nombre d'organes défaillants augmente et plus la durée de ces défaillances augmente. Le risque de sévérité de PA augmente également avec l'augmentation de comorbidités et avec le développement de calculs biliaires (18).

#### 1.4. Étiologie et pathophysiologie

Bien qu'environ 30% des cas diagnostiqués soient idiopathiques (16,19), il existe différentes étiologies connues pour la PA. Les principaux mécanismes connus pour induire la PA sont une activation (prématurée) des enzymes digestives pancréatiques dans le pancréas ou un dommage aux cellules acineuses (20). Les causes de PA peuvent être regroupées en trois catégories. D'une part, différents troubles mécaniques peuvent causer une PA, comme les calculs biliaires, une obstruction du canal pancréatique, un traumatisme physique ou une malformation congénitale (pancréas annulaire). Les calculs biliaires sont des dépôts solides, essentiellement composés de cholestérol, mais également de bilirubine ou de calcium, qui se forment dans la vésicule biliaire lorsque ces derniers sont sécrétés en excès par le foie. Ils représentent, toutes causes confondues, la cause la plus fréquente de PA (environ 40% des cas) (9,15,21). Les calculs biliaires peuvent se loger dans l'ampoule hépatopancréatique ou le canal pancréatique principal et bloquer le flux des sucs pancréatiques, provoquant un retour de ceux-ci vers le pancréas (22). Les enzymes protéolytiques peuvent alors être activées dans le pancréas et provoquer son autodigestion, ce qui peut induire l'inflammation et entraîner la nécrose des tissus du pancréas. D'autre part, certains troubles toxiques ou métaboliques peuvent causer une PA, comme la consommation régulière d'alcool, les hyperlipidémies, certains médicaments comme les statines, les traitements hormonaux substitutifs, les diurétiques et les agents hypoglycémiques (21), et l'hypercalcémie. Toutes causes confondues, l'alcool est la seconde cause la plus fréquente de PA (environ 30% des cas) (21). L'alcool peut avoir des effets directs sur les conduits pancréatiques et cellules acineuses ainsi que sur les enzymes pancréatiques et induire l'inflammation du pancréas via le stress oxydatif ou la formation de calculs. Enfin, d'autres troubles peuvent causer une PA, comme une ischémie, une infection bactérienne, la fibrose kystique, une réaction auto-immune ou une prédisposition génétique. Certains gènes sont connus pour être impliqués dans la pathophysiologie de la PA, soit via le transport des sucs pancréatiques ou dans le sentier d'activation de la trypsine.

#### 1.5. Facteurs de risque

La PA peut être causée par des facteurs génétiques ou des facteurs environnementaux dits modifiables, ou une combinaison des deux. L'âge et les antécédents familiaux sont les principaux facteurs de risque non-modifiables. Les facteurs de risque incluent, entre autres, les dyslipidémies, les malformations congénitales, les calculs biliaires, la prise de certains médicaments et les diabètes de

type 1 et 2 (DT2). Les facteurs de risque liés aux habitudes de vie et dits modifiables incluent le tabagisme, la consommation d'alcool et le syndrome métabolique, caractérisé par un tour de taille élevé et une résistance à l'insuline.

#### 1.5.1. L'âge

Bien que la PA soit présente chez tous les groupes d'âge, le risque de souffrir de PA augmente avec l'âge (16). Plus de 50% des individus atteints de PA sont âgés de plus de 55 ans et cette proportion semble être en augmentation (23). Cette hausse de l'incidence de la PA chez les personnes plus âgées peut s'expliquer, en partie, par une plus grande consommation de médicaments, comme les antihypertenseurs et les immunosuppresseurs, et par le plus haut risque de développer des calculs biliaires ou un DT2 à un âge plus avancé (23,24).

#### 1.5.2. Les antécédents familiaux

Les antécédents familiaux et facteurs génétiques sont des facteurs de risque indépendants de la PA et impliquent certains gènes pouvant être responsables de la maladie. Les dyslipidémies sont d'importants facteurs de risque de la PA. Des dyslipidémies héréditaires, comme les hypertriglycéridémie et hyperchylomicronémie familiales sont également des facteurs de risque pour la PA (25–27). L'hypertriglycéridémie (HTG) en particulier a une prévalence d'environ 1.7% dans la population adulte (28) et est présente chez près de 10% des cas de PA (29–31). Bien que le mécanisme par lequel l'HTG cause la PA n'est toujours pas bien compris, l'augmentation de la viscosité du plasma sanguin, via l'augmentation des niveaux d'acides gras libres et de chylomicrons dans la circulation sanguine causée par l'HTG, est suspectée d'induire l'ischémie des tissus pancréatiques et l'inflammation du pancréas (29,30,32). Enfin, des variations génétiques héréditaires dans les gènes codant pour l'inhibiteur de la trypsine (*SPINK1*), le modulateur régulateur transmembranaire de la fibrose kystique (*CFTR*) et la chymotrypsine C (*CTRC*) sont également connus comme étant d'importants facteurs de risque génétiques pour la PA (33) La prévalence de la pancréatite héréditaire causée par ces mutations oscille entre 0.13 et 0.6 pour 100 000 individus, mais celles-ci pourraient représenter près de 45% des cas de PA récurrente (34).

### 1.5.3. Les calculs biliaires

Les calculs biliaires, lesquels peuvent bloquer l'ampoule hépatopancréatique, sont la cause la plus fréquente de PA et leur prévalence aux États-Unis est estimée à 7% (35). Le risque de PA causée par les calculs biliaires augmente avec l'âge et est plus élevée chez les femmes (16). Le retrait des calculs biliaires par cholécystectomie permet généralement de résoudre la PA et de prévenir sa récurrence (36).

### 1.5.4. L'obésité

Enfin, l'accumulation de graisses dans le tissu adipeux abdominal est également associée à un risque plus élevé de PA (37). Cet effet est similaire entre les individus atteints de PA causée par les calculs biliaires ou non et les individus ayant un tour de taille de plus de 105 cm présentent un risque plus de 2 fois plus élevé de développer la maladie que ceux dont cette valeur se situe entre 75-85 cm. Le risque, la sévérité ainsi que la mortalité sont également accrus chez les individus dont l'indice de masse corporelle (IMC) est plus élevé que 30 kg/m<sup>2</sup> (16,38). L'accumulation de graisses dans les tissus adipeux abdominal et viscéral entraînerait une réponse inflammatoire systémique en réponse à une production accrue d'acides gras libres via la lipolyse (16).

### 1.5.5. Le diabète de type 2

Le DT2 est une maladie complexe impliquant des facteurs génétiques et environnementaux. Il s'agit d'un désordre métabolique se traduisant par une résistance à l'insuline et une hyperglycémie. Le DT2 augmente le risque de PA de 1,5 à 3 fois après ajustement pour d'autres facteurs de risque, et ce risque est plus élevé chez les individus âgés de moins de 45 ans (16,39-41). Avec la hausse du taux mondial de DT2, des facteurs de risque qui y sont associés et du vieillissement de la population, particulièrement dans les pays ayant adopté un mode de vie occidental, il est à prévoir que l'incidence globale de PA augmentera également (16). Cependant, une récente étude de randomisation mendélienne de la PA suggère que le DT2 influencerait la PA via son effet sur la formation de calculs biliaires (28).

### 1.5.6. La prise de médicaments

La PA peut être induite par la prise de médicaments, mais dans ce cas, elle est principalement idiosyncratique. Plusieurs médicaments sont suspectés de contribuer à la PA, incluant les inhibiteurs de l'enzyme de conversion de l'angiotensine, les statines, l'hormonothérapie substitutive et les contraceptifs oraux, les diurétiques, les traitements antirétroviraux hautement actifs et les agents hypoglycémiants, comme les analogues de GLP-1 (16,21,42)

### 1.5.7. Le tabagisme

Le tabagisme est un facteur de risque de la PA et accroît ce risque d'environ 2 fois lorsque celle-ci n'est pas causée par des calculs biliaires (43–45). De plus, ce risque augmente à plus de 4 fois lorsque le tabagisme est accompagné d'une consommation régulière d'alcool (16). Les mécanismes par lesquels le tabagisme pourrait influencer la PA incluent l'infiltration de lymphocytes dans le pancréas causant son inflammation, le stress oxydatif, le déséquilibre de la production des inhibiteurs de la trypsine en faveur de la trypsine et la réduction du flux sanguin, pouvant provoquer une ischémie dans le pancréas (45). Ces effets pourraient être également accrus par la consommation d'alcool (46).

### 1.5.8. La consommation d'alcool

L'alcool peut favoriser la précipitation de protéines sécrétées dans les conduits pancréatiques, lesquelles peuvent à leur tour se calcifier pour former des calculs pouvant causer l'ulcération des cellules épithéliales de ces canaux. Cette ulcération peut ensuite entraîner l'atrophie et la fibrose des cellules acineuses. L'alcool peut également affecter directement les enzymes digestives produites par le pancréas. Les produits des métabolismes oxydatif et non-oxydatif de l'alcool, comme les espèces réactives de l'oxygène (ROS) et les esters éthyliques d'acides gras, peuvent déstabiliser les membranes des granules de zymogènes et lysosomes des cellules acineuses, augmentant les risques que les enzymes de ces deux organites entrent en contact. L'interaction entre ces enzymes au sein des cellules acineuses peut conduire à l'activation prématurée des enzymes digestives et conduire à l'autodigestion du pancréas (47–51). Un autre effet potentiel du stress oxydatif causé par l'alcool serait une hausse des niveaux intracellulaires de calcium dans les cellules acineuses, ce qui entraînerait une surcharge de calcium dans la mitochondrie et, éventuellement, la mort de la cellule (52). Cependant,

l'effet de l'alcool sur le risque de PA pourrait être médié par son effet sur la formation de calculs biliaires (28).

## 2. L'héritabilité génétique

### 2.1. Généralités

Une question centrale dans le domaine de la biologie et de l'épidémiologie génétique concerne la part d'imputabilité qui revient aux facteurs biologiques (ou génétiques) et aux facteurs environnementaux dans l'apparition d'un trait ou d'une maladie. Par exemple, la langue parlée relève de l'influence de l'environnement (socio-culturel) sur un individu, tandis que la couleur des yeux est un trait qui relève de la biologie, des facteurs génétiques de l'individu, hérités de ses parents. En ce sens, l'héritabilité est un concept qui permet de déterminer à quel point la variation observée pour un trait, ou variation phénotypique, est dépendante de la variation génétique observée. Une large part des phénotypes observés chez l'humain sont influencés par une combinaison de facteurs génétiques et environnementaux. En ce sens, déterminer l'héritabilité d'une maladie représente une étape préliminaire dans l'identification, l'interprétation et l'importance des causes génétiques d'une maladie. Dans le domaine de l'épidémiologie génétique, l'héritabilité est en quelque sorte une mesure de l'importance à accorder à la génétique dans l'étude et la compréhension d'un trait ou, en d'autres termes, de la qualité de la prédiction d'un trait à partir des variations génétiques. L'héritabilité, en tant que ratio des variances génétique et phénotypique observée, est une valeur comprise entre 0 et 1, où une valeur nulle signifie que la génétique n'est impliquée en rien dans le trait observé et une valeur de 1 signifie qu'elle permet de l'expliquer à elle seule. L'héritabilité ne permet ni d'identifier un gène causal pour un trait ni s'il s'agit d'un seul gène, mais permet plutôt de déterminer si les corrélations entre les variations génétiques sont suffisamment informatives pour expliquer les variations observées pour un phénotype. L'héritabilité n'est pas une mesure individuelle, mais populationnelle, et ne permet pas d'expliquer la cause d'un phénotype chez un individu, mais plutôt à quel point la fréquence de ce phénotype dans la population est dépendante de facteurs génétiques. De même, l'héritabilité est spécifique à la population et l'époque où le trait a été mesuré et peut donc varier selon les caractéristiques génétiques d'une population et son environnement. L'héritabilité peut varier avec l'environnement, n'est pas une mesure de la capacité d'un individu d'affecter un trait (par exemple, l'IMC est héritable, mais cela ne signifie pas que l'on ne puisse pas le modifier par l'alimentation ou l'activité physique) et ne permet donc pas de déterminer que des variations inter-populationnelles sont dues à des différences génétiques (53).

Le concept d'héritabilité regroupe deux grands types d'héritabilités : l'héritabilité au sens large et l'héritabilité au sens restreint. L'héritabilité au sens large représente le ratio des variances génétique et phénotypique totales. Elle inclut les effets additifs (les effets de tous les gènes contribuant à un phénotype) et non-additifs (la dominance, l'épistasie). L'héritabilité au sens restreint ne considère que les effets génétiques additifs (54,55). L'héritabilité de polymorphismes d'un seul nucléotide (SNP) est un cas particulier d'héritabilité au sens restreint qui détermine la variance phénotypique à partir d'un sous-échantillon de variants génétiques (56). Un SNP est une substitution germinale (héritée) d'un nucléotide (allèle) à une position précise du génome et dont la fréquence dans une population donnée est d'au moins 1% (57). L'héritabilité de SNP est dépendante du déséquilibre de liaison (LD) entre les SNPs causaux et génotypés. La LD représente l'association non aléatoire d'allèles à un loci dans une population donnée, c'est-à-dire lorsque la fréquence de ces allèles est inférieure ou supérieure à la fréquence attendue s'ils étaient transmis aléatoirement, et est de fait une exception à la loi de l'assortiment indépendant des allèles énoncée par Mendel. Le concept de LD sous-tend le concept d'haplotype, c'est-à-dire de groupes d'allèles situés sur un même chromosome et transmis ensemble. L'héritabilité de SNP est généralement plus basse que l'héritabilité au sens large, étant donnée l'utilisation de variants génétiques communs plutôt que rares. Un variant génétique est dit commun s'il présente une fréquence allélique d'au moins 1% dans une population donnée, tandis qu'un variant rare a une fréquence allélique de moins de 1%. Les variants rares ont généralement un effet plus prononcé sur un phénotype que les variants communs et, bien qu'ils puissent expliquer une part importante de l'héritabilité d'un trait (58), les analyses mettant à profit les données de variants rares ont généralement une puissance statistique réduite causée par la rareté des individus possédant ces variants (59). De ce fait, l'héritabilité de SNP représente un reflet plus fidèle de la variation phénotype populationnelle (60).

## 2.2. Héritabilité génétique de la pancréatite aiguë

Jusqu'à présent, peu de gènes ont été associés à la pancréatite aiguë. En 1997, une étude de liaison génétique menée par Gorry *et al.* a montré qu'il existait une association entre la PA et la PC et des mutations dans le gène *PRSS1* (61). En 2012, Whitcomb *et al.* ont effectué une étude d'association génétique incluant plus de 5000 individus d'ascendance européenne sur la PC et ont identifié des variants génétiques associés à la PA et à la PC aux loci des gènes *PRSS1* et *PRSS2* (62). D'autres études d'association génétique ont également découvert que des variations génétiques dans les gènes

*CFTR* et *SPINK1* étaient surreprésentées chez des patients atteints de PC (63–66). Une étude effectuée en 2010 dans une cohorte européenne a également identifié une association entre un variant au locus *SPINK1* et la PA (67). En 2022, Schmidt *et al.* ont identifié une association entre la PC et des variants génétiques aux loci des gènes *CTRC* et *SPINK1* dans une cohorte de participants européens (68). Enfin, Rosendahl *et al.* ont effectué une étude d'association génétique de PC dans une cohorte de participants d'ascendance européenne et, en plus de répliquer des associations connues avec les gènes *CTRC*, *PRSS1*, *PRSS2* et *SPINK1*, ont également identifié qu'une inversion des gènes *CTRB1* et *CTRB2* modifiait le risque de PC (69). Les mécanismes de transmission de ces variations génétiques demeurent toutefois complexes et variables selon le gène et le variant (34). Les variations dans le gène *PRSS1* sont généralement transmises selon un mécanisme autosomal dominant, bien que certains cas résultent de mutations germinales. De plus, les mutations de ce gène ont une pénétrance réduite : il est estimé que 20% des individus présentant une mutation dans ce gène ne développeront jamais de pancréatite (70). Récemment, des variations génétiques associés à la pancréatite ont été identifiées dans les gènes *ABCG8* (71), de par son implication possible dans la formation de calculs biliaires, et *TRPV6* (72,73), impliqué dans la régulation de l'homéostasie du calcium dans le pancréas (73). Cependant, comparativement à d'autres maladies complexes, seuls quelques loci génétiques ont été identifiés, ce qui suggère que la PA présenterait une héritabilité relativement faible.

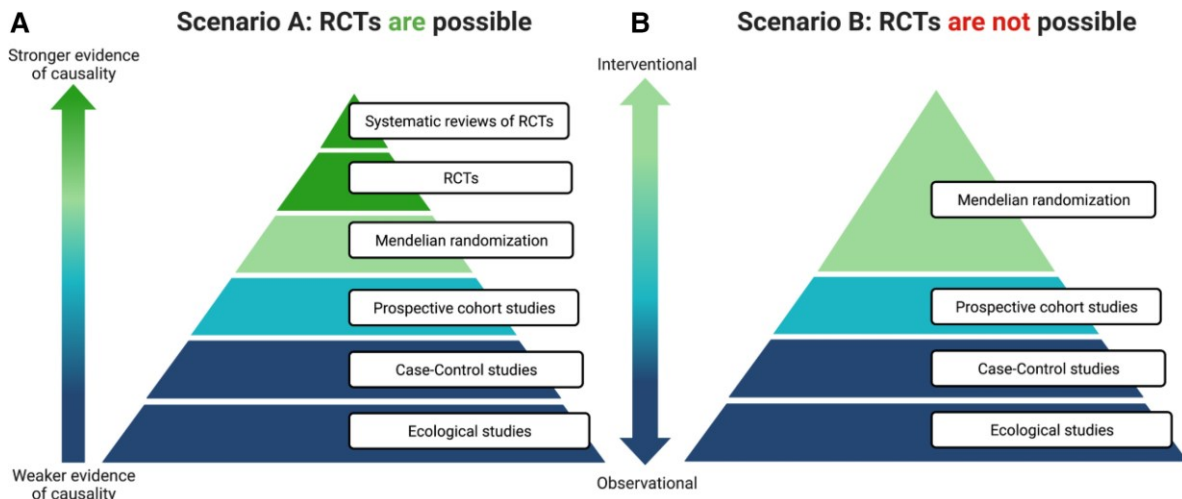
### 2.3. Corrélation génétique

En épidémiologie, la compréhension des relations parfois complexes qui existent entre certains traits et maladies est d'une grande importance pour comprendre les causes et mécanismes qui sous-tendent ceux-ci. Les premières méthodes établies pour évaluer les interactions génétiques entre certains traits étaient les études de familles, lesquelles sont limitées par le nombre de traits mesurables chez tous les individus et le nombre de ceux-ci. Au cours des dernières décennies, l'avènement des études d'association pangénomiques (GWAS) a permis d'étudier les interactions génétiques entre un large éventail de traits mesurés dans de larges cohortes, sans qu'il soit nécessaire de mesurer tous ces traits chez l'ensemble de ces individus. Une GWAS est une étude observationnelle qui vise à mieux comprendre les mécanismes biologiques d'un trait, catégorique ou continu, en identifiant des variants génétiques dont la fréquence allélique diffère significativement entre les individus en fonction du trait (74). L'une des méthodes les plus utilisées dans ce domaine est la randomisation mendélienne (MR), laquelle utilise des SNPs significativement associés à un trait mesuré afin d'établir une relation causale

entre celui-ci et un autre trait, par exemple entre un facteur de risque et une maladie. Dans le cas de traits complexes, ces méthodes sont efficaces lorsque les SNPs considérés expliquent une forte proportion de l'héritabilité, mais dans plusieurs cas, celle-ci est généralement faible, et il peut être préférable d'utiliser l'ensemble des variants du génome pour expliquer une plus grande part d'héritabilité (60,75–77). La corrélation génétique est une approche qui permet d'inclure l'effet de tous les variants du génome et d'établir la corrélation entre deux phénotypes à partir de l'effet des SNPs sur ceux-ci. Elle permet donc d'établir la similarité génétique entre ces traits et de mettre en lumière les similarités ou différences dans leur architecture génétique. Il existe deux grands types de méthodes permettant d'évaluer la corrélation génétique entre deux traits : celles qui requièrent des données génétiques individuelles sur les participants, comme la probabilité maximale restreinte basée sur la parenté génomique (GREML) et les scores de risque polygéniques (PRS), et celles qui requièrent seulement des statistiques sommaires de GWAS, comme la régression du score de LD (LDSC) (75,78). Bien que ces dernières méthodes facilitent le calcul de corrélations génétiques entre plusieurs traits de par l'accès plus facile aux statistiques sommaires de GWAS, elles sont généralement moins robustes que les méthodes qui utilisent des données génétiques individuelles de participants et leurs résultats doivent être interprétés avec prudence (78,79).

### 3. La randomisation mendélienne

En épidémiologie traditionnelle, les études observationnelles visent à évaluer l'association entre une intervention ou un facteur de risque et un trait d'intérêt dans une population sans manipuler en l'exposition chez les individus. Il existe quatre grands types d'études observationnelles, dont les études de cas-contrôle et les études longitudinales. Une étude observationnelle pourrait être menée pour observer l'association entre une exposition à de faibles concentrations sanguines de vitamine D et le risque de développer un cancer, par exemple. Cependant, ce type d'étude est sujet à d'importants biais introduits par une potentielle relation de causalité inverse, c'est-à-dire lorsque le trait d'intérêt cause le facteur de risque (par exemple, si le cancer affecte les niveaux de vitamine D, et non l'inverse) ou par la présence de facteurs confondants, qui affectent le trait d'intérêt indépendamment du facteur de risque (par exemple, le tabagisme) et qu'il est parfois difficile d'identifier dans leur entièreté. Les résultats des études observationnelles peuvent donc ne pas conduire à l'établissement d'un lien de causalité entre un facteur de risque ou une intervention et un trait d'intérêt, et sont généralement débattus dans la communauté scientifique, voire parfois contredits par les résultats d'essais cliniques randomisés (ECR) (77,80,81), lorsque ceux-ci sont possibles à réaliser. À la différence des études observationnelles, les ECR répartissent aléatoirement les participants dans un groupe contrôle, ne recevant pas l'intervention, et dans un groupe traité, recevant l'intervention. Si le nombre de participants est suffisant, la randomisation permet de réduire le biais associé aux facteurs confondants en répartissant ceux-ci de manière similaire entre les deux groupes (82). Cependant, bien que les ECR permettent généralement d'établir un lien de causalité plus robuste entre un facteur de risque et un trait d'intérêt, elles ne sont pas toujours faisables d'un point de vue temporel, financier ou éthique (83,84). Par exemple, un ECR visant à exposer les participants au tabagisme pour en déterminer les effets sur le risque de cancer ne serait évidemment pas recevable d'un point de vue éthique. La MR est comparable à un ECR, à la différence qu'elle répartit les participants en groupes selon leurs variants génétiques et, dans le cas d'une étude sur les effets du tabagisme sur le risque de cancer, utilise plutôt des variants génétiques associés à une prédisposition au tabagisme (Figure 3).

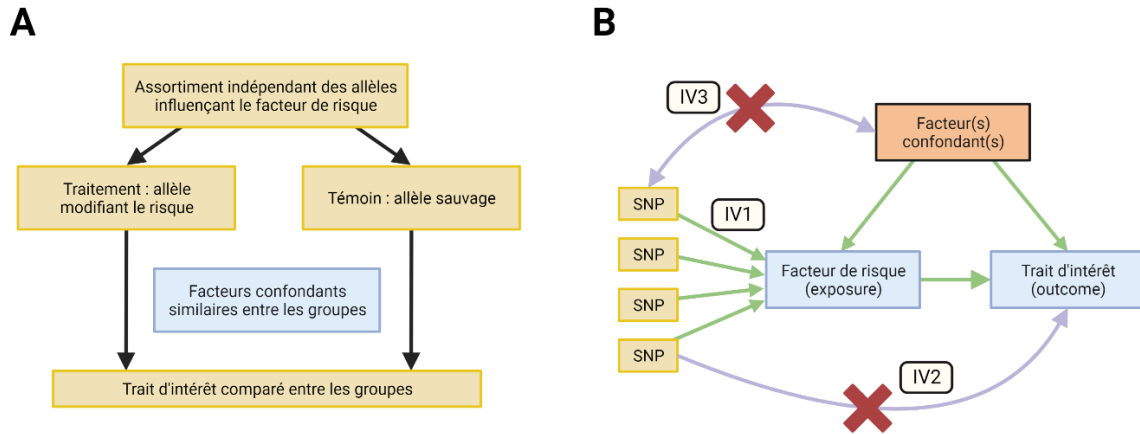


**Figure 3. Pyramide hiérarchique des méthodes d'inférence causales utilisées en épidémiologie.** (A) Lorsque les ECR peuvent être conduites et (B) lorsque les ECR ne peuvent pas être conduites. Adapté de Arsenault BJ. *European Heart Journal* (2022).

### 3.1. Principe et utilisation de la randomisation mendélienne

La MR est une méthode qui permet non seulement d'évaluer le potentiel lien de causalité entre un facteur de risque et un trait d'intérêt, mais également d'en quantifier l'effet (85). Elle s'avère particulièrement utile dans les cas où les facteurs confondants sont difficilement identifiables et les enjeux éthiques liés à l'exposition à un facteur de risque difficilement surmontables (77). La MR tire parti des lois de Mendel concernant la ségrégation et l'assortiment indépendant des allèles. Étant donnée la ségrégation aléatoire des allèles, les individus d'une population à l'étude peuvent être répartis aléatoirement en groupes selon les allèles hérités de leurs parents. À l'instar des ECR, cette randomisation des allèles permet de répartir facteurs confondants de manière similaire entre les groupes. De plus, puisque les variants sont acquis dès la conception, ils ne sont en pratique pas altérés par l'environnement ou par un trait d'intérêt. Les variants génétiques sont alors définis comme des variables instrumentales (IV) reflétant un facteur de risque modifiable sur un trait d'intérêt et les participants sont répartis en groupes selon l'effet de leurs variants génétiques sur ce facteur de risque (diminution ou augmentation du risque) (77,80,82,86,87). Le concept central derrière la MR est que les variants génétiques sont des facteurs non modifiables, répartis aléatoirement et qui influencent un trait d'intérêt au cours de la vie, comme une maladie, via leur effet sur un facteur de risque. En ce sens, elle est en pratique moins sujette aux biais associés aux facteurs confondants et à la causalité inverse. La MR est ainsi souvent comparée à une forme « naturelle » de ECR. En effet, la MR et le ECR sont

toutes deux des méthodes d'inférences causales utilisées en épidémiologie. Dans les deux cas, les participants sont répartis de manière aléatoire dans chaque groupe afin de réduire le plus possible les biais associés aux facteurs confondants.



**Figure 4. Représentation schématique d'une étude de randomisation mendélienne.** A) Comparaison de la MR avec un ECR. B) Postulats de la MR. Adapté de Hemani et al. *eLife* (2018).

Dans le cas de la MR, si l'on fait le postulat que l'allèle X provoque une augmentation du trait A tandis que l'allèle Y n'a aucun effet sur ce trait et que le trait A cause un trait B, alors les individus possédant l'allèle X présenteront des niveaux plus élevés du trait A que ceux possédant l'allèle Y. Les participants possédant l'allèle X sont alors regroupés dans ce qui correspond au groupe traité d'un ECR, tandis que ceux qui possèdent l'allèle Y forment le groupe témoin. Ce faisant, tout comme un ECR, la MR permet de quantifier l'effet de l'allèle X sur le trait B, via son influence sur le trait A et d'établir un potentiel lien de causalité entre la présence de ce variant génétique et la présence du trait B (Figure 4A). La MR peut être utilisée comme un outil de validation lorsque ses résultats sont corroborés avec les résultats de ECR ou d'études observationnelles (88). Sachant que la recherche et le développement de nouveaux médicaments est un processus coûteux en termes de temps et ressources financières, coûts principalement dus au faible taux d'approbation des médicaments développés, la MR peut également être utilisée afin d'identifier des cibles thérapeutiques prometteuses ou des effets secondaires liés à un traitement, et donc de réduire le champ d'exploration, et d'établir une relation de causalité entre celles-ci et une maladie. Ce faisant, la MR permet d'accélérer le développement de nouveaux médicaments et d'en réduire les coûts (89).

Comme pour toute étude épidémiologique, la MR doit satisfaire certains postulats pour que ses résultats puissent être considérés comme valides (80,86) (Figure 4B). Le premier postulat concerne la pertinence des IV, c'est-à-dire que celles-ci doivent être (robustement) associées au facteur de risque. Le second postulat indique que les IV ne doivent pas partager de cause commune avec le trait d'intérêt. Il s'agit du postulat d'indépendance des variables instrumentales. Le troisième postulat est celui d'exclusion-restriction et stipule que les IV ne doivent affecter le trait d'intérêt que par l'entremise du facteur de risque. En pratique, seul le premier postulat peut être prouvé hors de tout doute en mettant à profit les valeurs d'association statistique entre chaque variant génétique et le facteur de risque. Ceci permet d'obtenir une statistique (généralement la statistique F) qui reflète la force de l'association entre les IV et le facteur de risque (90). La force des IV dépend de l'ampleur de leur effet sur le facteur de risque (la variance du facteur de risque expliquée par celles-ci) et par le nombre d'individus à l'étude. Les autres postulats devraient être évalués à l'aide d'analyses de sensibilité. Ces analyses incluent, entre autres, le test d'hétérogénéité des variants, le test de pléiotropie (horizontale), le test de direction de l'effet et l'identification des variants aberrants. L'identification de variants ayant un effet disproportionné sur le trait d'intérêt permet ensuite de déterminer si ceux-ci ont un effet sur le trait d'intérêt via un autre trait que le facteur de risque (pléiotropie) ou si ceux-ci ont un effet direct sur le trait d'intérêt (hétérogénéité) (91,92). Le test de direction de l'effet vise à établir que la direction du lien de causalité, c'est-à-dire que le facteur de risque influence le trait d'intérêt, et non l'inverse (causalité inverse). La vérification du respect des postulats de la MR permet d'accroître la précision et le niveau de confiance par rapport à l'effet causal des IV sur le trait d'intérêt.

### 3.2. Randomisation mendélienne inverse

Bien que la MR permette de réduire les risques associés à la causalité inverse, ceci ne signifie pas pour autant qu'elle en soit exempte. En effet, les tests de direction de l'effet peuvent parfois être insuffisants pour déterminer la direction de l'effet causal (93,94), particulièrement dans le cas où les variants génétiques pris en compte prédisposent l'individu à un trait d'intérêt ou un facteur confondant qui, en retour, influence le facteur de risque. En ce sens, la MR inverse peut être considérée comme une analyse de sensibilité. Par exemple, une analyse de MR pourrait viser à déterminer si un métabolite X est associé à un risque plus élevé de développer une maladie Y. La présence de certains variants génétiques chez un individu, regroupés dans la IV G, pourrait prédisposer celui-ci à un plus haut risque de développer la maladie Y, laquelle affecterait à son tour les niveaux du métabolite X.

Ainsi, les résultats de la MR pourraient conclure que les variants génétiques identifiés sont associés à une augmentation du risque de développer la maladie Y via leur effet sur le métabolite X, alors que l'augmentation des niveaux de X est en fait due à la maladie Y. Une analyse de MR inverse consiste à utiliser les mêmes IV que l'analyse de MR « classique », mais en considérant que le trait d'intérêt est le facteur de risque, et vice-versa. Dans ce cas, s'il s'avère qu'il existe une association entre le facteur de risque et le trait d'intérêt, alors il est probable que la direction du lien de causalité soit inversée.

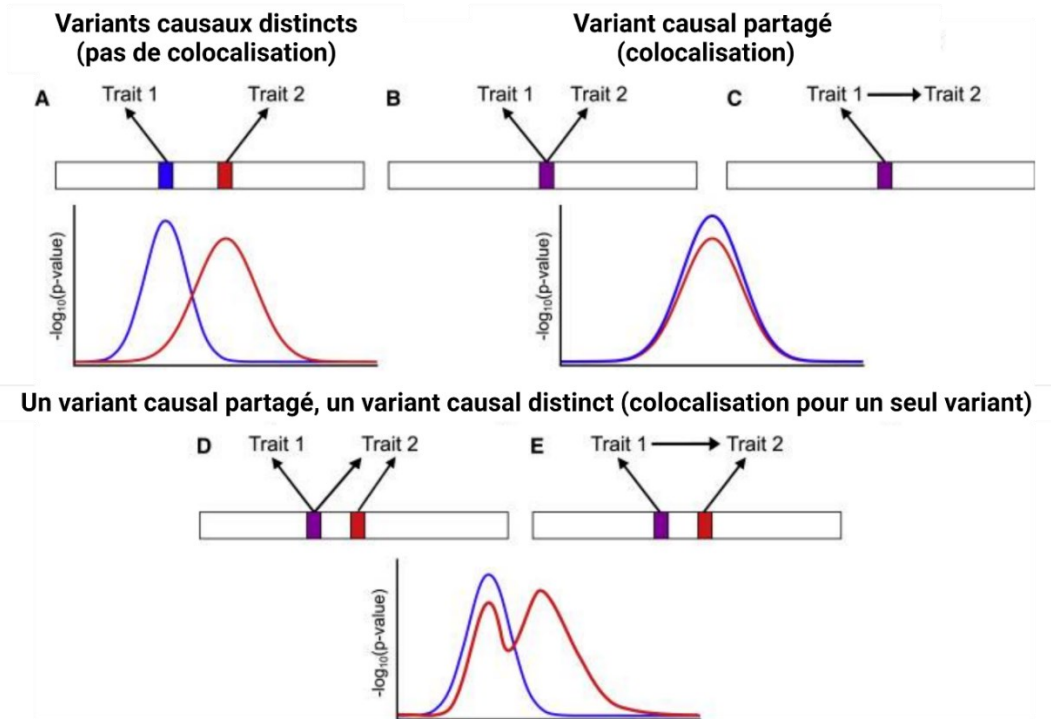
### 3.3. Limites de la randomisation mendélienne

Bien que la MR permette de réduire les risques associés à la causalité inverse et aux effets confondants affectant les études épidémiologiques, cette méthode comporte tout de même certaines limites. Les conclusions d'une étude de MR dépendent de la fiabilité de l'association établie entre les informations génotypiques et phénotypiques. Des facteurs comme la stratification de population ou des erreurs de génotypage peuvent invalider les conclusions d'une étude de MR. De plus, certains effets confondants, comme le LD et la pléiotropie, peuvent être réintroduits dans une étude de MR. Dans le cas du LD, un variant potentiellement causal peut être faussement identifié par son LD avec un autre variant. Ceci peut être problématique, surtout si les deux variants sont situés dans deux gènes différents, ce qui pourrait mener à identifier faussement un gène comme impliqué dans une maladie. Dans le cas de la pléiotropie, un variant peut influencer plusieurs phénotypes intermédiaires différents. Ce variant pourrait influencer un phénotype qui réduit le risque de développer une maladie, tout en influençant un second phénotype qui augmente ce risque. Ainsi, la pléiotropie peut compliquer ou fausser l'interprétation des résultats d'une étude de MR. Enfin, une limite importante de la MR est la présence de variants génétiques pouvant être utilisés comme IV pour un facteur de risque. Dans ce cas, une étude de MR peut être difficile, voire impossible à réaliser (81).

### 3.4. Colocalisation génétique

Les GWAS ont connu un essor considérable au cours des années et ont, entre autres, permis de détecter des associations et de calculer la corrélation génétique entre deux traits (75,78,95,96). La MR utilise des données de GWAS pour déterminer si un trait en influence un autre, car étant donnée la nature observationnelle d'une GWAS, les SNPs identifiés par ces études ne sont pas forcément causaux ou fonctionnels, mais l'identification de mêmes variants génétiques dans plusieurs GWAS

effectuées sur des traits différents suggère que les causes et l'héritabilité de traits complexes s'inscriraient dans un modèle dit omnigénique, plutôt que polygénique, où tous les gènes exprimés dans une cellule pourraient potentiellement affecter un trait donné (97,98). Ce paradigme soulève des inquiétudes quant au fait que le troisième postulat de la MR, soit que les variants génétiques ne doivent affecter le trait d'intérêt que par l'entremise de leur effet sur le facteur de risque, ne soit pas respecté, étant donné le possible LD entre les variants causaux (97,99). Étant donné le nombre grandissant de variants génétiques identifiés par GWAS, il est probable que certains traits soient affectés par des variants qui, bien que distincts, soient en LD, corrélés entre eux. Par exemple, un variant X fortement associé à un facteur de risque pourrait être en fort LD avec un autre variant Y qui affecte le trait d'intérêt, soit directement, soit via un autre facteur de risque. Ainsi, le lien de causalité établi entre le facteur de risque et le trait d'intérêt pourrait être dû à la présence du variant Y (à la corrélation entre X et Y) et donc être erroné. La colocalisation génétique est une méthode statistique bayésienne qui peut pallier ce problème en ce sens qu'elle vise à déterminer, non pas si un trait en cause un autre, mais si deux traits partagent un ou plusieurs mêmes variants causaux (100–102). Dans le cadre des analyses de colocalisation, un variant causal est défini comme un SNP dont les variations génotypiques influencent directement les mécanismes moléculaires qui influencent le trait d'intérêt (99). À la différence d'un variant causal, un variant marqueur est un variant pouvant être corrélé à un variant causal. Étant donné les coûts associés au séquençage du génome entier, les GWAS utilisent généralement des variants marqueurs, représentatifs d'une région du génome marquée par un fort LD (haplotype), obtenus par une analyse partielle du génome. Il existe deux familles de méthodes de colocalisation : la colocalisation proportionnelle et la colocalisation conditionnelle, ou énumérative. La colocalisation proportionnelle, basée sur l'association génétique entre deux traits, utilise un ensemble de SNPs considérés comme étant de bons prédicteurs d'au moins un des deux traits étudiés. Pour chaque trait, un coefficient de régression correspondant à l'effet des variants sur ce trait est calculé. À condition que la LD soit similaire entre les SNPs, cette méthode permet de conclure qu'il y a colocalisation si les coefficients de régression calculés pour chaque trait sont proportionnels (99,100,102). La colocalisation conditionnelle consiste à identifier les SNPs X et Y les plus fortement associés aux traits A et B, respectivement, et à déterminer s'il existe une association entre le trait A et chaque SNP appartenant à X, conditionnellement à chaque SNP appartenant à Y (102).



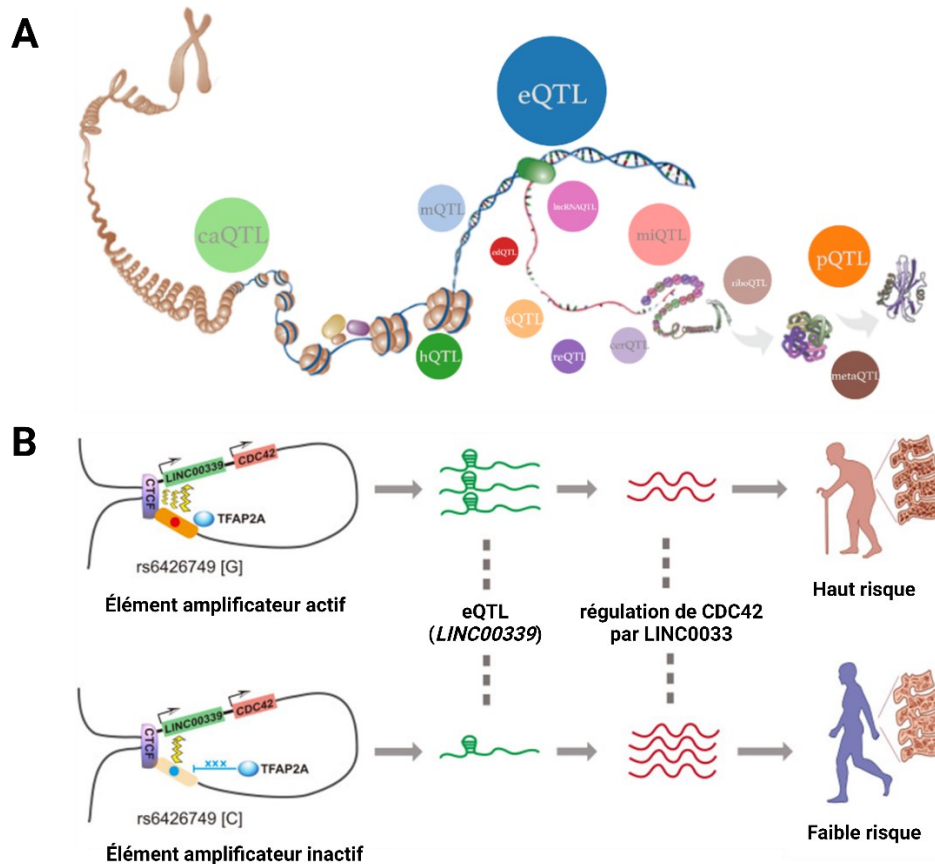
**Figure 5. Différents scénarios de la colocalisation génétique.** A) Absence de colocalisation : deux traits présentent des variants causaux distincts en LD, B) Colocalisation : deux traits partagent un même variant causal, C) Colocalisation : le trait 1 affecte le trait 2 par l'entremise d'un même variant causal, D) Colocalisation : les deux traits partagent un même variant causal, mais le trait 2 présente également un variant distinct et E) Colocalisation : le trait 1 affecte le trait 2 par l'entremise d'un même variant causal, mais le trait 2 présente également un variant distinct. La colocalisation conditionnelle simple ne permet pas de distinguer entre les scénarios B et D, et C et E. Adapté de Zuber et al. *Am J Hum Genet* (2022).

La colocalisation génétique conditionnelle permet de répondre à l'une des cinq hypothèses mutuellement exclusives qui la sous-tendent : (1) il n'existe aucun variant causal, (2,3) il existe un variant causal pour l'un des deux traits, (4) il existe deux variants causaux distincts et (5) il existe un même variant causal pour les deux traits (97). Son intérêt principal est qu'elle permet de déterminer s'il existe, pour deux traits, un seul variant partagé (Figure 5A,B,C) ou deux variants causaux possiblement en LD (99) (Figure 5D,E). La colocalisation génétique conditionnelle ne tient pas compte des postulats de la MR, mais examine plutôt toutes les combinaisons possibles de variants potentiellement causaux pour deux traits. En postulant qu'il existe un variant causal par trait au locus étudié, cette méthode permet de calculer un facteur de Bayes pour chaque combinaison, lequel représente la probabilité que chaque variant soit causal (97). Il existe plusieurs types de colocalisation conditionnelle permettant, entre autres, de déterminer si plus de deux traits partagent un même variant

causal (103) ou s'il existe plus d'un variant causal pour un même trait (104). Généralement, les analyses de colocalisation sont considérées comme des analyses de sensibilité visant à corroborer les résultats d'analyses de MR (99).

### 3.5. Locus de caractère quantitatif

Un caractère quantitatif est un trait dont les variations phénotypiques dues à des facteurs génétiques ou environnementaux sont mesurables. Un locus de caractère quantitatif (QTL) est un locus génétique dont les allèles sont responsables de variations phénotypiques. Les traits quantitatifs complexes sont influencés par plusieurs gènes et peuvent donc être influencés par plusieurs QTL. Un trait peut être affecté par un seul QTL situé sur un seul gène (un locus dit mendélien) qui contribue fortement à la variation phénotypique ou par plusieurs QTL situés sur différents gènes qui contribuent tous individuellement à de faibles variations phénotypiques. Par exemple, la taille est caractère quantitatif qui peut être affecté par la présence de mutations autosomales dans le gène *FGFR3* (achondroplasie) uniquement (105) ou par une multitude de variations génétiques situées sur de multiples gènes (106). En ce sens, les effets des QTL s'inscrivent dans un continuum. De plus, il est commun que des QTL soient en interaction (107). Durant les dernières décennies, les GWAS ont permis d'identifier de nombreux variants génétiques associés à un large éventail de traits, mais une large proportion de ces variants, surtout dans le cas de traits complexes, sont situés dans des régions non codantes du génome, ce qui suggère que les variations phénotypiques dont ils sont responsables ne sont pas dues à leur effet direct sur la fonction du gène ou de son produit, mais par d'autres mécanismes comme la régulation de l'expression (108–111). De même, les gènes ou mécanismes par lesquels ces variants influencent un trait donné sont souvent inconnus (112). Ce faisant, l'étude des QTL a connu un essor important au cours des dernières années et de nombreuses études ont cherché à caractériser et cataloguer des QTL dans divers types cellulaires, tissus ou populations (113–118). Cet effort a permis une cartographie plus fine des QTL étant données, par exemple, les fortes disparités d'expression d'un même gène parfois observées entre deux individus, deux tissus ou deux types cellulaires. Il existe plusieurs types de QTL tirant parti de données omiques diverses (épigénomique, transcriptomique, protéomique, métabolomique, métagénomique) et identifiés par leur effet sur l'accessibilité à la chromatine (caQTL), la méthylation de l'ADN (mQTL), les histones (hQTL), l'expression (eQTL), ou l'épissage alternatif de gènes (sQTL), la liaison à un microARN (miQTL), les niveaux de protéines (pQTL) ou de métabolites (metaQTL), ou le microbiote (mbQTL) (119–121) (Figure 6A).



**Figure 6. Types de QTL et exemple de mécanisme d'action.** A) Principaux types de QTL répertoriés. B) Exemple du mécanisme d'action d'un eQTL sur le risque d'ostéoporose. Adapté de Zheng *et al. Nucleic Acids Res* (2020) et Chen *et al. Am J Hum Genet* (2018).

La cartographie des QTL consiste généralement à effectuer des régressions linéaires pour identifier des associations statistiquement significatives entre des mesures phénotypiques et des variants génétiques situés à proximité ou non. Les QTL peuvent exercer leur effet sur un gène situé à proximité (*cis*), c'est-à-dire à plus ou moins 1 mégabase (Mb) du corps ou du site d'initiation de la transcription de ce gène, ou sur un gène distant (*trans*), voire situé sur un autre chromosome (108). Par exemple, un SNP intergénique (rs6426749) a récemment été identifié comme étant un (*cis*) eQTL affectant l'expression d'un long ARN non codant (*LINC00339*) (Figure 6B) (122). Chez les individus possédant l'allèle C de ce SNP, l'élément amplificateur du gène *LINC00339* est faiblement lié au promoteur, ce qui empêche le facteur de transcription TFAP2A d'activer l'élément amplificateur et entraîne une diminution de l'expression de *LINC00339*. Ce gène est suspecté de réguler à la baisse l'expression du gène *CDC42*, un important régulateur du métabolisme osseux impliqué dans l'ostéoporose.

Les analyses de MR et de colocalisation génétique peuvent également tirer parti des données sur les QTL afin de déterminer, non seulement un gène causal pour un trait d'intérêt, mais également le groupe d'individus, tissu ou type cellulaire dans lequel son effet est médié, s'il existe un même variant causal pour le facteur de risque et le trait d'intérêt (101,110,112,123–126), voire si ce gène est une cible thérapeutique potentielle (126). Par exemple, Richardson *et al.* ont montré que les effets des gènes *ADCY3* et *ADIPOQ* sur le risque de maladies cardiovasculaires seraient possiblement médiés par des modifications dans la méthylation de l'ADN (meQTL) à ces loci (127). De même, en utilisant des données de pQTL, Zhao *et al.* ont pu identifier des protéines dont les niveaux seraient causalement associés à diverses maladies, comme l'insuffisance cardiaque, ainsi que des médicaments ciblant ces protéines (126).

### 3.6. Le génome thérapeutique

Le génome humain comprend entre 20000 et 25000 gènes, lesquels renferment l'information nécessaire pour produire une protéine. Une pléthore de maladies ont été associées à un dérèglement de l'expression d'un gène ou une dysfonction de la protéine produite par celui-ci, faisant de ces protéines des cibles thérapeutiques potentielles pour le traitement de maladies. Il existe quatre grands types de macromolécules avec lesquels peuvent interférer des molécules thérapeutiques, soit les protéines, les polysaccharides, les lipides et les acides nucléiques. Cependant, la recherche de médicaments ciblant les polysaccharides, lipides et acides nucléiques est limitée par la toxicité, la spécificité ou l'efficacité de molécules ciblant celles-ci (128). Ainsi, la plupart des molécules médicamenteuses à ce jour ciblent des protéines, auxquelles elles se lient et dont elles modifient ou modulent la fonction. Le génome humain contiendrait entre 3000 et 5000 gènes faisant partie du génome thérapeutique, c'est-à-dire de gènes codant pour des protéines et dont il est possible de moduler l'expression ou la fonction à l'aide de molécules thérapeutiques (128–130). À ce jour, seul quelques centaines de protéines sont ciblées par des médicaments, tandis que plusieurs milliers de protéines pourraient être de potentielles cibles thérapeutiques (129). Une meilleure compréhension de ces protéines et de leur rôle dans la pathophysiologie de maladies pourrait permettre d'accroître les options thérapeutiques actuellement disponibles. Plusieurs méthodes ont été développées au cours des dernières années, non seulement pour permettre et faciliter l'identification de cibles thérapeutiques et de médicaments, mais également pour en estimer l'efficacité et la viabilité (131–134). Étant donnée la faible efficacité de la prédiction de cibles thérapeutiques basée sur des modèles animaux ou

cellulaires et le fait que la validité d'une nouvelle cible thérapeutique n'est généralement constatée que dans les phases avancées d'un ECR, seul 4% des médicaments testés expérimentalement sont approuvés (135–137). La recherche et l'identification de cibles thérapeutiques peut tirer profit d'études génétiques, comme les GWAS et la MR. Au sein de ces études, les SNPs agissant entre autres sur l'expression, la fonction ou les niveaux d'une protéine permettent d'imiter l'effet pharmacologique d'un médicament sur cette protéine, mais sans nécessiter l'administration du médicament, contrairement à un ECR (135). De fait, les médicaments dont l'identification a tiré profit de données génétiques obtiennent un plus haut taux d'approbation (138–141).

## 4. Problématique, hypothèses et objectifs

La PA est la cause la plus fréquente d'admissions en centre hospitalier liées à des troubles gastrointestinaux et représente un important fardeau économique pour la société (13,142). La sévérité de la maladie est également associée à des coûts plus élevés (143). Bien que l'incidence de la PA varie entre les différentes régions du monde (144), elle semble avoir connu une augmentation soutenue dans les pays occidentalisés au cours des dernières décennies (13,16,17,145). Cette augmentation devrait continuer de croître au cours des prochaines années (16) et semble aller de pair avec l'augmentation mondiale de l'incidence de l'obésité (38,146–148). De plus, malgré une réduction dans le taux de mortalité associé à la PA au cours des dernières années dû aux avancées dans le domaine médical, près du tiers des individus hospitalisés pour la PA développeront une forme récurrente de la maladie dans les mois suivant la première apparition, tandis que près de 10% développeront une forme chronique de la maladie (149–152). La PA, surtout lorsque sa sévérité est considérée comme légère ou modérée, a longtemps été négligée dans le domaine médical (153) et un sous-diagnostic de la maladie peut échouer à prévenir la récurrence de la PA ou, dans les cas sévères, sa mortalité (154). C'est pourquoi il est essentiel de bien comprendre les facteurs qui influencent la PA et sa sévérité en vue de prévenir l'apparition et la progression de cette maladie (12,150,154–156). À l'exception de la modification des habitudes de vie, comme la consommation d'alcool ou le tabagisme, ou du suivi des niveaux de TG sanguins, peu de traitements existent présentement pour prévenir l'apparition de la PA (12,157,158). Malgré le fardeau financier qu'elle représente, les facteurs génétiques de la PA demeurent peu caractérisés (159). En comparaison avec d'autres maladies complexes, les GWAS de PA n'ont identifié à ce jour que quelques loci de susceptibilité à la maladie (69,160–162). Pourtant, la caractérisation de l'architecture génétique des maladies est essentielle afin d'en comprendre les mécanismes et d'identifier de nouvelles cibles thérapeutiques (163). La PA est une maladie qui n'est pas seulement causée par des calculs biliaires ou une dysfonction des cellules acineuses du pancréas. En effet, de plus en plus d'indices suggèrent que des facteurs extérieurs au pancréas ou au foie pourraient être impliqués. Entre autres, l'inflammation systémique joue un rôle important dans la pathophysiologie de la PA (159,164–167). De plus, des facteurs sanguins, comme des protéines ou métabolites impliqués dans divers sentiers métaboliques ou inflammatoires pourraient également contribuer à l'apparition et la sévérité de la PA (159,168–172). Des études de MR mettant à profit les données du protéome sanguin humain ont contribué à mieux caractériser les facteurs génétiques impliqués dans plusieurs maladies complexes, surtout celles qui sont influencées par des protéines

sanguines, et à identifier de nouvelles cibles thérapeutiques (173–175). L'identification de protéines causalement associées à la PA pourrait permettre de mieux en comprendre l'architecture génétique tout en permettant d'identifier de potentielles cibles thérapeutiques pour cette maladie.

Notre **hypothèse** est qu'il est possible d'identifier des protéines circulantes associées à la pancréatite aiguë en mettant à profit les résultats d'études d'association pangénomiques dans une analyse de randomisation mendélienne à l'échelle du protéome.

Notre **premier objectif** consiste à caractériser l'architecture génétique de la PA à l'aide d'une méta-analyse de GWAS de PA. Pour ce faire, nous avons effectué deux nouvelles GWAS de PA à partir de données de participants de la UK Biobank et de la Estonian Biobank, respectivement. Nous avons ensuite combiné ces résultats à ceux d'une GWAS effectuée dans la cohorte FinnGen (Finlande) dans une méta-analyse comprenant 10 630 cas de PA et 844 679 témoins. Les cas de PA de ces études ont tous été identifiés par dossiers de patients électroniques et nous avons exclu, des cas et des témoins, les individus qui présentaient d'autres pathologies du pancréas ou associées au pancréas. Nous avons ensuite utilisé cette méta-analyse de GWAS pour calculer l'héritabilité de la PA et la corrélation génétique avec plusieurs autres traits potentiellement impliqués dans ou associés à la PA.

Notre **second objectif** consiste à identifier des protéines circulantes associées à la PA. Pour ce faire, nous avons utilisé les données de protéines (pQTL) des cohortes deCODE (176) (Islande) et Fenland (177) (Royaume-Uni). Ces cohortes incluent les données de 4907 protéines mesurées chez 35 559 individus et 4979 protéines mesurées chez 10 708 individus, respectivement. Nous avons ensuite effectué une série d'analyses de MR dans ces deux cohortes séparément pour déterminer le potentiel effet causal de chaque protéine, considérée comme un facteur de risque, sur la PA, considérée comme le trait d'intérêt. Pour corroborer les résultats obtenus par MR, nous avons effectué une série d'analyses de sensibilité et de colocalisation génétique pour chaque protéine pour laquelle nous détectons une association statistiquement significative avec la PA. Pour chacune de ces protéines identifiées, nous avons également effectué des analyses de MR inverse pour valider la direction du lien de causalité.

Notre **troisième objectif** consiste à identifier de nouvelles cibles thérapeutiques pour la PA. Pour ce faire, nous avons évalué le potentiel thérapeutique des gènes et protéines identifiés précédemment à

l'aide de la plateforme Open Targets (178), une plateforme qui recoupe des données publiques sur les facteurs génétiques d'une maladie, ses cibles thérapeutiques et leurs médicaments connus.

## 5. Résumé des méthodes

### 5.1. Cohortes

La méta-analyse de GWAS inclut trois cohortes distinctes : la UK Biobank, la Estonian Biobank et FinnGen. La UK Biobank est une grande cohorte prospective incluant plus de 500 000 participants du Royaume-Uni âgés entre 40 et 69 et dont les données ont été récoltées entre 2006 et 2010. Elle comprend, entre autres, des données anthropométriques, de génotypage à l'échelle du génome, de séquençage du génome (WGS) et l'exome entiers (WES), de mesures de métabolites, etc. Elle comprend également les dossiers médicaux électroniques provenant des bases de données de la Hospital Episode Statistic et des registres nationaux sur le décès et sur le cancer pour tous les participants (179). La Estonian Biobank comprend les données de plus de 200 000 participants de l'Estonie représentant l'âge, le sexe et la distribution géographique de la population adulte de l'Estonie. Elle comprend, entre autres, des données personnelles, généalogiques, liées au mode de vie et de génotypage de tous les participants, ainsi que les données de séquençage de l'ARN et de l'exome et de métabolomique de plusieurs d'entre eux (180). FinnGen comprend les données de près de 350 000 participants de la Finlande et comprend des données anthropométriques, sur le mode de vie et de génotypage de tous les participants (181).

Les analyses de MR à l'échelle du protéome (PWMR) incluent deux cohortes distinctes : la cohorte deCODE et la cohorte Fenland. La cohorte de deCODE genetics est une cohorte islandaise qui a généré des données de WGS en utilisant des puces Illumina pour 49 708 participants dont les données génotypiques et phénotypiques sont disponibles pour imputer le génome de 166 281 Islandais génotypés, pour un total de plus de 27 millions de variants imputés. Les protéines sanguines ont été mesurées chez 35 559 de ces individus. Au total, 4719 protéines sanguines uniques ont pu être mesurées à partir de 4907 aptamères. Les loci de caractère quantitatif de protéine (pQTL) ont été générés à partir d'associations génétiques à l'échelle du génome pour les niveaux de chacun des 4907 aptamères. Une GWAS a ensuite été effectuée pour chacune des protéines. Un pQTL était considéré en *cis* s'il était situé à tout au plus 1 Mb du site d'initiation de la transcription du gène encodant la protéine ciblée (176). La cohorte Fenland a généré des données génotypiques et phénotypiques pour 12 435 participants d'ascendance britannique, pour un total de 17 millions de variants génotypés ou imputés. Au total, plus de 4775 protéines uniques ont pu être mesurées à partir de 4979 aptamères. Les pQTL ont été générés à partir d'un modèle additif et les résultats ont ensuite été combinés pour

tous les réseaux de génotypage dans une méta-analyse à effets fixes. Un pQTL était considéré en *cis* s'il était situé à tout au plus 500 000 kb du corps du gène encodant la protéine ciblée (177).

L'analyse de colocalisation génétique a utilisé, entre autres, les données du consortium GTEx (version 8). GTEx comprend les données de 17 382 échantillons de 54 tissus prélevés chez 948 individus décédés âgés entre 20 et 70 ans, parmi lesquels 838 sont également génotypés. Elle inclut entre autres des données sur le sexe, l'âge, l'ethnie, la cause de la mort et le nombre de tissus prélevés chez chaque individu. GTEx a produit des données de WGS pour 838 participants, pour un total de plus de 10 millions de variants génotypés ou imputés ayant une fréquence allélique plus grande ou égale à 1%. Le séquençage de l'ARN a été effectué à l'aide du séquenceur Illumina TruSeq ou du réseau d'expression Affymetrix Human Gene 1.1 ST. Les transcrits ont été alignés au génome de référence (GRCh38/hg38) à l'aide du logiciel STAR (182). La quantification des transcrits et les valeurs de transcrits par million (TPM) ont été produites à l'aide du logiciel RNA-SeqQC (version 1.1.9) (183), puis les eQTL ont été produits à partir de ces valeurs et des données de génotypage à l'aide du logiciel FastQTL (108,184).

## 5.2. Études d'association pangénomique et méta-analyse

Afin d'effectuer la méta-analyse de GWAS de PA, nous avons effectué deux nouvelles GWAS dans la UK Biobank et la Estonian Biobank. Pour ce faire, nous avons utilisé la méthode implémentée dans le logiciel SAIGE (185). Cette méthode, basée sur des modèles mixtes généralisés, a été développée pour tenir compte du déséquilibre entre le nombre de cas et de témoins, la parenté entre les individus et la structure populationnelle. Dans chacune des deux nouvelles GWAS produites, afin de tenir compte de potentiels facteurs confondants et d'accroître la précision des résultats, nous avons inclus dans notre modèle le sexe, l'âge et les 10 premières composantes principales basées sur l'ascendance en tant que covariables. Cette méthode a également été utilisée dans la cohorte FinnGen. Nous avons ensuite effectué une méta-analyse à effets fixes de ces trois cohortes en utilisant le logiciel METAL (186). Au total, cette méta-analyse inclut 10 630 cas de PA et 844 679 témoins. Nous avons également effectué une méta-analyse à effets aléatoires afin de tenir compte de l'hétérogénéité génétique des variants inclus. En tout, cette méta-analyse a permis l'étude de 9 570 209 SNPs ayant une fréquence allélique plus grande ou égale à 1%. Nous avons testé la possible inflation des résultats obtenus étant donné que les études de cas-contrôle sont susceptibles d'être affectées par la

stratification de population, laquelle survient lorsque différentes proportions de cas et témoins sont issus de différentes populations. Celle-ci peut accroître le nombre de faux positifs parmi les résultats lorsque les associations détectées sont dues aux différences entre les populations et non aux différences observées pour le trait d'intérêt (187). Le test d'inflation des résultats d'une GWAS, également appelé « contrôle génomique », est fondé sur la distribution du khi carré et compare les valeurs statistiques (valeurs p) observées pour chaque SNP de la GWAS par rapport à celles attendues sous une distribution théorique du khi carré. Si les valeurs observées divergent rapidement de ce qui est attendu, il est possible que les résultats de la GWAS soient biaisés par la stratification de population (188,189). Pour vérifier si les résultats de la GWAS de PA pouvait être biaisés par la stratification de population, nous avons utilisé l'extension *qqman* (190) du logiciel R. Pour chacun des SNPs les plus significatifs à un locus donné, nous avons utilisé l'extension *gassocplot* (191) du logiciel R afin de visualiser leur position sur le génome ainsi que les variants en LD avec ceux-ci.

### 5.3. Héritabilité et corrélation génétique

Étant donné l'absence de variants rares dans la GWAS de PA, l'héritabilité de SNP a été sélectionnée pour calculer l'héritabilité de la PA et le point d'intersection du score de régression du LD à l'aide du logiciel LDSC (75,96). D'une part, l'héritabilité permet de déterminer la part de variation phénotypique pour être attribuée à des facteurs génétiques, d'autre part, le point d'interception du score de régression de LD reflète l'influence de facteurs confondants sur les résultats obtenus (192). Afin d'établir la corrélation génétique entre la PA et d'autres traits similaires ou potentiellement impliqués dans la PA ou dans les mêmes sentiers biologiques, une analyse croisée de phénotypes a été effectuée à l'aide de la base de données interactive de l'outil en ligne Cross-Phenotype Analyse of GWAS database (iCPAGdb) (193). iCPAGdb met à profit des associations génétiques spécifiques à la structure de LD de populations déterminées pour calculer des valeurs d'enrichissement entre différents phénotypes, c'est-à-dire de phénotypes qui partagent une architecture génétique similaire. Cet outil effectue un test exact de Fisher, corrigé pour un taux potentiel de 5% de fausses découvertes (FDR), afin de quantifier la significativité de la proportion de SNPs partagés entre deux traits, et calcule les indices de similarité de Jaccard, Sorensen et Chao-Sorensen pour quantifier la similarité entre ces traits. Étant donné l'absence de données individuelles disponibles pour les traits à l'étude ainsi que le fait que certains des traits compris dans iCPAGdb sont des traits plus généraux (par exemple, la mesure de tous les phytostérois, plutôt que la mesure individuelle de différents phytostérois), nous

avons ensuite utilisé le logiciel LDSC pour calculer la corrélation génétique entre la PA et des traits inclus ou similaires, mais plus spécifiques, à ceux utilisés par iCPAGdb. L'analyse de corrélation génétique a ainsi été employée pour déterminer s'il existe une relation entre deux traits, c'est-à-dire si un trait affecte l'autre ou si les deux traits s'affectent mutuellement, mais cette interaction peut être causale ou due à une interaction non prise en compte avec un tiers trait (194). Le potentiel lien de causalité entre les traits corrélés significativement a ensuite été analysé dans une étude de MR.

#### 5.4. Analyses de randomisation mendélienne à l'échelle du protéome

Les analyses de PWMR ont tiré parti des données de pQTL des cohortes deCODE et Fenland. Dans les deux cas, les pQTL obtenus par GWAS pour chaque protéine ont été considérés comme les facteurs de risque et la PA a été considérée comme le trait d'intérêt. En d'autres termes, il s'agissait de déterminer, pour chaque protéine prise individuellement, si les variants génétiques significativement associés à une modulation de ses niveaux sanguins avaient un effet causal sur la PA. La même stratégie analytique a été employée pour les données de chacune des deux cohortes. Pour chaque aptamère (ou SOMAmer) mesuré, seuls les SNPs ayant un effet en *cis*, c'est-à-dire situés à plus ou moins 1 Mb du corps du gène encodant la protéine cible, ont été inclus dans l'étude de PWMR. Parmi ces SNPs, seuls ceux présentant une fréquence allélique plus grande ou égale à 1%, significatifs à l'échelle du génome, c'est-à-dire dont la valeur p d'association est plus petite ou égale à  $5 \times 10^{-8}$ , et considérés comme étant indépendants, c'est-à-dire dont le coefficient de détermination de la LD avec d'autres SNPs à proximité est plus petit que 0.1, ont été conservés. Ces SNPs constituaient alors l'instrument génétique (variable instrumentale) utilisée pour tester l'association entre la protéine et la PA. En fonction du nombre de SNPs restants après ce tri, l'estimé de l'effet causal de ces SNPs sur la PA a été calculé par la méthode de la variance inverse pondérée (IVW) ou du test statistique de Wald à l'aide de l'extension *TwoSampleMR* (87) du logiciel R. Étant donné la multiplicité des analyses effectuées simultanément et à agréger dans une même conclusion (195), une correction pour un FDR de 5% a été appliquée aux valeurs de significativité de chaque association. Des analyses de sensibilité ont ensuite été effectuées pour vérifier la robustesse de chaque association. La statistique Q de Cochran a été calculée pour mesurer l'hétérogénéité des SNPs inclus et une analyse de régression de Egger a été effectuée pour mesurer les potentiels effets pléiotropes de ceux-ci. Ces tests ont été considérés comme étant significatifs si leur valeur p était plus petite que 0.05. Dans le cas contraire, il était possible de conclure que les estimés causaux n'étaient pas biaisés par l'hétérogénéité ou la

pléiotropie. La statistique F de Cragg-Donald (90) a été calculée pour chaque instrument génétique afin d'en établir la force. Afin de valider ces résultats, différentes méthodes ont été utilisées pour déterminer les SNPs devant être priorisés dans l'analyse de PWMR, c'est-à-dire ceux qui, parmi des multiples variants corrélés, sont le plus susceptibles d'être responsables des variations phénotypiques mesurées. Ces méthodes incluent JAM-MR (196) et l'analyse par composante principale (PCA). JAM-MR est une méthode bayésienne stochastique basée sur un algorithme Monte-Carlo par chaînes de Markov à sauts réversibles. Différents paramètres de calibrage ont été utilisés pour la pénalisation des SNPs ayant un effet pléiotrope afin de créer différents modèles permettant chacun de calculer un estimé causal. Cette méthode agrège ensuite tous les estimés causaux en un seul, représentatif de l'effet causal moyen. La méthode PCA calcule les composantes principales d'une matrice de corrélation entre les différents SNPs afin d'identifier ceux qui permettent d'expliquer la plus grande part de la variance phénotypique observée pour le facteur de risque. Cette méthode met à profit les données de référence de LD spécifique à une population, européenne en l'occurrence, pour réduire les risques d'induire un biais associé à la faible force des instruments, causée par une forte corrélation entre les variants. Toutes les composantes principales, donc les SNPs, expliquant jusqu'à 99% de la variance du facteur de risque ont été incluses. Ces deux méthodes ont chacune permis de calculer un estimé causal tenant compte des différents biais potentiels pouvant affecter l'analyse de PWMR et les différents effets causaux obtenus ont ensuite été comparés entre eux pour vérifier la robustesse du lien de causalité. En vue de corroborer les résultats des analyses de PWMR et déterminer si le facteur de risque et le trait d'intérêt sont causés par un même variant génétique, trois analyses de colocalisation ont été effectuées : l'analyse de colocalisation simple, l'analyse de colocalisation multiple et l'analyse de colocalisation conditionnelle pairée. Ces trois analyses ont été effectuées à l'aide des extensions *COLOC* (97) et *HyPrColoc* (103) du logiciel R et du logiciel *PWCoCo* (197), respectivement. La colocalisation simple est une méthode bayésienne n'utilisant que des statistiques sommaires de GWAS et aucune information sur la LD entre les SNPs. Cette méthode a été utilisée pour déterminer la probabilité que chaque protéine prise individuellement partage un même SNP causal avec la PA. *HyPrColoc* est un algorithme bayésien de classification divisionnaire qui vise à identifier des groupes de traits qui partagent un même ou différents variants causaux à un même locus génétique. Les traits considérés incluent généralement le facteur de risque et le trait d'intérêt ainsi que d'autres traits sélectionnés en fonction de l'interprétation biologique de l'association détectée, de l'hypothèse de recherche, etc. *PWCoCo* est une méthode basée sur *COLOC*, mais prenant en compte les possibles

effets confondants liés au LD dans l'identification de SNPs causaux partagés. Cette méthode compare tous les SNPs conditionnellement indépendants du facteur de risque et conditionne leur effet en fonction de leur corrélation (LD) avec chacun des SNPs conditionnellement indépendants du trait d'intérêt. Cette méthode présente l'intérêt de réduire les risques d'inférence causale fautive due à la présence de plusieurs variants corrélés situés à un même locus. En d'autres termes, elle permet de déterminer si un SNP a un effet causal pour les deux traits ou si son effet détecté est dû à sa corrélation avec le véritable variant causal. Pour toutes les analyses de colocalisation, seuls les variants ayant une fréquence allélique supérieure à 1% et situés à 500 kb du corps du gène encodant la protéine cible ont été inclus.

Dans le cas de l'analyse de MR inverse, chaque protéine a été individuellement considérée comme le trait d'intérêt tandis que la PA a été considérée comme le facteur de risque. Dans ce cas, seuls les SNPs inclus dans la GWAS de PA présentant une fréquence allélique plus grande ou égale à 1%, significatifs à l'échelle du génome et considérés comme étant indépendants ont été conservés. L'estimé de l'effet causal des SNPs associés à la PA sur les niveaux de chaque protéine individuellement a été calculé en fonction de la IVW à l'aide de l'extension *TwoSampleMR*. Les analyses de sensibilité incluaient le test d'hétérogénéité de Cochran, le test de pléiotropie de Egger et une analyse de colocalisation simple, effectuée avec *COLOC*.

## 5.5. Loci de traits quantitatifs d'expression

L'analyse de colocalisation multiple avec HyPrColoc a mis à profit les données d'expression hépatique du consortium GTEx. Étant donné que GTEx utilise les données de tous les donneurs, et non des donneurs d'ascendance européenne seulement, les eQTLs propres à la population européenne ont dû être recalculés. Au total, les données d'expression hépatique de 177 donneurs d'ascendance européenne étaient disponibles. Les valeurs d'expression en TPM ont d'abord été normalisées pour tous les sujets européens seulement et tous les gènes, puis les eQTLs pour les gènes d'intérêt ont été calculés à l'aide du logiciel QTLtools (198). Ce calcul a pris en compte une correction pour différentes covariables, incluant les effets discontinus liés aux différentes plateformes utilisées pour le séquençage de l'ADN et de l'ARN, le sexe des participants, les 30 premiers facteurs individuels et les 10 premières composantes principales. Les facteurs individuels et composantes principales sont utilisés pour

accroître la sensibilité et l'interprétabilité des résultats en identifiant de potentiels facteurs confondants et leur effet sur l'expression des gènes (198,199).

## 5.6. Identification de cibles thérapeutiques

Les données publiquement accessibles de la base de données du consortium Open Targets (178) ont été utilisées afin d'identifier des cibles thérapeutiques potentielles pour la PA. Open Targets est une initiative à grande échelle et pluriannuelle née d'un partenariat public-privé qui utilise des données génétiques et génomiques humaines pour l'identification de cibles thérapeutiques. Cette base de données contient, entre autres, des informations sur les associations connues entre des cibles thérapeutiques et des maladies, des informations les cibles thérapeutiques connues, comme leurs expression et interactions avec d'autres cibles, et des données sur une vaste gamme de médicaments, comme leurs mécanismes d'action, statut clinique et effets secondaires. Ces données ont été interrogées systématiquement pour chaque protéine significativement et robustement associée à la PA. Parmi les associations entre une protéine et un médicament, seules celles dont les valeurs d'expression et du niveau de la protéine spécifiques au tissu d'intérêt, de spécificité à un type cellulaire du tissu d'intérêt et d'effets secondaires pour lesquelles la métrique calculée était dans le 95<sup>e</sup> percentile et non nulle ont été conservées. Enfin, parmi ces associations conservées, toutes celles dont le mécanisme d'action du médicament était incohérent avec l'effet estimé par l'analyse de PWMR ont été rejetées.

# Chapitre 1. Proteome-Wide Mendelian Randomization Identifies Causal Links Between Blood Proteins and Acute Pancreatitis

Jérôme Bourgault<sup>1</sup>, Erik Abner<sup>2</sup>, Hasanga D. Manikpurage<sup>1</sup>,  
Natàlia Pujol-Gualdo<sup>2</sup>, Triin Laisk<sup>2</sup>, Estonian Biobank Research Team, Émilie Gobeil<sup>1</sup>,  
Eloi Gagnon<sup>1</sup>, Arnaud Girard<sup>1</sup>, Patricia L. Mitchell<sup>1</sup>, Sébastien Thériault<sup>1,3</sup>, Tõnu Esko<sup>2</sup>,  
Patrick Mathieu<sup>1,4</sup> and Benoit J. Arsenault<sup>1,5\*</sup>

- 1) Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec, Québec (QC), Canada
- 2) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Riia 23b, 51010, Estonia
- 3) Department of Molecular Biology, Medical Biochemistry and Pathology, Faculty of Medicine, Laval University, Quebec (QC), Canada
- 4) Department of Surgery, Faculty of Medicine, Université Laval, Québec (QC), Canada
- 5) Department of Medicine, Faculty of Medicine, Université Laval, Québec (QC), Canada

## 1.1. Résumé

**Mise en contexte et objectifs** : La PA est une maladie complexe et la cause principale des admissions en centre hospitalier liés à des troubles gastrointestinaux. Peu d'options thérapeutiques existent présentement pour la prévention de la PA. Des protéines sanguines causalement impliquées dans la PA pourraient représenter des cibles thérapeutiques prometteuses, mais peu ont été causalement associées à la PA à ce jour. Notre objectif était d'identifier des protéines sanguines associées à la PA en combinant les résultats d'une méta-analyse de GWAS dans des analyses de PWMR.

**Méthodes** : Nous avons effectué une méta-analyse de GWAS comprenant 10 630 cas de PA et 844 679 témoins ainsi qu'une série d'analyses de MR estimant les effets causaux par IVW et utilisant les variants en *cis* pour 4719 protéines sanguines de l'étude deCODE (N = 33 559) et 4979 protéines sanguines de l'étude Fenland (N = 10 708).

**Resultats** : La méta-analyse a identifié des variants significatifs à l'échelle du génome ( $P < 5 \times 10^{-8}$ ) situés à 5 loci (*ABCG5/8*, *TWIST2*, *SPINK1*, *PRSS2* et *MORC4*). Les analyses de PWMR ont identifié 68 protéines sanguines uniques pouvant être causalement associées à la PA, incluant 29 protéines validées dans les deux jeux de données. L'annotation fonctionnelle de ces protéines a confirmé l'expression de plusieurs d'entre elles dans des tissus métaboliques responsables de la digestion et du métabolisme énergétique, dont l'œsophage, le tissu adipeux et le foie, ainsi que dans les cellules acineuses du pancréas. Les analyses de colocalisation génétique et l'étude du génome thérapeutique ont également identifié de potentielles cibles thérapeutiques pour la PA.

**Conclusions** : Cette vaste méta-analyse de GWAS pour la PA a identifié de nouveaux variants génétiques associés à la PA ainsi que plusieurs protéines sanguines pouvant être causalement associées à la PA. Cette étude fournit de nouvelles informations sur l'architecture génétique de cette maladie et a identifié des sentiers biologiques impliqués dans la PA, lesquels pourront être explorés pour en évaluer le potentiel thérapeutique pour la prévention ou le traitement de la PA.

## 1.2. Abstract

**Background & Aims :** Acute pancreatitis (AP) is a complex disease and the leading cause of gastrointestinal disease–related hospital admissions. Few therapeutic options exist for AP prevention. Blood proteins with causal evidence may represent promising drug targets, but few have been causally linked with AP. Our objective was to identify blood proteins linked with AP by combining genome-wide association meta-analysis and proteome-wide Mendelian randomization (MR) studies.

**Methods :** We performed a genome-wide association meta-analysis totalling 10,630 patients with AP and 844,679 controls and a series of inverse-variance weighted MR analyses using cis-acting variants on 4719 blood proteins from the deCODE study (N = 35,559) and 4979 blood proteins from the Fenland study (N = 10,708).

**Results :** The meta-analysis identified genome-wide significant variants ( $P < 5 \times 10^{-8}$ ) at 5 loci (ABCG5/8, TWIST2, SPINK1, PRSS2 and MORC4). The proteome-wide MR analyses identified 68 unique blood proteins that may causally be associated with AP, including 29 proteins validated in both data sets. Functional annotation of these proteins confirmed expression of many proteins in metabolic tissues responsible for digestion and energy metabolism, such as the esophagus, adipose tissue, and liver as well as acinar cells of the pancreas. Genetic colocalization and investigations into the druggable genome also identified potential drug targets for AP.

**Conclusions :** This large genome-wide association study meta-analysis for AP identified new variants linked with AP as well as several blood proteins that may be causally associated with AP. This study provides new information on the genetic architecture of this disease and identified pathways related to AP, which may be further explored as possible therapeutic targets for AP.

### 1.3. Introduction

AP is a common and complex disease. AP is the leading cause of gastrointestinal disease–related hospital admissions, and its frequency is increasing worldwide (1). Population prevalence increases in hypertriglyceridemia and metabolic syndrome are suspected to contribute to the rise in the number of individuals with AP, which is responsible for ~200,000 hospitalizations per year in the United States alone (2).

Few, if any, preventative treatments currently exist for AP. Despite the financial burden and clinical importance of AP, the genetic factors linked with this disease remain poorly characterized (3). In comparison with other complex diseases, GWASs have only identified a handful of risk loci (ABCG5/8, SPINK1, CTSC) linked to chronic or AP and have suggested low to moderate heritability (4).

AP is not a disease of pancreatic acinar cells alone, and increasing evidence suggests that factors outside the pancreas or the liver may be involved. Systemic inflammation plays a crucial role in the pathophysiology of AP (5,6). Blood factors, such as proteins from various metabolic or inflammatory pathways, may also be involved in AP (3). Leveraging data from the human blood proteome in MR studies has contributed to a deeper characterization of the genetic factors implicated in several complex diseases, especially those that may be influenced by circulating factors (7,8). Circulating proteins with causal evidence may represent promising drug targets. Identifying proteins causally associated with AP may accelerate our understanding of the genetic architecture of this disease while simultaneously identifying drug targets for this disease.

Here, we performed 2 new GWASs of AP (UK Biobank and Estonian Biobank) and combined these results in a genome-wide association meta-analysis with those of FinnGen for a total sample size of 10,630 cases and 844,679 controls. Leveraging GWAS summary statistics data from the deCODE (4719 proteins measured in 35,559 individuals; deCODE Genetics) and the Fenland cohorts (4979 proteins measured in 10,708 individuals), we further performed PWMR studies that identified several new blood proteins that may be causally linked with AP.

## 1.4. Methods

### Genome-Wide Association Meta-analysis of Acute Pancreatitis

We performed a GWAS meta-analysis of 3 cohorts: the UK Biobank, the Estonian Biobank, and FinnGen. We performed 2 new GWASs in the Estonian Biobank and the UK Biobank. In the UK Biobank, we included 3247 cases and 405,102 controls (White British). All participants of UK Biobank present extensive electronic health records coming from 3 national databases: Hospital Episode Statistic data and national death and cancer registries. These databases use normalized diagnostic codes from the International Classification of Diseases (ICD)-10th revision and surgical procedure codes of OPSC-4 (Office of Population, Censuses and Surveys Classification of Interventions and Procedures, version 4). To define AP, we selected participants with ICD-10 entries for AP (K85). In the Estonian Biobank analysis, we included data from 2735 cases (ICD-10 K85) and 166,135 controls of European ancestry (ICD-10 K80-K87 excluded from controls). We then combined these results with the FinnGen data freeze 7 (June 2022), comprising 4648 AP cases (ICD-10 K85) and 273,442 controls.

We used the SAIGE (9) (Scalable and Accurate Implementation of Generalized Mixed Models) method to perform the GWAS in the UK Biobank and in the Estonian Biobank. This method is based on generalized mixed models and was developed to control for case-control imbalance, sample relatedness, and population structure. In this analysis, sex, age, and the 10 main ancestry-based principal components were used as covariates. SAIGE was also used to obtain GWAS summary statistics of the FinnGen cohort.

In the UK Biobank genome-wide genotyping was available for >28 million genetic markers directly genotyped or imputed by the Haplotype Reference Consortium, UK10K, and 1000 Genomes Project (1000G) panels. In the Estonian Biobank, genome-wide genotyping was available for >26 million genetic markers, using a custom imputation panel (10). In FinnGen, GWAS was performed using >16 million genetic markers genotyped with the Illumina or Affymetrix (Thermo Fisher Scientific) arrays or imputed using the population-specific SISu 3 reference panel (THL Biobank). The genome-wide association meta-analysis performed from these data sets totaled 10,630 cases and 844,679 controls.

We performed a fixed-effect GWAS meta-analysis of the UK Biobank, FinnGen, and Estonian Biobank cohorts using the METAL (11) package. A random-effect analysis was also done to account for heterogeneity. A total of 9,570,209 single-nucleotide polymorphisms (SNPs) with a minor allele

frequency of  $\geq 0.01$  were investigated. The quantile-quantile plot was generated using the qqman package (12) in R software (R Foundation for Statistical Computing). The SNP-based heritability and linkage disequilibrium (LD) score regression intercept were computed using the LDSC software (13). The R package gassocplot (14) was used to plot regional association plots for the top SNP at each of the genome-wide significant loci identified.

### **Institutional Review Board Approval**

The activities of the Estonian Biobank are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of the Estonian Biobank. Individual-level data analysis in the Estonian Biobank was performed under ethical approval 1.1-12/624 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 3-10/GI/20956 from the Estonian Biobank. The UK Biobank analyses were performed using data application number 25205. The UK Biobank received approval from the British National Health Service, North West—Haydock Research Ethics Committee (16/NW/0274). All participants provided informed consent at the baseline assessment. GWAS summary statistics for FinnGen were obtained from the FinnGen portal ([https://www.finnngen.fi/en/access\\_results](https://www.finnngen.fi/en/access_results)).

### **Cross-Phenotype Genetic Correlation of Acute Pancreatitis**

The GWAS for AP was investigated for genetic correlation with other traits using iCPAGdb (15). iCPAGdb uses ancestry LD-specific association data to compute cross-phenotype enrichment analyses. We used the National Human Genome Research Institute-European Bioinformatics Institute GWAS catalog data for 3793 traits to investigate genetic correlations with AP. iCPAGdb reports pairwise traits signals and shared signals from traits associations with LD proxy SNPs. The output data include Fisher's exact test with adjustment for 5% FDR and Bonferroni's, and Jaccard's, Sorensen's, and Chao-Sorensen similarity indexes. We then used LDSC to compute genetic correlation between AP and 58 traits included in or similar to those used by iCPAGdb. Only SNPs with minor allele frequency  $>1\%$  were retained, and the genetic correlation was computed according to the 1000G LD reference data set for Europeans provided by LDSC by default. Finally, we performed a series of inverse-variance weighted (IVW)-MR analyses using only genome-wide significant ( $P \leq 5 \times 10^{-8}$ ) and independent ( $R^2 < 0.1$ ) SNPs to assess whether these traits could be causally linked with AP and vice versa.

### **Proteome-Wide Mendelian Randomization Study Using deCODE Protein Quantitative Trait Loci**

deCODE Genetics generated whole-genome sequencing data using Illumina SNP chips for 49,708 Icelanders with genotype and phenotype information and was used to impute the genomes of 166,281 genotyped Icelanders, which resulted in >27 million imputed variants. Of these individuals, 35,559 also had proteins measurements. The 4719 unique plasma proteins were measured from 4907 aptamers using the SomaScan multiplex aptamer assay (version 4), which only includes aptamers with high correlation between samples and excludes nonhuman aptamers or aptamers mapping to multiple genes.

The pQTLs were generated from genome-wide association testing using the 4907 aptamer levels adjusted for rank-inverse normal transformed for age, sex, and sample age as phenotypes and 27.2 million imputed variants as genotypes. The 4907 aptamers were used to perform the PWMR study. For each aptamer, we included cis-acting SNPs (ie, SNPs located within  $\pm 1$  Mb) window from the gene body of the target gene. From these SNPs, only those with a minor allele frequency of  $\geq 1\%$  that were genome-wide significant ( $P < 5.0 \times 10^{-8}$ ) and considered independent ( $LD R^2 < 0.1$  in 1000G) were retained.

Depending on the number of SNPs remaining, we then performed IVW or Wald's ratio MR analyses on each protein individually using the TwoSampleMR package (16) in R. We used a 5% FDR correction to correct for multiple testing. Sensitivity analyses were performed using Cochran's Q test and Egger's regression intercept and were considered significant if the  $P$  value was  $< .05$ . Joint analysis of marginal summary statistics (JAM)-MR software performs a Bayesian stochastic search through a reversible-jump Markov Chain Monte Carlo algorithm to identify variants to include in the MR analysis. Different tuning parameters were used for pleiotropy penalization, each of which was used to create a model to compute causal effect estimates. JAM-MR then averages the estimates across all models and returns an aggregate causal effect estimate. The principal component analysis used the LD reference for Europeans to alleviate the potential weak instrument bias caused by correlated variants. We included all principal components explaining up to 99% of the variance.

Hypothesis Prioritisation in multi-trait Colocalization (HyPrColoc) is a Bayesian divisive clustering algorithm that aims at identifying clusters of traits that colocalize at distinct causal variants in a same genomic region. We used 4 traits, including AP, UGT1A1 liver expression quantitative trait loci (eQTLs) and pQTLs, and gallstones. The data for the UGT1A1 eQTLs and pQTLs were obtained from

Genotype-Tissue Expression (GTEx) European participants and deCODE GWAS on UGT1A1, respectively. The gallstones GWAS was the same as the one used in the genetic correlation analysis and was obtained from the GWAS catalog (accession number GCST90044196).

Pairwise conditional and colocalization (PWCoCo) tests all conditionally independent exposure instruments against all conditionally independent association signals in the outcome to avoid unreliable inference from colocalization due to the presence of multiple conditionally distinct association signals within the same genomic region. We used only SNPs that had a minor allele frequency >1% and that were located within a 500-Mb window from the gene body of the target gene. Finally, the F statistic of our instruments strength was calculated using the Cragg-Donald F statistic formula (17). The PW-MR, sensitivity, and colocalization analyses all used the 1000G LD reference panel in Europeans.

### **Proteome-Wide Mendelian Randomization Study Using Fenland Protein Quantitative Trait Loci**

The Fenland study generated genotype data using Affymetrix UK Biobank Axiom array, Affymetrix SNP 5.0, or Illumina Infinium Core-Exome 24v1, and phenotype data for 12,435 participants of White British ancestry, which resulted in >17 million genotyped or imputed variants with a minor allele frequency of  $\geq 0.1\%$ . The relative concentrations of the 4775 unique proteins from blood samples were measured from 4979 slow off-rate modified aptamers (SOMAmers) by SomaLogic Inc using the V4 platform by an aptamer-based approach, which only included human protein targets and SOMAmers deemed acceptable by SomaLogic and that had a value between 0.25 and 4 for all scaling factors. The cis-pQTLs were computed from an additive model using BGENIE 1.3 (<https://jmarchini.org/software/>), and the results were then combined for the 3 genotyping arrays using a fixed-effects meta-analysis in METAL. A pQTL was considered to be cis-acting if it was located within  $\pm 500$  kilobase of the gene body of the protein-encoding target gene.

The 4979 SOMAmers were used to perform the association study. For each SOMAmer, we used all the SNPs located within  $\pm 1$  Mb of the gene body of the target gene. From these SNPs, only those that were genome-wide significant ( $P < 5.0 \times 10^{-8}$ ) and considered independent ( $LD R^2 < 0.1$  in 1000G) were retained. We then used the same analytical strategy as described for the deCODE data set to identify proteins linked with AP.

### **Reverse Mendelian Randomization Study Using deCODE Protein Quantitative Trait Loci**

We used each significant aptamer from the forward MR analysis using deCODE and Fenland pQTL data as outcomes and AP as exposure. From our AP meta-analysis, we only retained SNPs that were genome-wide significant ( $P < 5.0 \times 10^{-8}$ ) and considered independent ( $LD R^2 < 0.1$ ). We then performed IVW MR analyses for AP on each protein individually using the TwoSampleMR package (16) in R. Sensitivity analyses were performed using Cochran's Q test and Egger's regression intercept. Colocalization analyses were performed using the coloc (18) software.

### **Post Genome-Wide Association Study Analyses**

We used the FUMA (19) web application programming interface 1.3.8 to perform MAGMA (v1.08) tissue expression analysis using the SNP2GENE tool. MAGMA tests relationships between tissue-specific gene expression and disease-gene associations from gene-property analyses. For this analysis, we only used GWAS variants with  $P < 5 \times 10^{-5}$  and data from 54 GTEx v8 tissues. We also performed per-data set cell type specificity on these same variants using Cell Type tool and the GSE81547 human pancreas single-cell transcriptomic data (20).

We then used the GENE2FUNC tool to investigate prioritized genes and test for enrichment of the genes coding for the 68 unique proteins identified by the PW-MR approach. An enrichment was considered significant if it passed the 5% FDR  $P$ -value threshold and if at least 2 genes overlapped with the gene sets. We investigated the differentially expressed genes in each of the 30 GTEx v8 general tissues vs all the remaining tissues, and the results are presented as the  $\log_2$ -transformed average expression of the 68 unique protein-coding genes in each tissue. For the GWAS hits and top 10 biologically relevant protein-coding genes identified by the PW-MR approach, we investigated the tissue-specific expression using the transcript per million data for each protein-coding gene in the tissue of interest. Using European participants data only, we calculated the median transcript per million for each gene in the tissue of interest and calculated the mean transcript per million for each gene across all 54 tissues. We then calculated the  $\log_2$  fold-change expression for each gene in the tissue of interest.

### **Drug Targets Analysis**

The drug targets analysis was performed using the OpenTargets (21) publicly available data (v.22.11). We selected all drugs for which there was evidence for an association with the protein of interest. We

only retained the tissue-specific expression and protein levels, cell type specificity, and adverse effects for which the metric (level or likelihood ratio) was in the top 5% value and >0.

### Code and Data Availability

The GWAS meta-analysis was performed using the METAL 2011-03-25 software (<http://csg.sph.umich.edu/abecasis/metal>), and the LD score regression analysis was performed using the LDSC 1.0.1 tool (<https://github.com/bulik/ldsc>). The GWAS summary statistics of blood pQTLs from deCODE can be found at <https://www.decode.com/summarydata/>, and those of the Fenland study can be found at <https://omicscience.org/apps/pgwas/>. Statistical analyses were performed in R 4.0.0 using publicly accessible functions from the SAIGE 0.44 (<https://github.com/weizhouUMICH/SAIGE>), the TwoSampleMR 0.5.6 (<https://github.com/MRCIEU/TwoSampleMR>), the MendelianRandomization 0.5.1 (<https://cran.r-project.org/web/packages/MendelianRandomization>), the data.table 1.14.0 (<https://github.com/Rdatatable/data.table>), the JAM-MR (<https://github.com/pjnewcombe/R2BGLiMS/>), the coloc (<https://cran.r-project.org/web/packages/coloc/>), the HyPrColoc (<https://github.com/jrs95/hyprcoloc>), and the PWCoCo (<https://github.com/jwr-git/pwcoco>) packages in R. Finally, the OpenTargets 22.11 data were downloaded directly from the OpenTargets website (<https://platform.opentargets.org/downloads/data>).

## 1.5. Results

### Meta-analysis of the Genome-Wide Association Studies for Acute Pancreatitis

To identify genetic variants associated with AP, we first performed 2 new GWASs in the Estonian Biobank and UK Biobank data sets. Subsequently, we performed a meta-analysis of 3 GWASs (adding FinnGen) for a total of 10,630 AP cases and 844,679 controls all identified through electronic health records (Graphical Abstract). We identified 5 genetic risk loci that harbored at least 1 SNP that passed the genome-wide significance threshold of  $P \leq 5 \times 10^{-8}$  (*ABCG5/8*, *TWIST2*, *SPINK1*, *PRSS2*, and *MORC4*) (Figure 1). Among these loci, *TWIST2* and *MORC4* were identified as possible novel susceptibility loci for AP. The associated quantile-quantile plot is presented in Supplementary Figure 1. The genomic inflation factor ( $\lambda$ ) was 1.051. The LD score regression intercept was 1.0123 (SE = 0.0067), and the SNP-based heritability ( $h^2$ ) was 0.0034 (SE = 0.0006), suggesting no genomic inflation. These values were obtained using LDSC (13,22). The associations and assessment of heterogeneity of SNPs that passed the genome-wide P-value threshold at these loci with AP are

presented in Supplementary Table 1A and B. Because there was some evidence of heterogeneity, the  $P$  values are presented from the random-effects meta-analysis. The LocusZoom plots of the lead SNPs at *ABCG5/8*, *TWIST2*, *SPINK1*, *PRSS2*, and *MORC4*, their genomic location, GWAS  $P$  values, and recombination rate with surrounding SNPs are shown in Supplementary Figure 2. Altogether, this GWAS meta-analysis identified genetic variation at 3 loci known to be associated with AP and possibly 2 novel loci and provides a robust data set for MR analyses.

### **Cross-Trait and Genetic Correlation Analysis of Acute Pancreatitis**

Cross-phenotype genetic association analyses were performed for AP genome-wide significant SNPs ( $P \leq 5 \times 10^{-8}$ ) using the iCPAGdb (15). The iCPAGdb provides enrichment and similarity metrics using ancestry LD-specific association data by integrating genetic data across >3700 traits from the National Human Genome Research Institute–European Bioinformatics Institute GWAS catalog. After Bonferroni's correction, the cross-phenotype analysis revealed that 6 traits were significantly associated with AP (Figure 2A and Supplementary Table 2). The highest enrichments were for phytosterol measurement ( $P = 1.01 \times 10^{-10}$ ), gallstones ( $P = 1.37 \times 10^{-8}$ ), 3-hydroxy-3-methylglutaryl–coenzyme A reductase inhibitor use ( $P = 1.57 \times 10^{-7}$ ), apolipoprotein B measurement ( $P = 1.14 \times 10^{-6}$ ), total cholesterol measurement ( $P = 2.90 \times 10^{-6}$ ), and low-density lipoprotein cholesterol measurement ( $P = 5.90 \times 10^{-6}$ ). Genetic correlation analyses were then performed using GWAS summary statistics of 56 traits downloaded from the GWAS catalog (23). We used the LDSC software (13,22) to calculate the SNP-based heritability and genetic correlation with AP for each trait (Figure 2B and Supplementary Table 3A). After correcting for multiple testing, 15 traits passed the 5% FDR  $P$ -value threshold, including gallstones ( $\rho = 0.84$ ;  $P = 7.44 \times 10^{-12}$ ), coronary artery disease ( $\rho = 0.38$ ;  $P = 7.24 \times 10^{-11}$ ), antithrombotic use ( $\rho = 0.51$ ;  $P = 5.04 \times 10^{-9}$ ), nonalcoholic fatty liver disease ( $\rho = 0.91$ ;  $P = 3.00 \times 10^{-4}$ ), and total high-density lipoprotein cholesterol ( $\rho = -0.16$ ;  $P = 9.39 \times 10^{-3}$ ). To further support the potential causal relationship between the 56 traits and AP, we performed MR. After correcting for multiple testing (5% FDR), we identified 31 significant associations between potential risk factors and AP, among which 25 were phytosterol-related (Supplementary Table 3B). Significant associations also included gallstones ( $\beta = 0.34$ ;  $P = 2.14 \times 10^{-10}$ ) and total esterified cholesterol ( $\beta = -0.12$ ;  $P = 4.95 \times 10^{-3}$ ).

### **Proteome-Wide Mendelian Randomization Studies of Acute Pancreatitis**

We investigated the associations of 4907 and 4979 plasma protein levels with AP using genetic association summary statistics from 35,559 Icelanders included in deCODE Genetics (24) and 10,708

Europeans included in Fenland (25), respectively. Our genetic instrument selection strategy allowed for the investigation of 1763 and 1735 proteins from the deCODE and Fenland studies, respectively. The PW-MR approach using cis-acting variants (Methods) identified 68 unique plasma proteins associated with AP, after correcting for multiple testing using a 5% FDR. A total of 46 proteins were identified in deCODE (Figure 3A), and 44 proteins were identified in Fenland (Figure 3B). The results of multiple sensitivity analyses, including MR-Egger (26), weighted median, MR-PRESSO (Pleiotropy RESidual Sum and Outlier) (27), and contamination mixture (28), showed little evidence of pleiotropy. JAM-MR (29) and principal components analysis (30) methods were further applied to assess possible correlation between instrumental variables (Supplementary Table 4A and B). The F statistics of the genetic instruments for each protein showed solid instrument strength (Supplementary Table 4C and D). This analysis yielded slightly more results with deCODE than with Fenland, including 2 more proteins that have a potential protective effect and the same number that have a deleterious effect on AP. After FDR correction, 19 proteins were identified in both data sets (Figure 3C). An additional 10 proteins that passed FDR correction in 1 data set were nominally associated with AP ( $P < .05$ ) in the second data set, suggesting cross-validation for 29 proteins (Supplementary Table 4E). A PhenoGram representing the chromosomal location of the 68 unique identified proteins as well as that of the susceptibility loci identified by GWAS meta-analysis is depicted in Figure 3D. We also used coloc software (18) to assess genetic colocalization of each protein with AP. UGT1A1 showed strong evidence of colocalization (posterior probability of hypothesis 4 [PPH4] of 1 causal variant shared between both traits) in deCODE (PPH4 = 0.93) and Fenland (PPH4 = 0.89). This analysis identified 1 causal variant in deCODE (rs6742078) and 1 causal variant in Fenland (rs887829). These variants are in LD with one another ( $R^2 = 1$ ) in the European population (1000G). To account for possible confounding due to LD between neighboring signals, we applied PWCoCo (31) on each of the significant proteins identified. In both cohorts, UGT1A1 (PPH4<sub>deCODE</sub> = 0.897; PPH4<sub>Fenland</sub> = 0.885) and CTRB2 (PPH4<sub>deCODE</sub> = 0.977; PPH4<sub>Fenland</sub> = 0.974) showed strong evidence of colocalization (Supplementary Table 4F and G).

Because UGT1A1 is a liver-expressed gene, we sought to further investigate shared genetic etiology across traits involving UGT1A1 using HyPrColoc (32), a Bayesian method used to detect colocalization across multiple traits simultaneously. We investigated colocalization across 4 traits: AP, UGT1A1 eQTLs in the liver, UGT1A1 pQTLs and gallstones, highlighting the SNP (rs6742078) showing the

highest PPH4 in deCODE and providing evidence for a potential role of liver-expressed UGT1A1 in the etiology of both gallstones and AP (Figure 4).

### **Reverse Mendelian Randomization Studies of Acute Pancreatitis Using deCODE**

We performed a “reverse” PW-MR analysis using genetically predicted AP as the exposure and the 4907 and 4979 proteins from deCODE and Fenland, respectively, to determine whether AP could causally influence variation in blood levels of these proteins. None of the blood proteins from deCODE or Fenland appeared to be causally influenced by genetically predicted AP (at the 5% FDR *P*-value threshold) (Supplementary Table 4H and I), suggesting that the 68 proteins may represent new blood proteins causally linked with AP rather than consequences of AP.

### **Gene-Based Association and Pathway Analyses**

To gain insight on the biological relevance of the previous findings, we investigated the functional annotations of the genome-wide significant genetic variants as well as of the genes encoding proteins linked with AP using the MAGMA tool of the FUMA web application programming interface (19). These genes appeared to be significantly enriched in metabolic pathways, such as carbohydrate metabolism, and digestion and absorption (Figure 5A). The significant SNPs ( $P \leq 5 \times 10^{-5}$ ) from the GWAS are enriched in the acinar pancreatic cell type (20) (Supplementary Figure 3). The 68 unique genes encoding these 71 proteins may be differentially expressed in the esophagus and adipose tissue and upregulated in the liver and pancreas (Figure 5B) and are predominantly expressed in the blood, liver, lymphocytes, and pancreas (Figure 5C and D). This analysis revealed that an important proportion of genes harboring genome-wide significant SNPs as well as blood proteins associated with AP (pancreatic colipase [CLPS], chromotrypsinogen B2 [CTRB2], alpha-amylase 2A and 2B [AMY2A and AMY2B], Kirre-like nephrin family adhesion molecule 2 [KIRREL2], and glutathione s-transferase alpha 1 and zeta 1 [GSTA1 and GSTZ1]) appeared to be predominantly expressed in the pancreas (33) (Supplementary Table 5). The liver is also another organ inside which genes harboring genome-wide significant SNPs (*ABCG5/8*) and blood proteins associated with AP (angiopoietin-like protein 3 [ANGPTL3], lipopolysaccharide-binding protein [LBP], Klotho beta [KLB], complement factor H related 2 and 5 [CFHR2/5], and serine hydroxymethyltransferase 1 [SHMT1]) may be more predominantly expressed.

## Druggable Genome Analyses

To leverage this genetic information into new potential therapeutic avenues for AP, we investigated the potential druggability of genes and proteins identified in the previous analyses using OpenTargets (21). We identified 16 unique drugs targeting 10 unique proteins (Supplementary Table 6). Among these 10 unique proteins identified, those for which the corresponding drugs effect are in the same direction as the PW-MR results are ANGPTL3 (vupanorsen/evinacumab), KLB (pegbelfermin), R-spondin 3 (RSP03) (rosmantuzumab), NQ01 (vatiquinone), PDCD1LG2 (rHIgM12B7), and AMY2A (acarbose). For each protein, we included tissue-specific expression level(s), tissue-specific protein level(s), and cell type specificity. For each drug, we included the disease(s) targeted, the mechanism of action, and the potential adverse effects.

## 1.6. Discussion

To gain insight on the genetic architecture of AP, we performed new GWAS in the UK Biobank and the Estonian Biobank and combined these results with those available from the publicly available FinnGen GWAS. Our meta-analysis of 10,630 AP cases and 871,679 controls identified 3 known<sup>4</sup> (*ABCG5/G8*, *SPINK1* and *PRSS2*) and 2 possibly novel (*TWIST2* and *MORC4*) AP susceptibility loci. This GWAS is the largest genetic association study on AP to our knowledge. *ABCG5/G8* (ATP binding cassette subfamily G member 8) encodes a transmembrane protein involved in sterol and cholesterol excretion into bile and sterol transport into the intestinal lumen (34,35). *ABCG8* is also known to be involved in cholelithiasis (36) (gallstone disease) and sitosterolemia-1 (34). *SPINK1*, serine protease inhibitor Kazal-type 1, codes for a secreted protein, which, in the pancreas, protects against premature activation of zymogens catalyzed by trypsin (37). *SPINK1* is known to be involved in chronic and hereditary pancreatitis (38,39).

Our analysis revealed several variants associated with AP at 7q34. Two genes in at this locus have previously been linked with AP, transient receptor potential cation channel subfamily V member 6 (*TRPV6*) and serine protease 2 (*PRSS2*). *TRPV6* codes for a selective calcium channel that is involved in Ca<sup>2+</sup> ion homeostasis by mediating Ca<sup>2+</sup> uptake in various tissues, including the gastrointestinal tract, pancreas, and salivary gland (40). Functional defects of *TRPV6* have previously been associated with chronic pancreatitis (41). *PRSS2* codes for an anionic trypsinogen that stimulates trypsin-mediated degradation of anionic trypsinogen (42,43). *TWIST2* is a transcription factor regulating the epithelial-

mesenchymal transition (44). Recently, *TWIST2* has been shown to regulate the fibroblast growth factor 21 and the 5' adenosine monophosphate-activated protein kinase/mechanistic target of rapamycin pathway. It has also been found to induce oxidative stress in lung cancer cells (45). Genetic association studies with a larger sample size will be required to determine whether these 2 genes represent causes of AP and to identify new susceptibility loci for AP.

A PW-MR study revealed several blood proteins that may be causally related to AP and represent potential new therapeutic targets for the prevention or treatment of AP. Using cis-acting pQTLs on thousands of proteins measured in deCODE (24) and Fenland, 68 proteins significantly associated with AP after FDR correction, many of which play key roles in biological processes that can lead to AP, were identified. The strongest association was with pancreatic amylase alpha 2B (*AMY2B*), an enzyme that catalyzes the endohydrolysis of 1,4- $\alpha$ -glucoside bonds in oligosaccharides and polysaccharides, therefore catalyzing an early step in digestion of dietary starch and glycogen. *AMY2B* is expressed in the pancreas and salivary glands (33) and is an important paralog of *AMY2A*, which was also significantly associated with AP. We also identified an association with the salivary amylase alpha 1A (*AMY1A*). *AMY1* is very similar to *AMY2*, both in sequence and size and location of the exons (16). Chymotrypsinogen B2 (*CTRB2*), which is the principal precursor of pancreatic proteolytic enzyme, is specifically synthesized in the acinar cells of the pancreas and activated in the small intestine. Genetic variation at this locus has been associated with chronic pancreatitis and pancreatic cancer (4,46,47). Pancreatic colipase (*CLPS*) is activated by trypsin in the intestinal lumen and is required for optimal enzyme activity of pancreatic lipase. Pancreatic lipase plays a key role in digestion through the hydrolysis of dietary triglycerides into monoglycerides and free-fatty acids. *UGT1A1* and *UGT1A6* are part of the uridine diphosphate glycosyltransferase (UGT) family, which catalyze glucuronidation, a process involved in the biotransformation of endobiotics and xenobiotics. *UGT1A1* detoxifies bilirubin by conjugating it with glucuronic acid and is involved in the detoxification and excretion of endogenous and exogenous lipophilic compounds in the liver and gastrointestinal tract (48). Motilin (*MLN*) is a small peptide hormone that plays an important role in the regulation of interdigestive gastrointestinal motility. High circulating levels have been associated with type 2 diabetes in individuals after AP (49). There is also evidence that low motilin release secondary to pancreatic exocrine deficiency can lead to disordered gastrointestinal motor activity in patients with chronic pancreatitis (50).

Reduced nicotinamide adenine dinucleotide phosphate (NAD[P]H) quinone dehydrogenase 1 (*NQO1*) catalyzes the oxidation of NAD(P)H to NAD(P)<sup>+</sup> by various quinones, thereby preventing the production of radical species by the 1 electron reduction of quinones. NQO1 has been shown to have a protective effect on AP through the regulation of NAD(P)H oxidase (51). The genes encoding for the cell-surface glycoproteins R-spondin 3 (*RSP03*) and signal peptide, CUB domain and epidermal growth factor–like domain-containing 1 (*SCUBE1*), which binds to calcium (52), are both mainly expressed in the esophagus (33) and are involved in tissue homeostasis (53) and inflammatory response (54), respectively.

Results of the bidirectional PW-MR analysis provide valuable new information on the causal direction of AP biomarkers. For instance, amylases and colipase are often used as biomarkers of AP and are helpful in the diagnosis of AP. By showing that these proteins are causally associated with AP and that the genetically predicted AP is not causally linked with elevation in the blood levels of these enzymes, our bidirectional analysis suggests that these enzymes do not simply reflect the presence of AP but may even be directly implicated in the pathophysiological processes leading to AP. Many of the proteins identified in the PW-MR study may represent therapeutic targets in AP. For instance, we provide evidence that higher blood levels of angiopoietin-like protein 3 (*ANGPTL3*) was associated with higher AP. *ANGPTL3* is a critical regulator of lipoprotein lipase. Patients with lipoprotein lipase deficiency and subsequent familial chylomicronemia have very high plasma triglyceride levels and a high risk of AP (55). RNA-interfering therapies targeting *ANGPTL3* hepatic gene expression (56), as well as *ANGPTL3*-neutralizing antibodies (57), are currently under investigation for the reduction of blood lipid levels.

On top of identifying *ANGPTL3*, our analysis of the druggable genome with the OpenTargets platform identified other possible therapeutic targets for AP such as *KLB* (which may be influenced by the fibroblast growth factor 21 analogue pegbelfermin), *RSP03* (targeted by the monoclonal antibody rosmantuzumab), *NQO1* (targeted by the oral compound vatiquinone), *PDCD1LG2* (targeted with the anti–programmed cell death 1 ligand 2 monoclonal antibody [rHlgM12B7]), and *AMY2A* (targeted by the compound drug acarbose). Several inflammatory markers also causally linked with AP suggest that the modulation of the interleukins 1, 2, and 6 pathways, as well as adhesion molecules, such as intercellular adhesion molecule 5, may influence the course of AP. Whether agents modulating these pathways could be repurposed for the prevention of AP still needs to be examined in clinical trials. In

support of the validity of the present study findings, functional annotation of these proteins confirmed expression of many proteins in metabolic tissues responsible for digestion and energy metabolism, such as the esophagus, adipose tissue, liver, and pancreas, and a potential role of pancreatic acinar cells as revealed by single-cell sequencing analyses.

This study has limitations. For example, the GWAS and proteome data sets only included individuals of European ancestry. The heritability of AP was also low, which hampers the power to perform cross-trait analyses as well as reverse MR (whereby AP is used as the study exposure instead of the study outcome). Our assessment of the human blood proteome was also based only on summary-level data relying on one technology (SOMAmer-based) to measure blood proteins. Other techniques, such as proximity extension assays (Olink), could also be valuable to identify blood proteins linked to disease traits such as AP. Finally, our analysis only included SNPs with a minor allele frequency of  $\geq 1\%$ . Therefore, rare variants with potentially stronger effects could not be identified using this approach. Rare sequence variants may contribute to AP, as recently shown in an exome-wide sequencing study of AP (41).

## **1.7. Conclusion**

We performed the largest genome-wide association meta-analysis and, to our knowledge, the first large-scale PW-MR study of AP. These new analyses confirmed an important role of digestive tissues, such as the liver and the pancreas, in the etiology of AP and that genetic variation altering digestive pathways may be functionally linked with AP. Our study revealed the complex genetic architecture of AP and identified several new blood proteins that may be causally linked with AP, opening the door to possible therapeutic strategies.

### **Acknowledgments**

The authors would like to thank all study participants as well as all investigators of the studies that were used throughout the course of this investigation (Estonian Biobank, UK Biobank, FinnGen, deCODE, and Fenland).

### **CRedit Authorship Contributions**

Jérôme Bourgault, BSc, MSc (Conceptualization: Equal; Data curation: Lead; Formal analysis: Lead; Writing – original draft: Equal). Erik Abner, PhD (Formal analysis: Supporting; Writing – review &

editing: Supporting). Hasanga D. Manikpurage, PhD (Formal analysis: Supporting; Writing – review & editing: Equal). Natàlia Pujol-Gualdo, PhD (Writing – review & editing: Equal). Triin Laisk, PhD (Writing – review & editing: Equal). Estonian Biobank Research Team (Data acquisition and curation: Supporting). Émilie Gobeil, MSc (Writing – review & editing: Equal). Eloi Gagnon, PhD (Writing – review & editing: Equal). Arnaud Girard, MSc (Writing – review & editing: Equal). Patricia L. Mitchell, PhD (Writing – review & editing: Equal). Sébastien Thériault, MD (Writing – review & editing: Equal). Tõnu Esko, PhD (Writing – review & editing: Equal). Patrick Mathieu, MD (Writing – review & editing: Equal). Benoit J. Arsenault, PhD (Conceptualization: Equal; Project administration: Lead; Writing – original draft: Equal; Writing – review & editing: Equal).

### **Conflicts of interest**

These authors disclose the following: Benoit J. Arsenault is a consultant for Novartis, Editas Medicine, and Silence Therapeutics and has received research contracts from Pfizer, Ionis Pharmaceuticals, and Silence Therapeutics. The remaining authors disclose no conflicts.

### **Funding**

The work of Estonian Genome Center, University of Tartu has been supported by the European Regional Development Fund and grants No. GENTRANSMED (2014-2020.4.01.15-0012), and MOBERA5 (Norface Network project no 462.16.107). The work of the Estonian Genome Center, University of Tartu was also funded by the European Union through Horizon 2020 Research and Innovation Program under grants no. 810645 and 894987, through the European Regional Development Fund projects GENTRANSMED (2014-2020.4.01.15-0012), MOBEC008, and Estonian Research Council Grant PRG1291. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by National Cancer Institute, National Human Genome Research Institute, National Heart, Lung and Blood Institute, National Institute on Drug Abuse, National Institute of Mental Health, and National Institute of Neurological Disorders and Stroke. Jérôme Bourgault and Émilie Gobeil hold a Masters' Research Award from the Canadian Institutes of Health Research. Eloi Gagnon holds a Doctoral Research Award and Benoit J. Arsenault holds a Senior Scholar Award from the Fonds de Recherche du Québec–Santé.

## Data Availability

The data used for the analyses described in this manuscript were obtained from the GTEx Portal on August 4, 2022.

The AP GWAS summary statistics are available at the GWAS catalog (study access number GCST90255375). The code is available at [https://github.com/LaboArsenault/2022\\_PW-MR\\_pipeline](https://github.com/LaboArsenault/2022_PW-MR_pipeline).

## 1.8. References

1. Iannuzzi JP, King JA, Leong JH, et al. Global Incidence of Acute Pancreatitis Is Increasing Over Time: A Systematic Review and Meta-Analysis. *Gastroenterology*. 2022;162(1):122-134. doi:10.1053/j.gastro.2021.09.043
2. Ingraham NE, King S, Proper J, et al. Morbidity and Mortality Trends of Pancreatitis: An Observational Study. *Surg Infect*. 2021;22(10):1021-1030. doi:10.1089/sur.2020.473
3. Mayerle J, Sendler M, Hegyi E, Beyer G, Lerch MM, Sahin-Tóth M. Genetics, Cell Biology, and Pathophysiology of Pancreatitis. *Gastroenterology*. 2019;156(7):1951-1968.e1. doi:10.1053/j.gastro.2018.11.081
4. Rosendahl J, Kirsten H, Hegyi E, et al. Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut*. 2018;67(10):1855-1863. doi:10.1136/gutjnl-2017-314454
5. Mayer J, Rau B, Gansauge F, Beger HG. Inflammatory mediators in human acute pancreatitis: clinical and pathophysiological implications. *Gut*. 2000;47(4):546-552. doi:10.1136/gut.47.4.546
6. Zhang D, Wang T, Dong X, et al. Systemic Immune-Inflammation Index for Predicting the Prognosis of Critically Ill Patients with Acute Pancreatitis. *Int J Gen Med*. 2021;14:4491-4498. doi:10.2147/IJGM.S314393
7. Folkersen L, Gustafsson S, Wang Q, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab*. 2020;2(10):1135-1148. doi:10.1038/s42255-020-00287-2
8. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021;374(6569):eabj1541. doi:10.1126/science.abj1541
9. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335-1341. doi:10.1038/s41588-018-0184-y
10. Mitt M, Kals M, Pärn K, et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur J Hum Genet*. 2017;25(7):869-876. doi:10.1038/ejhg.2017.51
11. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191. doi:10.1093/bioinformatics/btq340
12. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw*. 2018;3(25):731. doi:10.21105/joss.00731
13. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211

14. Staley J. gassocplot. Published online June 17, 2021. Accessed February 1, 2022. <https://github.com/jrs95/gassocplot>
15. Wang L, Balmat TJ, Antonia AL, et al. An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome Med.* 2021;13(1):83. doi:10.1186/s13073-021-00904-z
16. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Loos R, ed. *eLife.* 2018;7:e34408. doi:10.7554/eLife.34408
17. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol.* 2011;40(3):755-764. doi:10.1093/ije/dyr036
18. Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLOS Genet.* 2020;16(4):e1008720. doi:10.1371/journal.pgen.1008720
19. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
20. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, Kim SK, Quake SR. Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell.* 2017 Oct 5;171(2):321-330.e14. doi: 10.1016/j.cell.2017.09.004. Epub 2017 Sep 28. PMID: 28965763; P22.
21. Ochoa D, Hercules A, Carmona M, et al. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Res.* Published online November 18, 2022: gkac1046. doi:10.1093/nar/gkac1046
22. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241. doi:10.1038/ng.3406
23. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012. doi:10.1093/nar/gky1120
24. Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021;53(12):1712-1721. doi:10.1038/s41588-021-00978-w
25. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun.* 2021;12(1):6822. doi:10.1038/s41467-021-27164-0
26. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol.* 2017;32(5):377-389. doi:10.1007/s10654-017-0255-x
27. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet.* 2018;50(5):693-698. doi:10.1038/s41588-018-0099-7
28. Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nat Commun.* 2020;11(1):376. doi:10.1038/s41467-019-14156-4
29. Newcombe PJ, Conti DV, Richardson S. JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. *Genet Epidemiol.* 2016;40(3):188-201. doi:10.1002/gepi.21953
30. Burgess S, Zuber V, Valdes-Marquez E, Sun BB, Hopewell JC. Mendelian randomization with fine-mapped genetic data: Choosing from large numbers of correlated instrumental variables. *Genet Epidemiol.* 2017;41(8):714-725. doi:10.1002/gepi.22077
31. Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020;52(10):1122-1131. doi:10.1038/s41588-020-0682-6MCID: PMC6047899.

32. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun.* 2021;12(1):764. doi:10.1038/s41467-020-20885-8
33. Consortium TGte, Ardlie KG, Deluca DS, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* Published online May 8, 2015. doi:10.1126/science.1262110
34. Berge KE, Tian H, Graf GA, et al. Accumulation of Dietary Cholesterol in Sitosterolemia Caused by Mutations in Adjacent ABC Transporters. *Science.* Published online December 1, 2000. doi:10.1126/science.290.5497.1771
35. Lu K, Lee MH, Hazard S, et al. Two Genes That Map to the STSL Locus Cause Sitosterolemia: Genomic Structure and Spectrum of Mutations Involving Sterolin-1 and Sterolin-2, Encoded by ABCG5 and ABCG8, Respectively. *Am J Hum Genet.* 2001;69(2):278-290. doi:10.1086/321294
36. Buch S, Schafmayer C, Völzke H, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet.* 2007;39(8):995-999. doi:10.1038/ng2101
37. Serine Peptidase Inhibitor Kazal Type - an overview | ScienceDirect Topics. Accessed May 5, 2022. <https://www.sciencedirect.com/topics/medicine-and-dentistry/serine-peptidase-inhibitor-kazal-type>
38. Chen JM, Mercier B, Audrezet MP, Ferec C. Mutational analysis of the human pancreatic secretory trypsin inhibitor (PSTI) gene in hereditary and sporadic chronic pancreatitis. *J Med Genet.* 2000;37(1):67-69. doi:10.1136/jmg.37.1.67
39. Witt H, Luck W, Hennies HC, et al. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat Genet.* 2000;25(2):213-216. doi:10.1038/76088
40. Peng JB, Suzuki Y, Gyimesi G, Hediger MA. TRPV5 and TRPV6 Calcium-Selective Channels. In: Kozak JA, Putney JW, eds. *Calcium Entry Channels in Non-Excitable Cells.* CRC Press/Taylor & Francis; 2018. Accessed January 17, 2022. <http://www.ncbi.nlm.nih.gov/books/NBK531440/>
41. Masamune A, Kotani H, Sörgel FL, et al. Variants That Affect Function of Calcium Channel TRPV6 Are Associated With Early-Onset Chronic Pancreatitis. *Gastroenterology.* 2020;158(6):1626-1641.e8. doi:10.1053/j.gastro.2020.01.005
42. Zhou J, Sahin-Tóth M. Chymotrypsin C (CTRC) mutations in chronic pancreatitis. *J Gastroenterol Hepatol.* 2011;26(8):1238-1246. doi:10.1111/j.1440-1746.2011.06791.x
43. Witt H, Sahin-Tóth M, Landt O, et al. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. *Nat Genet.* 2006;38(6):668-673. doi:10.1038/ng1797
44. Fang X, Cai Y, Liu J, et al. Twist2 contributes to breast cancer progression by promoting an epithelial-mesenchymal transition and cancer stem-like cell self-renewal. *Oncogene.* 2011;30(47):4707-4720. doi:10.1038/onc.2011.181
45. Song Y, Zhang W, Zhang J, et al. TWIST2 inhibits EMT and induces oxidative stress in lung cancer cells by regulating the FGF21-mediated AMPK/mTOR pathway. *Exp Cell Res.* 2021;405(1):112661. doi:10.1016/j.yexcr.2021.112661
46. Jermusyk A, Zhong J, Connelly KE, et al. A 584 bp deletion in CTRB2 inhibits chymotrypsin B2 activity and secretion and confers risk of pancreatic cancer. *Am J Hum Genet.* 2021;108(10):1852-1865. doi:10.1016/j.ajhg.2021.09.002
47. Seltsam K, Pentner C, Weigl F, et al. Sequencing of the complex CTRB1-CTRB2 locus in chronic pancreatitis. *Pancreatol.* 2020;20(8):1598-1603. doi:10.1016/j.pan.2020.09.017

48. Sugatani J. Function, genetic polymorphism, and transcriptional regulation of human UDP-glucuronosyltransferase (UGT) 1A1. *Drug Metab Pharmacokinet.* 2013;28(2):83-92. doi:10.2133/dmpk.dmpk-12-rv-096
49. Gold-Smith FD, Singh RG, Petrov MS. Elevated Circulating Levels of Motilin are Associated with Diabetes in Individuals after Acute Pancreatitis. *Exp Clin Endocrinol Diabetes.* 2020;128(01):43-51. doi:10.1055/a-0859-7168
50. OWYANG C, FUNAKOSHI A, VINIK AI. Evidence for Modulation of Motilin Secretion by Pancreatico-Biliary Juice in Health and in Chronic Pancreatitis\*. *J Clin Endocrinol Metab.* 1983;57(5):1015-1020. doi:10.1210/jcem-57-5-1015
51. Shen A, Kim HJ, Oh GS, et al. Pharmacological stimulation of NQO1 decreases NADPH levels and ameliorates acute pancreatitis in mice. *Cell Death Dis.* 2018;10(1):5. doi:10.1038/s41419-018-1252-z
52. Tu CF, Yan YT, Wu SY, et al. Domain and Functional Analysis of a Novel Platelet-Endothelial Cell Surface Protein, SCUBE1\*. *J Biol Chem.* 2008;283(18):12478-12488. doi:10.1074/jbc.M705872200
53. Nagano K. R-spondin signaling as a pivotal regulator of tissue development and homeostasis. *Jpn Dent Sci Rev.* 2019;55(1):80-87. doi:10.1016/j.jdsr.2019.03.001
54. Yang RB, Ng CKD, Wasserman SM, et al. Identification of a novel family of cell-surface proteins expressed in human vascular endothelium. *J Biol Chem.* 2002;277(48):46364-46373. doi:10.1074/jbc.M207410200
55. O'Dea LSL, MacDougall J, Alexander VJ, et al. Differentiating Familial Chylomicronemia Syndrome From Multifactorial Severe Hypertriglyceridemia by Clinical Profiles. *J Endocr Soc.* 2019;3(12):2397-2410. doi:10.1210/js.2019-00214
56. Volanesorsen and Triglyceride Levels in Familial Chylomicronemia Syndrome | NEJM. Accessed July 22, 2022. [https://www.nejm.org/doi/10.1056/NEJMoa1715944?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%20pubmed](https://www.nejm.org/doi/10.1056/NEJMoa1715944?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed)
57. Ahmad Z, Banerjee P, Hamon S, et al. Inhibition of Angiopoietin-Like Protein 3 With a Monoclonal Antibody Reduces Triglycerides in Hypertriglyceridemia. *Circulation.* 2019;140(6):470-486. doi:10.1161/CIRCULATIONAHA.118.039107

## 1.9. Figure Legend

**Graphical abstract.** Overview of the study design. Study design summary identifying cohorts and methods applied. The ultimate endpoint of the study is to identify new genetic variants, blood proteins and therapeutic targets for acute pancreatitis.

**Figure 1.** Manhattan plot of acute pancreatitis genome-wide association study meta-analysis. Genetic loci harboring top SNPs associated with acute pancreatitis ( $p < 5.0 \times 10^{-8}$ ) are shown. The red dashed line represents the genome-wide significance threshold of  $5.0 \times 10^{-8}$ .

**Figure 2.** Cross-trait analysis and genetic correlation of acute pancreatitis with other traits. A) Significant cross-phenotype associations with AP (5% FDR). B) Lollipop plot of genetic correlation between AP and 58 traits. Nominal (0.05), Bonferroni and 5% FDR thresholds are indicated by a full black, blue dashed and red dotted line, respectively. The size of the dots represents the absolute magnitude of the genetic correlation and the red and blue color indicate a negative and positive correlation, respectively. The y-axis represent the p-value of the correlation.

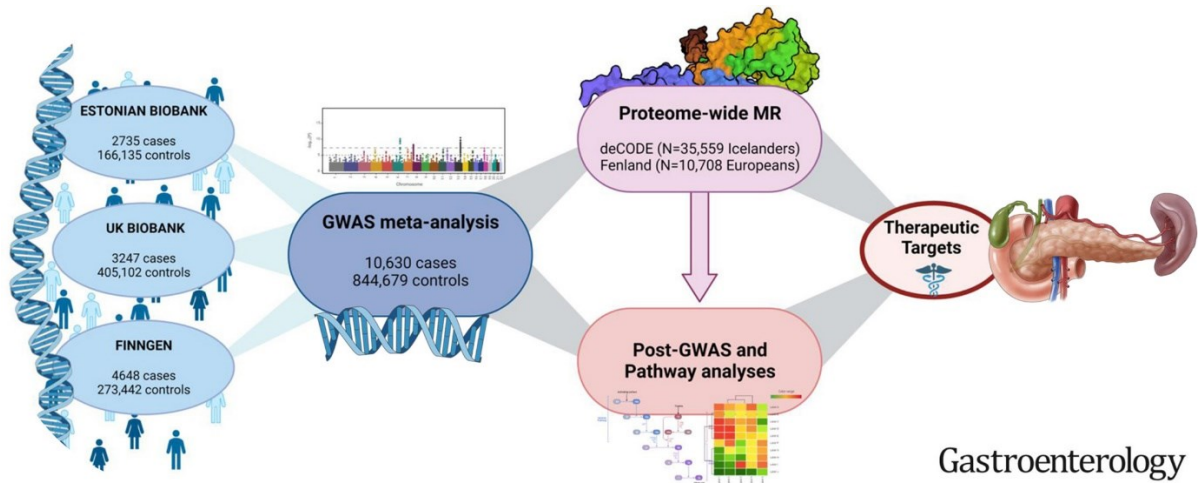
**Figure 3.** Proteome-wide Mendelian randomization studies of acute pancreatitis. A) Volcano plot of acute pancreatitis proteome-wide Mendelian randomization study using A) deCODE and B) Fenland data. Association strength between proteins and acute pancreatitis according to their effect. Annotated proteins passed the 5% FDR inverse-variance weighted p-value threshold. The green and red colors represent a negative and positive effect, respectively. C) Venn diagram depicting proteins associated with AP in Fenland only, in deCODE only or in both. D) PhenoGram of acute pancreatitis genome-wide association study meta-analysis and proteome-wide mendelian randomization study significant associations. Blue dots represent the top genome-wide significant loci, the green dots represent the PW-MR significant proteins and the red dot represents both simultaneously.

**Figure 4.** HyPrColoc plot of UGT1A1 variants. HyPrColoc stacked association plots and prior sensitivity heatmap for UGT1A1 association with AP, its blood protein levels and liver expression and gallstones.

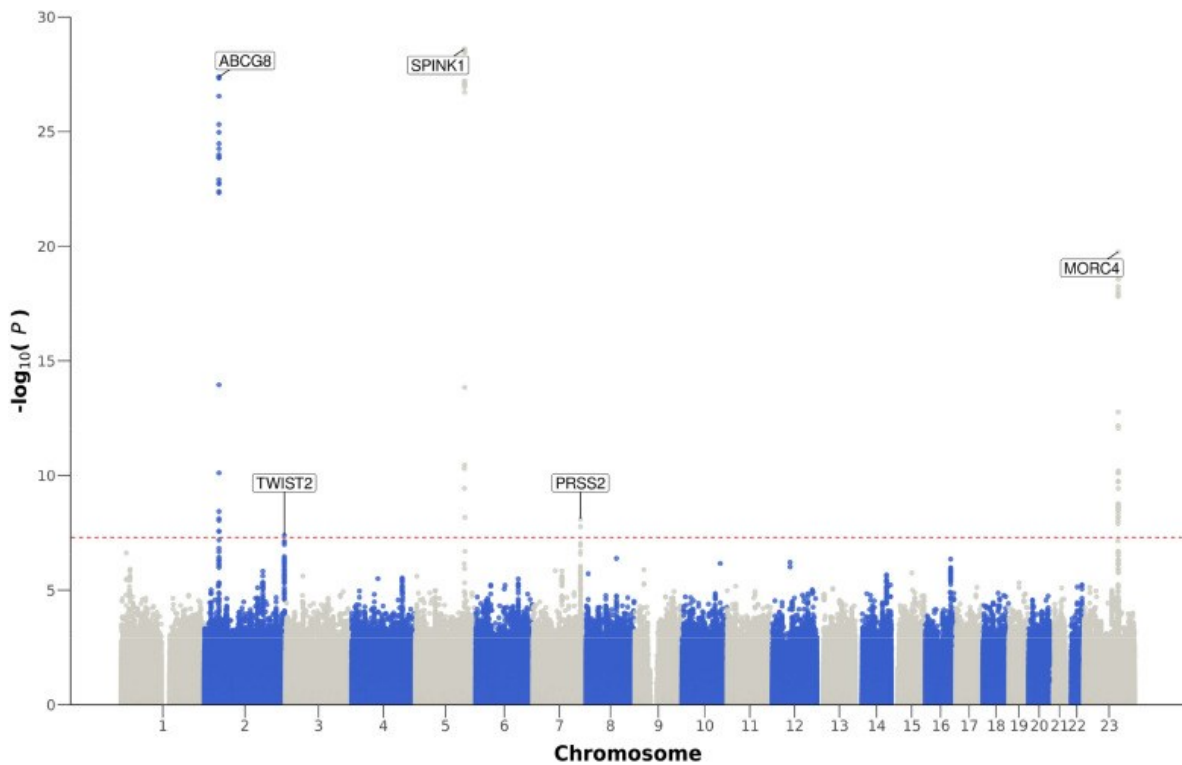
**Figure 5.** Functional annotation of the genetic architecture of AP. A) Pathway analysis of the 68 unique proteins identified in deCODE and Fenland. Pathways and functions involving the proteins identified by the PW-MR approach in deCODE and Fenland are presented. An enrichment was considered significant if it passed the 5% FDR p-value threshold and if at least 2 genes overlapped with the gene

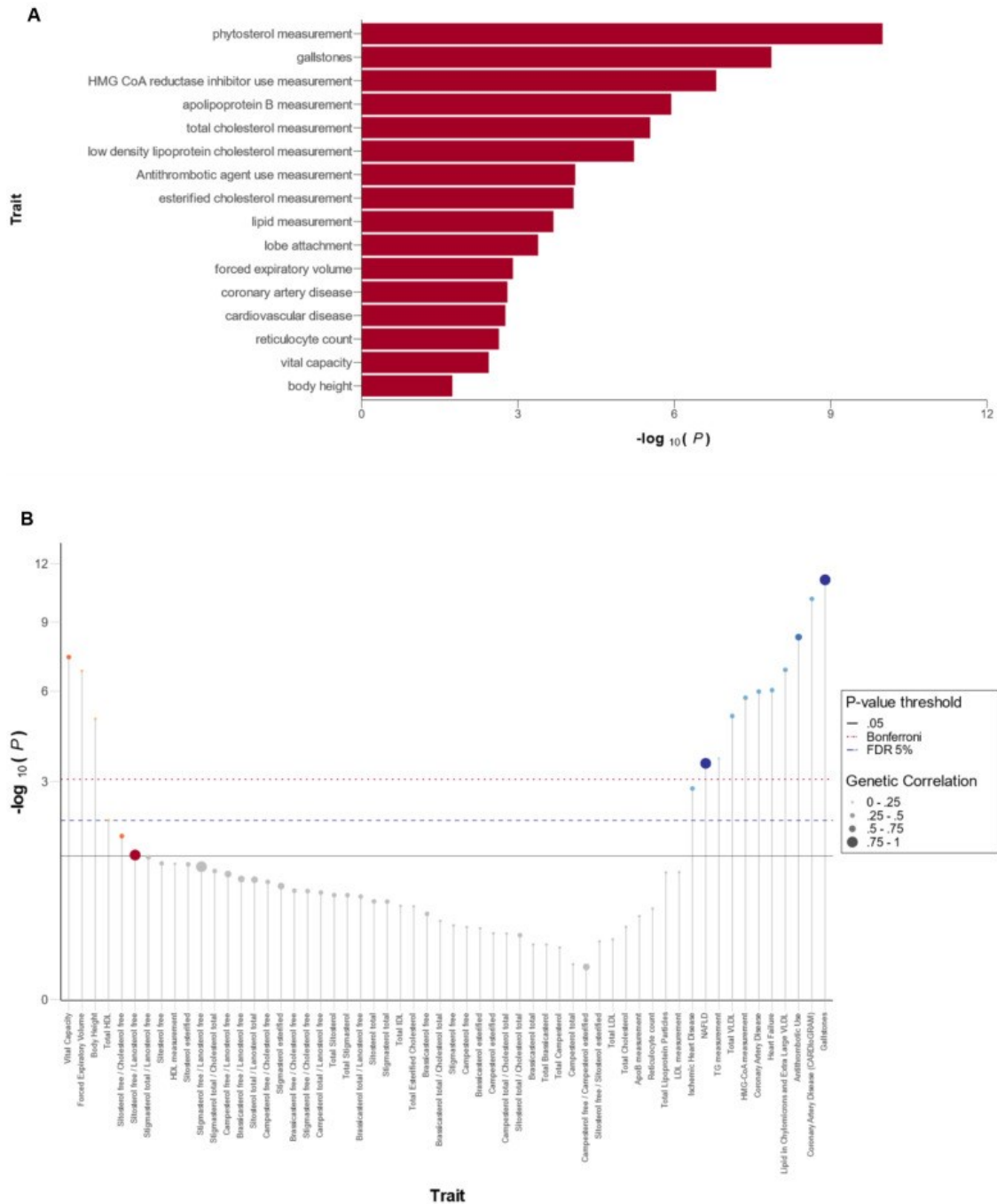
sets. B) Differential expression analysis of the top genome-wide significant loci and protein-coding genes identified in deCODE and Fenland. Differential expression of significant genes identified by GWAS and PW-MR in 30 GTEx v8 tissues. Red bars represent tissues in which the gene set is differentially expressed. C) Gene expression heatmap of the top genome-wide significant loci. Log<sub>2</sub>-transformed average expression in 54 GTEx v8 tissues. D) Gene expression heatmap of the 68 unique PW-MR significant proteins in deCODE and Fenland. Log<sub>2</sub>-transformed average expression in 54 GTEx v8 tissues.

## 1.10. Figures

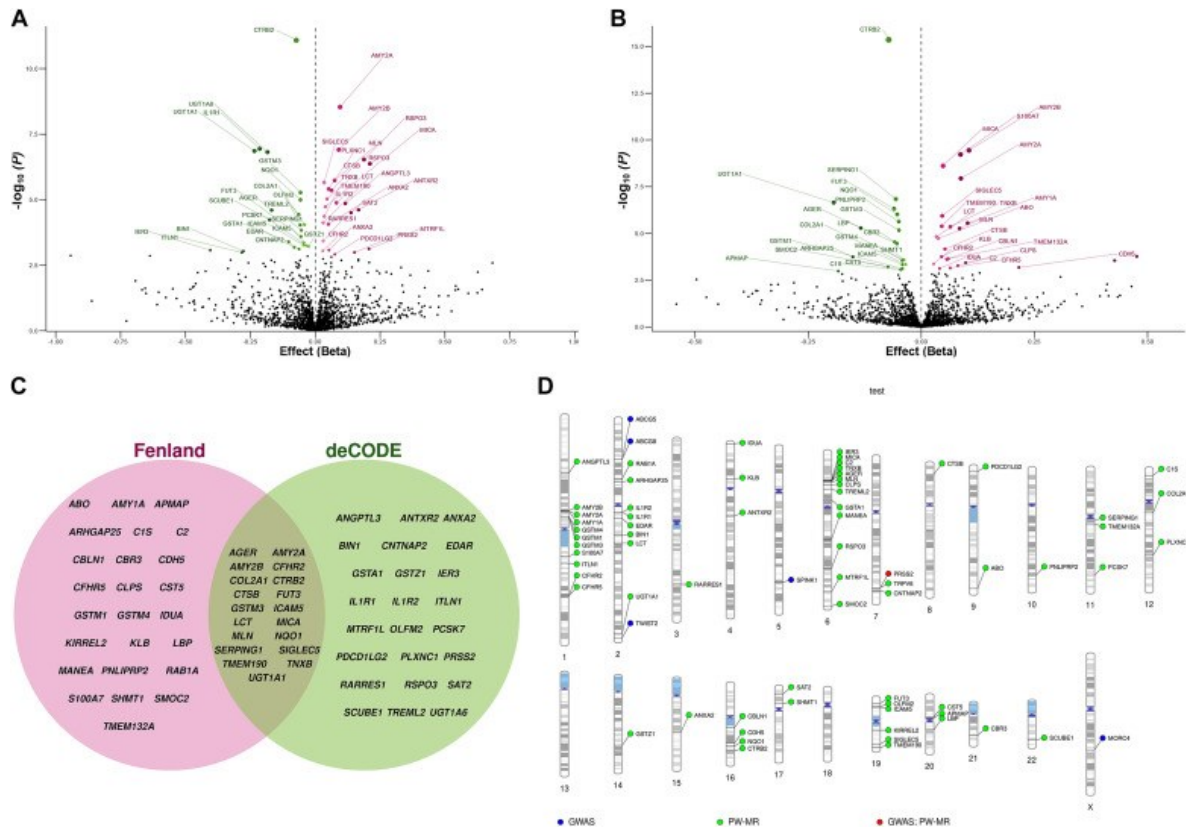


**Graphical abstract 1. Overview of the study design.** Study design summary identifying cohorts and methods applied. The ultimate endpoint of the study is to identify new genetic variants, blood proteins and therapeutic targets for acute pancreatitis.

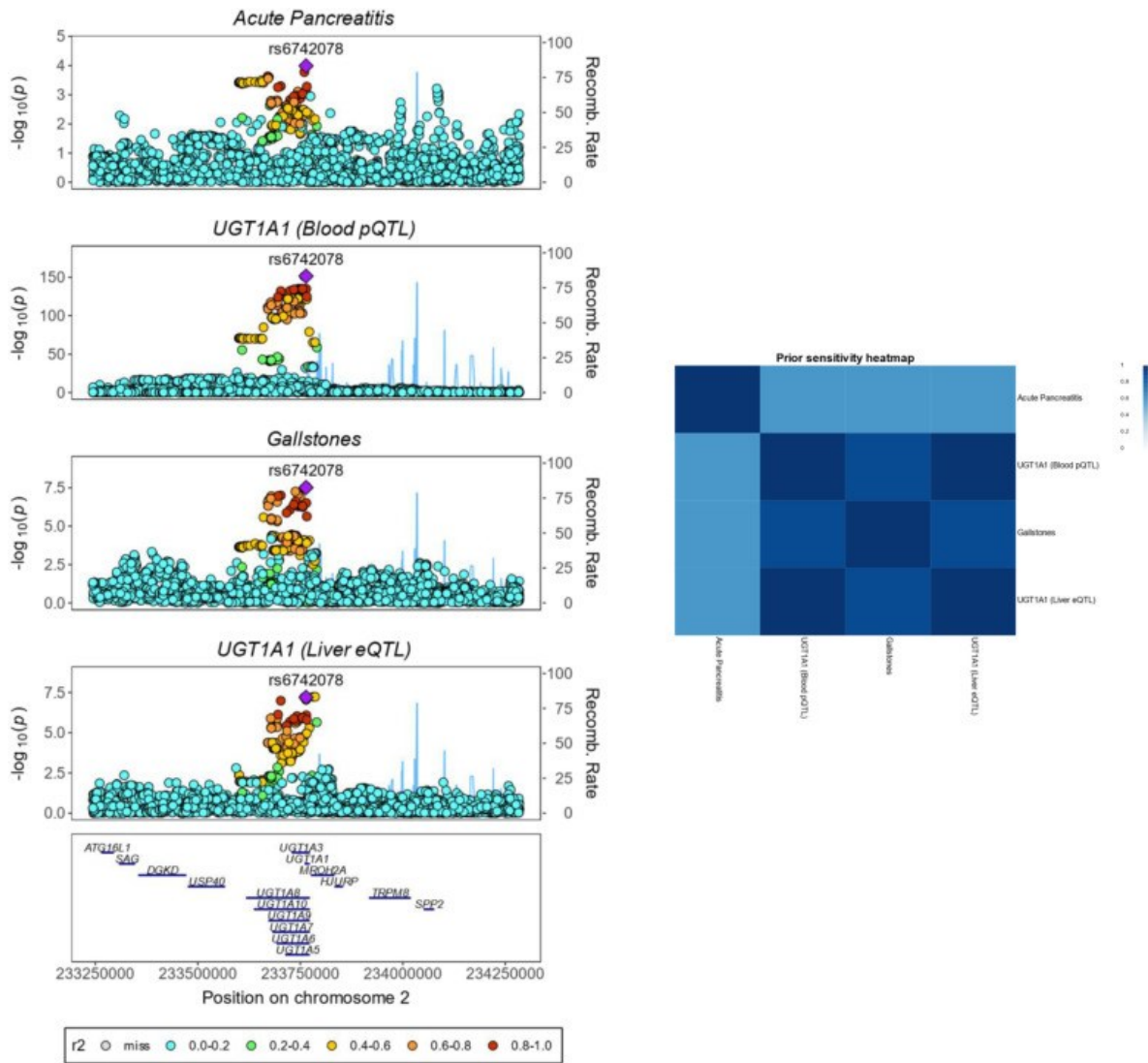




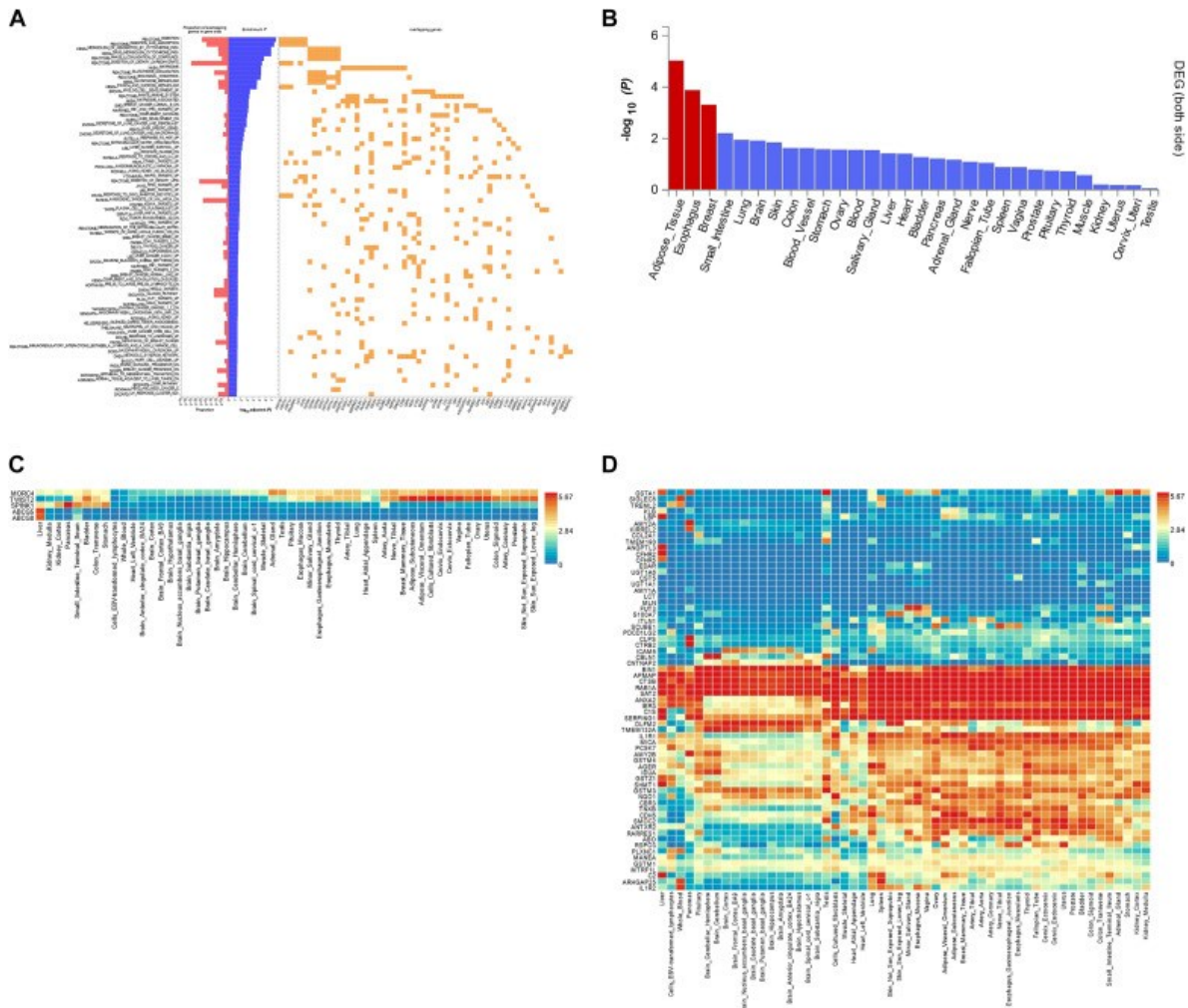
**Figure 2. Cross-trait analysis and genetic correlation of acute pancreatitis with other traits.** (A) Significant cross-phenotype associations with AP (5% FDR). (B) Lollipop plot of genetic correlation between AP and 58 traits. Nominal (0.05), Bonferroni's, and 5% FDR thresholds are indicated by a full black, blue dashed, and red dotted line, respectively. The size of the dots represents the absolute magnitude of the genetic correlation and the red and blue color indicates a negative and positive correlation, respectively. The y-axis represents the P value of the correlation. HDL, high-density lipoprotein; HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; NAFLD, nonalcoholic fatty liver disease; TG, triglyceride; VLDL, very low-density lipoprotein.



**Figure 3. PW-MR studies of acute pancreatitis.** Volcano plot of AP PW-MR study using (A) deCODE and (B) Fenland data. Association strength between proteins and AP according to their effect. Annotated proteins passed the 5% FDR IVW P-value threshold. The green and red colors represent a negative and positive effect, respectively. (C) Venn diagram depicting proteins associated with AP in Fenland only, in deCODE only, or in both. (D) PhenoGram of AP GWAS meta-analysis and PW-MR study significant associations. The blue dots represent the top genome-wide significant loci, the green dots represent the PW-MR significant proteins, and the red dot represents both simultaneously.

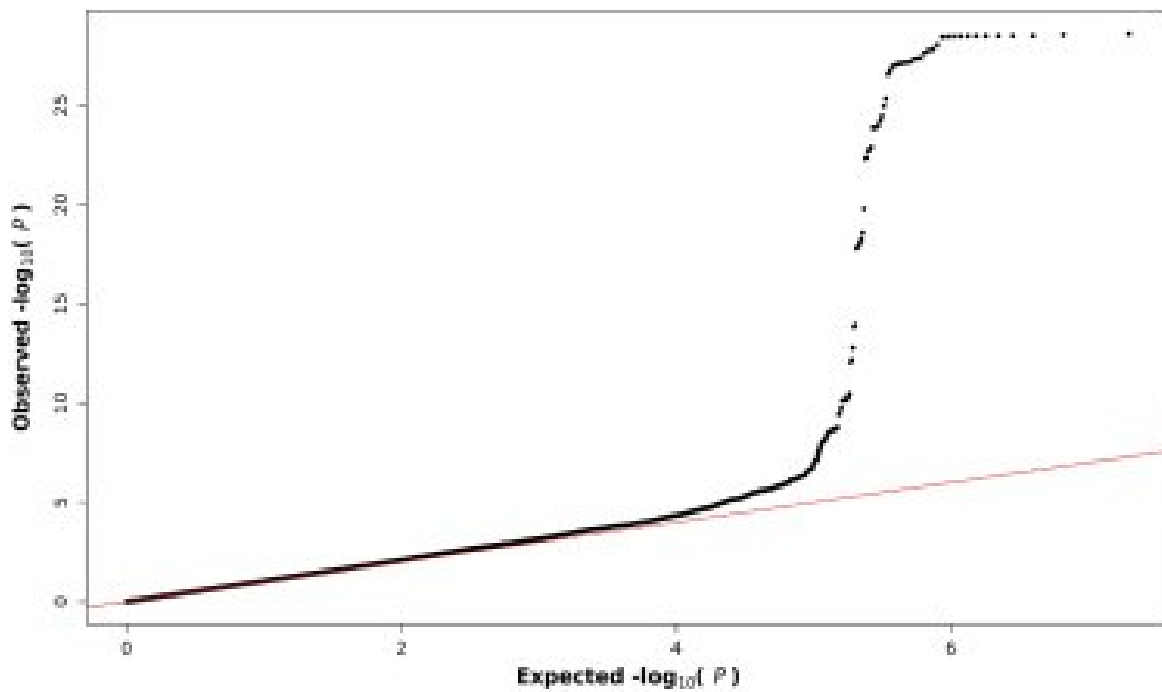


**Figure 4. HyPrColoc plot of UGT1A1 variants.** HyPrColoc stacked association plots and prior sensitivity heat map for UGT1A1 association with AP, its blood protein levels, and liver expression and gallstones.

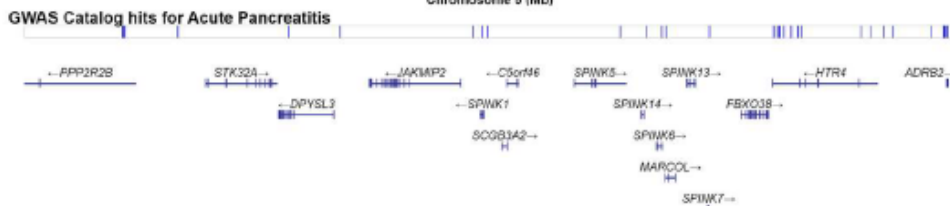
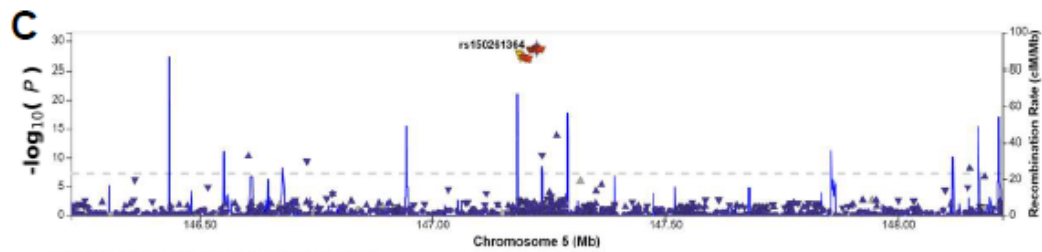
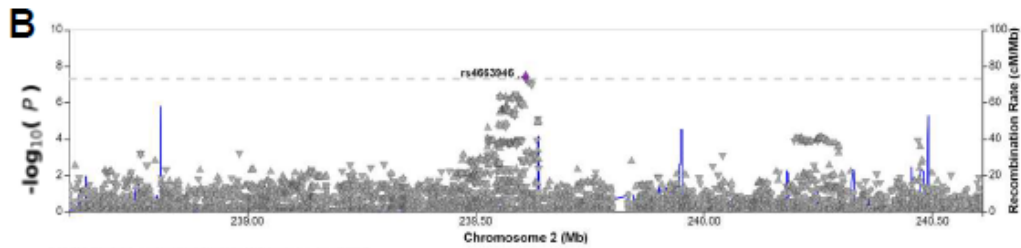
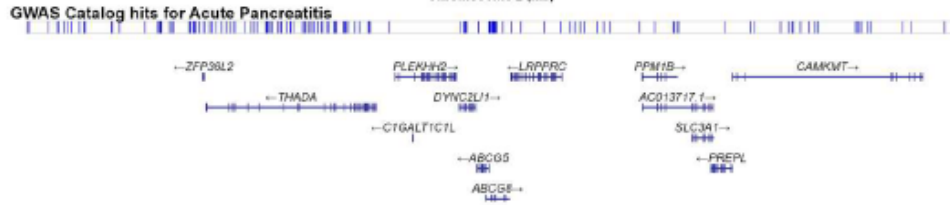
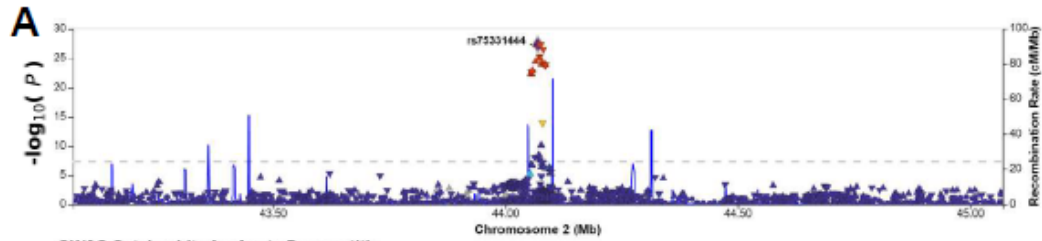


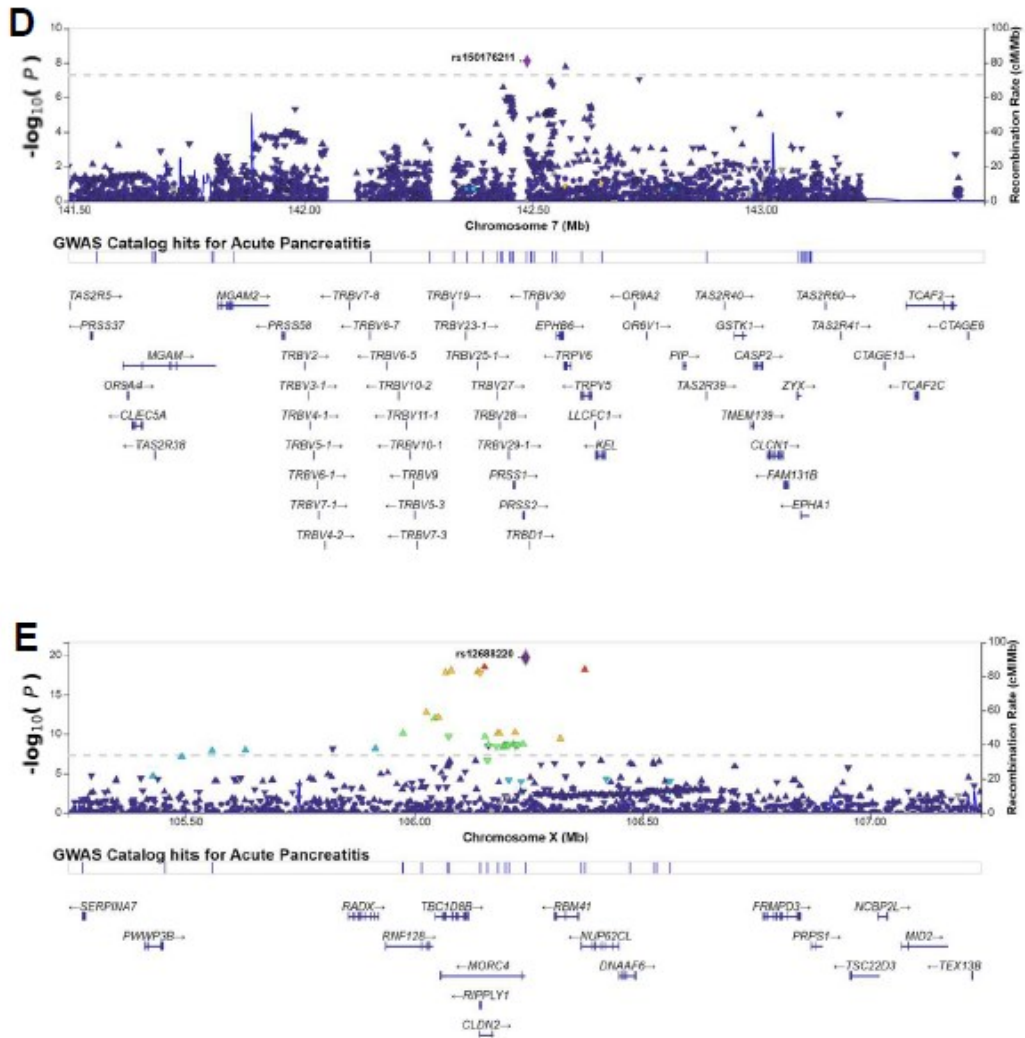
**Figure 5. Functional annotation of the genetic architecture of AP.** (A) Pathway analysis of the 68 unique proteins identified in deCODE and Fenland. Pathways and functions involving the proteins identified by the PW-MR approach in deCODE and Fenland are presented. An enrichment was considered significant if it passed the 5% FDR P-value threshold and if at least 2 genes overlapped with the gene sets. (B) Differential expression analysis of the top genome-wide significant loci and protein-coding genes identified in deCODE and Fenland. Differential expression of significant genes identified by GWAS and PW-MR in 30 GTEx v8 tissues. The red bars represent tissues in which the gene set is differentially expressed. (C) Gene expression heat map of the top genome-wide significant loci. Log<sub>2</sub>-transformed average expression in 54 GTEx v8 tissues. (D) Gene expression heat map of the 68 unique PW-MR significant proteins in deCODE and Fenland. Log<sub>2</sub>-transformed average expression in 54 GTEx v8 tissues.

## 1.11. Supplementary Figures

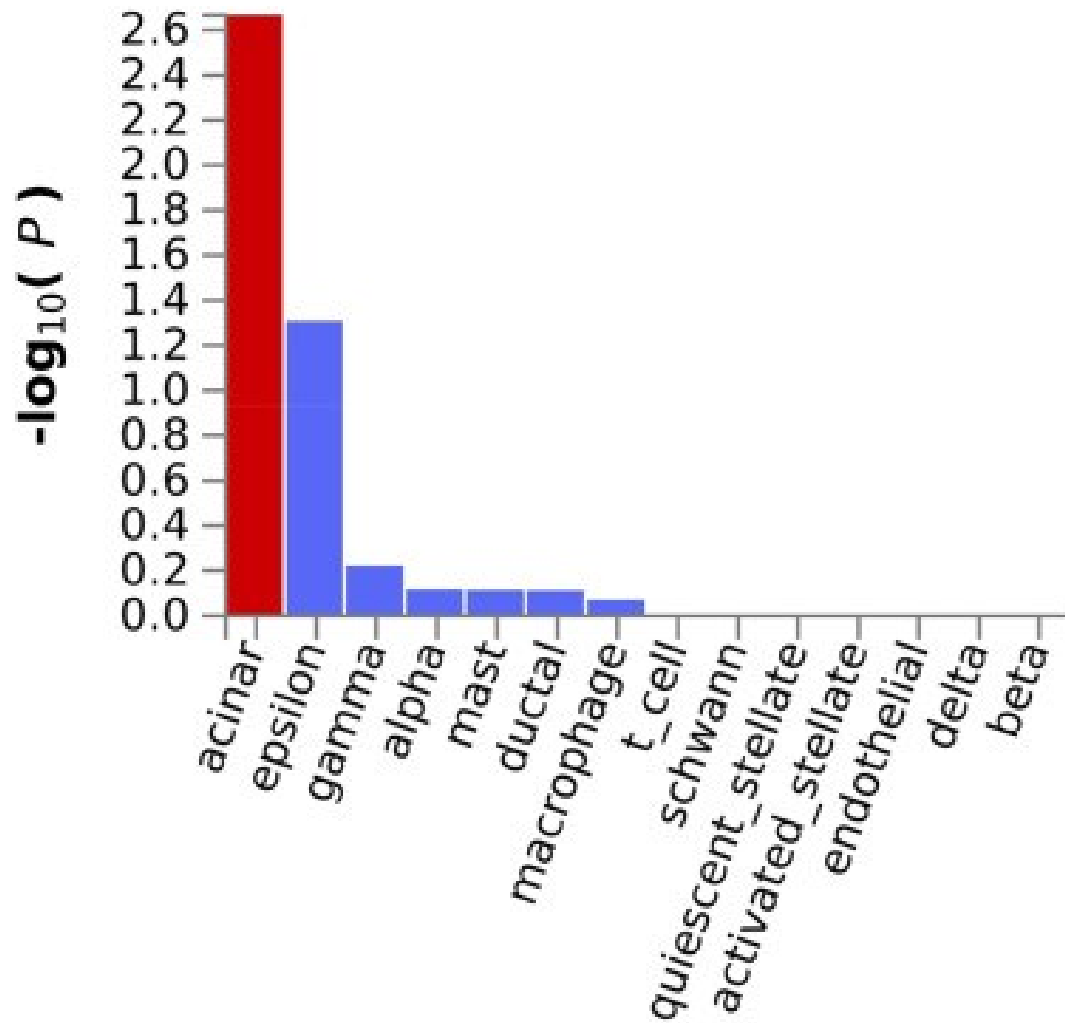


**Supplementary Figure 1. Quantile-quantile plot of the AP GWAS.** The quantile-quantile plot represents the deviation of every GWAS SNP P values (observed) from the P values of a theoretical  $\chi^2$  distribution (expected).





**Supplementary Figure 2. LocusZoom plots of GWAS top SNPs.** Genetic loci harboring top SNPs at the (A) ABCG5/8, (B) TWIST2, (C) SPINK1, (D) PRSS2, and (E) MORC4 loci are depicted. LD value with the top SNP is represented by colors spanning from dark blue (low) to red (high), whereas gray indicates missing LD information. The dashed gray line represents the genome-wide significance threshold of  $5 \times 10^{-8}$ . Neighboring genes are shown at the bottom of the figure.



**Supplementary Figure 3. Bar plot of single-cell specificity.** Cell-type specificity of GWAS significant loci. A red bar indicates a significant association after Benjamini-Hochberg multiple testing correction.

## Discussion

La PA est une maladie complexe et multifactorielle faisant intervenir des processus pro-inflammatoires complexes. Elle peut être causée ou exacerbée par la présence d'une hypertriglycéridémie (200), par l'âge avancé, lequel est généralement accompagné de fibrose, d'accumulation de lipides dans le pancréas, d'une atrophie du pancréas ou d'une dilatation du conduit pancréatique principal (201,202), par l'obésité et la sédentarité (203,204) ou par une prédisposition génétique à la maladie (162,205). La PA peut se présenter sous plusieurs scénarios cliniques différents, impliquer plusieurs autres organes et entraîner un syndrome de défaillance multiviscérale, caractérisé par la défaillance de deux organes ou plus et accompagné d'un déséquilibre homéostatique, dont le risque augmente avec la consommation d'alcool et le syndrome métabolique (206,207). Les complications systémiques de la PA incluent, entre autres, un dérèglement du rythme circadien (208), les thrombose de la veine porte et thrombose veineuse splanchnique (209,210) et le diabète pancréatogénique (211). En outre, de plus en plus d'études suggèrent que l'accumulation de graisses dans le pancréas pourrait causer la PA via le développement d'une stéatose pancréatique non alcoolique et d'une insuffisance pancréatique exocrine, et que cette maladie serait également associée à la stéatose hépatique non alcoolique (212–214). Étant donné les différentes formes, causes et complications de la PA et en l'absence de thérapies pharmacologiques efficaces pour la prévention ou le traitement de cette maladie, les travaux de ce mémoire ont cherché à déterminer les facteurs génétiques de la PA et à mettre en évidence le potentiel thérapeutique des gènes et sentiers biologiques impliqués dans cette maladie.

L'étude présentée dans le chapitre 1 fournit de nouvelles informations sur l'architecture génétique de la PA. Les résultats de la méta-analyse de GWAS effectuée ont permis de mettre en évidence cinq loci génétiques comprenant au moins un SNP significativement associé à la PA à l'échelle du génome. Parmi ceux-ci, trois loci, soit *ABCG5/8*, *SPINK1* et *PRSS2*, sont connus pour être associés à la PA, et deux loci, soit *TWIST2* et *MORC4*, sont de potentiels nouveaux loci de susceptibilité à la PA. Le gène *ABCG8* pourrait causer la PA via son implication dans la formation de calculs biliaires (cholélithiase) (215,216) et dans l'accumulation plasmatique de stérols végétaux, ou phytostérols (sitostérolémie) (217,218), et est soupçonné d'être impliqué dans une forme congénitale d'hyperinsulinisme (219). Alors que le rôle des calculs biliaires dans la pathophysiologie de la PA est bien établi, celui des phytostérols et de l'hyperinsulinisme demeure mal compris. D'une part, les phytostérols permettraient de réduire les concentrations sanguines de LDL cholestérol (220), de freiner la progression du DT2

(221–224) et auraient un effet antioxydant dans le pancréas (224). Une sitostérolémie pourrait influencer la PA en entraînant une accumulation de stérols végétaux dans les tissus et induire une réaction inflammatoire ainsi que l'athérosclérose (225). La sitostérolémie chez les individus présentant des mutations de perte de fonction dans les gènes *ABCG5/G8* peut également être accompagnée de niveaux plus élevés de LDL cholestérol et d'un risque accru de maladies cardiovasculaires (226). D'autre part, l'hyperinsulinisme et l'insulinorésistance associée au DT2 peuvent causer une hypertension artérielle (223,227,228). L'hypertension artérielle, en augmentant le stress oxydatif dans le pancréas, peut causer une PA (229,230). De plus, des déficiences dans les gènes *ABCG5/8* causeraient une HTG chez la souris (231). Le gène *SPINK1* est connu pour être impliqué dans les pancréatites aiguë, chronique et héréditaire (26,66,232). Une perte de fonction de la protéine encodée par ce gène prédisposerait à la PA en entraînant une réponse inflammatoire due à l'activation prématurée de la trypsine dans le pancréas. Notre méta-analyse nous a également permis d'identifier plusieurs variants associés à la PA situés sur le bras q34 du chromosome 7, situés à proximité ou dans le corps des gènes *TRPV6* et *PRSS2*. Le trypsinogène anionique codé par le gène *PRSS2* est le principal isoforme du trypsinogène cationique codé par le gène *PRSS1*, un gène connu pour être associé à la PA et à la PC (233,234). *PRSS2* est plus sensible à la dégradation autocatalytique et moins sensible à l'autoactivation que *PRSS1*. Des mutations entraînant une hausse de l'expression de *PRSS2*, concomitante à une baisse des niveaux de *PRSS1*, ou des mutations de perte de fonction introduisant un nouveau site de clivage dans la protéine auraient un effet protecteur contre la PC (233,235,236), tandis que des mutations de gain de fonction de la protéine pourraient causer la maladie (237). Bien qu'une association entre le gène *TRPV6* et la PC ait récemment été mise en lumière dans des cohortes japonaises et européennes grâce à des données de WES (73,238–240), qui ont permis de mettre en lumière la plus grande fréquence de mutations de perte de fonction du gène *TRPV6* chez des patients souffrant de PC, les mécanismes par lesquels il serait impliqué dans la maladie demeurent mal compris. La protéine codée par ce gène joue un rôle crucial dans l'absorption et le maintien de l'homéostasie du  $Ca^{2+}$  (241). Dans le pancréas, ce gène est principalement exprimé dans les cellules épithéliales des conduits pancréatiques, mais il est également exprimé dans les cellules acineuses (73), où il serait impliqué dans la clairance du  $Ca^{2+}$  relâché par les granules de zymogènes et la recapture de  $Ca^{2+}$  dans la cellule. Une perte de fonction de la protéine codée par *TRPV6* pourrait causer la PA via une augmentation du  $Ca^{2+}$  dans les cellules acineuses, lequel est connu pour stimuler l'activation prématurée de la trypsine (73,242–244). La méta-analyse effectuée dans le chapitre 1 a

également identifié *TWIST2* et *MORC4*, mais les mécanismes par lesquels ces gènes influenceraient la PA demeurent incertains. *TWIST2* est impliqué dans la régulation de la transition épithélio-mésenchymateuse (245). Récemment, il a été montré que *TWIST2* est impliqué dans la stéatose hépatique non alcoolique, que des variants dans ce gène sont associés à de faibles concentrations sanguines de HDL cholestérol, lesquelles permettraient de prédire la défaillance d'organes, la nécrose pancréatique et la mortalité associées à la PA (246), et qu'une surexpression de la protéine entraîne une expression accrue de cytokines pro-inflammatoires et une diminution des niveaux de glycogène dans le muscle squelettique (247–249). Il a également été montré que *TWIST2* régule FGF21, un facteur de croissance de fibroblastes, et le sentier AMPK/mTOR, respectivement impliquées dans le métabolisme énergétique et l'apoptose, en plus d'induire le stress oxydatif dans des cellules cancéreuses du poumon (250). FGF21 est également connue pour améliorer la sensibilité à l'insuline, réduire les niveaux de TG (251) et stimuler la sécrétion d'enzymes digestives par le pancréas (252). *TWIST2* pourrait influencer le risque de PA via le stress oxydatif. Le stress oxydatif et la production d'espèces réactives de l'oxygène pourrait causer des dommages aux mitochondries des cellules du pancréas via l'induction de stress dans le réticulum endoplasmique des cellules du pancréas et le déséquilibre de l'homéostasie du  $Ca^{2+}$ . Le dommage aux mitochondries mènerait ainsi à l'activation du sentier AMPK et à l'inhibition du sentier mTOR, ce qui entrainerait une réponse pro-apoptotique (253). *MORC4* est associé à un risque accru de PC, mais les mécanismes par lesquels ce gène influencerait le risque de la maladie demeurent incertains. La protéine codée par ce gène régulerait la formation de corps nucléaires, jouerait un rôle dans la progression de la phase S du cycle cellulaire et est suspectée d'interagir avec la sous-unité alpha 1 de AMPK (254,255). Cependant, des analyses de génétique fonctionnelle devront être effectuées afin de déterminer si ces deux gènes sont causalement impliqués dans la PA.

L'analyse de PWMR bidirectionnelle a mis en évidence plusieurs protéines sanguines associées à la PA, dont la plupart jouent un rôle clé dans les processus biologiques pouvant mener à la PA. Les associations les plus significatives dans les deux cohortes (deCODE et Fenland) incluent les  $\alpha$ -amylases pancréatiques 2A et 2B (*AMY2A/2B*), le chymotrypsinogène B2 (*CTRB2*), la colipase pancréatique (*CLPS*), l'UDP-glucuronosyltransférase 1A1 (*UGT1A1*), la motiline (*MLN*) et la NAD(P)H quinone déshydrogénase 1 (*NQO1*). Bien que les amylases et la colipase pancréatiques soient couramment utilisées en tant que biomarqueurs pour établir un diagnostic de PA, nos analyses montrent que les concentrations sanguines de ces enzymes seraient causalement associées à la PA,

et non l'inverse, donc qu'elles ne reflètent pas seulement la présence de PA, mais qu'elles sont directement impliquées dans sa pathophysiologie. En outre, nos analyses ont confirmé les associations connues entre la PA ou la PC et *CTRB2* (69,256), l'un des principaux précurseurs des enzymes protéolytiques pancréatiques, *MLN* (257,258), régulant la motilité gastrointestinale, et *NQO1* (169), prévenant la production d'espèces réactives en l'oxygène en catalysant l'oxydation de NAD(P)H et en régulant la NADPH oxydase. Elles ont également mis en évidence le rôle causal de *UGT1A1* dans la PA. *UGT1A1* est une enzyme exprimée dans le foie et qui détoxifie la bilirubine, un produit de la dégradation de l'hémoglobine, en la conjuguant avec l'acide glucuronique. La bilirubine conjuguée est ensuite dissoute dans la bile et excrétée dans le duodénum via le canal cholédoque. Un dysfonctionnement ou une mauvaise régulation de l'enzyme *UGT1A1* pourrait causer la PA en entraînant un excès de bilirubine dans la bile laquelle est, après un excès de cholestérol dans la bile, la seconde cause la plus fréquente de la formation de calculs biliaires (259,260). De plus, les analyses de colocalisation effectuées dans le chapitre 1 ont montré que *UGT1A1* et la PA partagent un même variant causal, indépendant, qui affecte à la fois les niveaux sanguins de *UGT1A1*, son expression dans le foie, les calculs biliaires et la PA. Ce variant (rs6742078) est déjà connu pour être associé à la formation de calculs biliaires (261,262). De plus, des variants dans le gène *UGT1A1* sont connus pour être associés à un risque plus faible de maladies cardiovasculaires (263). Le variant rs6742078 est associé à de plus hautes concentrations sanguines de bilirubine, laquelle, malgré son implication dans la formation de calculs biliaires, est un antioxydant efficace pour éliminer les ROS et empêcher l'oxydation de lipoprotéines dont la LDL (264), laquelle est connue pour son implication dans le risque de maladies cardiovasculaires athérosclérotiques (265). D'autres associations incluent les gènes codant pour la glycoprotéine de surface *RSPO3* et pour le peptide signal *SCUBE1*, impliqués dans l'homéostasie tissulaire et la répartition du tissu adipeux et dans la réponse inflammatoire, respectivement (266–268). De plus, il a récemment été montré que des mutations de perte de fonction dans le gène *NQO1* sont associées à un plus grand risque de maladies liées au stress oxydatif, dont le syndrome métabolique, le DT2 et l'hypertriglycémie (269). De même, les analyses de corrélation génétique effectuées dans l'étude présentée dans le chapitre 1 montrent qu'il existe une corrélation entre la PA et les calculs biliaires, les phytostérols, la stéatose hépatique non alcoolique, les niveaux de cholestérol (LDL, VLDL, chylomicrons) et TG, les maladies cardiovasculaires et l'utilisation de statines et d'agents antithrombotiques, utilisés dans le traitement de l'hypercholestérolémie et des maladies cardiovasculaires athérosclérotiques, respectivement. Ces résultats suggèrent qu'en plus

des calculs biliaires et de l'augmentation des niveaux sanguins ou de l'activation prématurée d'enzymes protéolytiques pancréatiques, le DT2 et les dyslipidémies seraient impliqués dans la PA via leur rôle dans la réponse inflammatoire. La stéatose hépatique non alcoolique pourrait également être impliquée dans la formation de calculs biliaires et contribuer à la PA (270).

Plusieurs des protéines identifiées dans l'analyse de PWMR présentée dans le chapitre 1 pourraient représenter des cibles thérapeutiques pour la PA. Entre autres, cette analyse a montré que des niveaux sanguins élevés de la protéine angiopoïétin-like protein 3 (ANGPTL3) sont associés à un risque plus élevé de PA. ANGPTL3 est un régulateur critique de la lipoprotéine lipase (LPL). Les individus présentant des déficiences en LPL, accompagnant une chylomicronémie familiale, ont des niveaux de TG sanguins plus élevés et sont plus à risque de développer une PA (271). Les thérapies basées sur des petits ARN interférents ciblant l'expression hépatique du gène ANGPTL3 (272) ainsi que sur des anticorps neutralisant ANGPTL3 (273) sont actuellement à l'étude pour la réduction des taux de lipides sanguins. En plus d'identifier ANGPTL3, notre analyse du génome thérapeutique avec la plateforme OpenTargets a identifié d'autres cibles thérapeutiques possibles pour la PA telles que la Klotho bêta (KLB) (qui peut être influencé par la pegbelfermine, un analogue de FGF21), RSPO3 (ciblé par l'anticorps monoclonal rosmantuzumab), NQO1 (ciblé par le composé oral vatiquinone), PDCD1LG2 (ciblé par l'anticorps monoclonal anti PD-L2 (rHlgM12B7) et AMY2A (ciblée par le composé acarbose). L'intérêt de la pegbelfermine et de KLB passerait par leur action sur FGF21, la première permettant d'accroître sa demi-vie, la seconde étant essentielle à son fonctionnement optimal et à l'homéostasie du cholestérol et des acides biliaires (274). Bien que la pegbelfermine soit sous investigation pour le traitement de la stéatose hépatique non alcoolique (275), de faibles niveaux de FGF21 ont été associés à la PA (276) et de hauts niveaux de KLB ont été associés aux carcinomes hépatocellulaires (277). RSPO3 est impliqué dans la fibrose en réponse à l'inflammation d'un tissu, où ses niveaux sont particulièrement élevés (278). Le rosmantuzumab est un anticorps contre RSPO3 utilisé dans le traitement de cancers et l'inhibition de cette protéine pourrait permettre de traiter la PA en freinant la progression de la fibrose et l'insuffisance endocrine et exocrine du pancréas qu'elle entraîne (279,280). La vatiquinone est un composé oral utilisé principalement dans le traitement de maladies associées à des dysfonctionnement mitochondriaux et qui cible NQO1 directement afin d'améliorer la régulation du métabolisme énergétique et le contrôle du stress oxydatif en favorisant la biosynthèse de glutathione (281–283). L'effet de l'anticorps monoclonal recombinant rHlgM12B7 sur la PA, bien qu'incertain, pourrait être dû à son effet modulateur sur la réponse inflammatoire (284,285). L'acarbose, un

médicament anti-diabétique, aurait un effet protecteur sur la PA en inhibant les  $\alpha$ -amylases pancréatiques, surtout chez des patients atteints de DT2 (40), mais son efficacité pour le traitement de la PA n'a pas été étudié (286). L'association entre plusieurs marqueurs inflammatoires et la PA dans notre analyse suggère que la modulation des interleukines 1, 2 et 6 et de molécules d'adhésion telles que la molécule d'adhésion intercellulaire 5 peut influencer l'évolution de la PA. Mais leur réutilisation dans le cadre de la prévention de la PA devra encore faire l'objet d'essais cliniques.

### **Limites**

Cette étude a des limites. D'une part, les données utilisées dans cette étude ne comprenaient que des individus d'ascendance européenne. L'héritabilité de la PA était également faible, ce qui limite la capacité d'effectuer des analyses de traits croisés ainsi que la MR inverse. Notre évaluation du protéome sanguin était également basée uniquement sur des données de GWAS reposant sur une seule technologie (SOMAmer) pour mesurer les protéines sanguines. D'autres techniques telles que les tests d'extension de proximité (Olink) pourraient également être utilisées pour identifier les protéines sanguines liées à la PA. Enfin, notre analyse n'incluait que des SNP ayant une fréquence allélique égale ou supérieure à 1 %. Par conséquent, les variants rares ayant des effets potentiellement plus prononcés n'ont pas pu être identifiés.

## Conclusion

La PA est une maladie complexe faisant intervenir des processus pro-inflammatoires complexes et dont les causes et complications sont multiples. Elle représente une charge importante pour les systèmes de santé, et aucun traitement médical préventif ou curatif n'est actuellement disponible pour cette maladie. Cependant, les progrès dans la compréhension de l'architecture génétique complexe de cette maladie ouvrent de nouvelles voies pour en explorer les thérapies potentielles.

Les travaux de ce mémoire apportent de nouvelles informations sur les déterminants génétiques de la PA et sur l'implication de plusieurs protéines sanguines dans l'apparition de cette maladie. Ils ont également confirmé l'expression de nombreuses protéines dans les tissus métaboliques responsables de la digestion et du métabolisme énergétique tels que l'œsophage, le tissu adipeux, le foie et le pancréas et un rôle potentiel des cellules acineuses pancréatique, venant ainsi appuyer la validité de nos résultats. Cependant, des analyses d'annotation génétique fonctionnelle seront nécessaires pour déterminer l'implication de certains gènes et protéines dans l'étiologie de la PA.

Bien que la présente étude mette en évidence un rôle potentiellement causal de la protéine UGT1A1 dans la pathophysiologie de cette maladie, aucun médicament prometteur ciblant UGT1A1 n'a pu être identifié. En outre, bien que cette étude ait permis d'identifier plusieurs médicaments ciblant certaines des protéines associées à la PA, ceux-ci devront toutefois faire l'objet d'essais cliniques pour en évaluer le potentiel thérapeutique dans le cadre du traitement ou de la prévention de la PA.

L'ensemble de ces résultats met en évidence de potentiels nouveaux déterminants génétiques, protéines et sentiers métaboliques impliqués dans la PA, et suggère que des médicaments existants et ciblant certaines des protéines identifiées pourraient être repositionnés afin de prévenir ou traiter la PA.

## Bibliographie

1. Liu W, Du JJ, Li ZH, Zhang XY, Zuo HD. Liver injury associated with acute pancreatitis: The current status of clinical evaluation and involved mechanisms. *World J Clin Cases*. 2021 Dec 6;9(34):10418–29.
2. Das SLM, Kennedy JIC, Murphy R, Phillips ARJ, Windsor JA, Petrov MS. Relationship between the exocrine and endocrine pancreas after acute pancreatitis. *World J Gastroenterol*. 2014 Dec 7;20(45):17196–205.
3. Noguchi GM, Huising MO. Integrating the inputs that shape pancreatic islet hormone release. *Nat Metab*. 2019 Dec;1(12):1189–201.
4. El Sayed SA, Mukherjee S. Physiology, Pancreas. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 [cited 2023 Apr 25]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK459261/>
5. Banerjee A, Onyuksel H. HUMAN PANCREATIC POLYPEPTIDE IN A PHOSPHOLIPID BASED MICELLAR FORMULATION. *Pharm Res*. 2012 Jun;29(6):1698–711.
6. Washabau RJ. Chapter 1 - Integration of Gastrointestinal Function. In: Washabau RJ, Day MJ, editors. *Canine and Feline Gastroenterology* [Internet]. Saint Louis: W.B. Saunders; 2013 [cited 2023 Apr 25]. p. 1–31. Available from: <https://www.sciencedirect.com/science/article/pii/B9781416036616000018>
7. Merck Manuals Professional Edition [Internet]. [cited 2023 Apr 25]. Chronic Pancreatitis - Gastrointestinal Disorders. Available from: <https://www.merckmanuals.com/en-ca/professional/gastrointestinal-disorders/pancreatitis/chronic-pancreatitis>
8. Merck Manuals Professional Edition [Internet]. [cited 2023 Apr 25]. Acute Pancreatitis - Gastrointestinal Disorders. Available from: <https://www.merckmanuals.com/en-ca/professional/gastrointestinal-disorders/pancreatitis/acute-pancreatitis>
9. Forsmark CE, Vege SS, Wilcox CM. Acute Pancreatitis. *N Engl J Med*. 2016 Nov 17;375(20):1972–81.
10. Ingraham NE, King S, Proper J, Siegel L, Zolfaghari EJ, Murray TA, et al. Morbidity and Mortality Trends of Pancreatitis: An Observational Study. *Surg Infect (Larchmt)*. 2021 Dec;22(10):1021–30.
11. Canadian Digestive Health Foundation [Internet]. [cited 2023 Apr 26]. Pancréatite. Available from: <https://cdhf.ca/fr/digestive-conditions/pancreatitis/>
12. Petrov MS, Yadav D. Global epidemiology and holistic prevention of pancreatitis. *Nat Rev Gastroenterol Hepatol*. 2019 Mar;16(3):175–84.
13. Iannuzzi JP, King JA, Leong JH, Quan J, Windsor JW, Tanyingoh D, et al. Global Incidence of Acute Pancreatitis Is Increasing Over Time: A Systematic Review and Meta-Analysis. *Gastroenterology*. 2022 Jan 1;162(1):122–34.
14. Roberts S, Akbari A, Thorne K, Atkinson M, Evans P. The incidence of acute pancreatitis: impact of social deprivation, alcohol consumption, seasonal and demographic factors. *Aliment Pharmacol Ther*. 2013 Sep;38(5):539–48.
15. Yadav D, Lowenfels AB. Trends in the Epidemiology of the First Attack of Acute Pancreatitis: A Systematic Review. *Pancreas*. 2006 Nov;33(4):323.
16. Yadav D, Lowenfels AB. The Epidemiology of Pancreatitis and Pancreatic Cancer. *Gastroenterology*. 2013 Jun;144(6):1252–61.
17. Fagenholz PJ, Castillo CF del, Harris NS, Pelletier AJ, Camargo CA. Increasing United States hospital admissions for acute pancreatitis, 1988-2003. *Ann Epidemiol*. 2007 Jul;17(7):491–7.

18. Nesvaderani M, Eslick GD, Vagg D, Faraj S, Cox MR. Epidemiology, aetiology and outcomes of acute pancreatitis: A retrospective cohort study. *International Journal of Surgery*. 2015 Nov 1;23:68–74.
19. Shayesteh S, Fouladi DF, Blanco A, Fishman EK, Kawamoto S. Idiopathic recurrent acute pancreatitis in the context of pancreas divisum: A case report. *Radiology Case Reports*. 2020 Nov 1;15(11):2255–8.
20. Mohy-ud-din N, Morrissey S. Pancreatitis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 [cited 2022 Jan 17]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK538337/>
21. Jones MR, Hall OM, Kaye AM, Kaye AD. Drug-Induced Acute Pancreatitis: A Review. *Ochsner J*. 2015;15(1):45–51.
22. Gallstone Pancreatitis [Internet]. 2019 [cited 2023 Apr 26]. Available from: <https://www.hopkinsmedicine.org/health/conditions-and-diseases/gallstone-pancreatitis>
23. Baeza-Zapata AA, García-Compeán D, Jaquez-Quintana JO, Scharrer-Cabello SI, Cueto-Aguilera ÁND, Maldonado-Garza HJ. Acute Pancreatitis in Elderly Patients. *Gastroenterology*. 2021 Dec 1;161(6):1736–40.
24. Pharmacy MAEZ PharmD, MS, CGP Senior Care Consultant Pharmacist and President of MZ Associates, Inc , Norwich, New York [www.mzassociatesinc.com](http://www.mzassociatesinc.com) Recipient of the Excellence in Geriatric Pharmacy Practice Award from the Commission for Certification in Geriatric. Acute Pancreatitis: Risks, Causes, and Mortality in Older Adults [Internet]. [cited 2023 Apr 26]. Available from: <https://www.uspharmacist.com/article/acute-pancreatitis-risks-causes-and-mortality-in-older-adults>
25. Molnár G, Gyarmathy VA, Zádori N, Hegyi P, Kanizsai P. Severe Hypertriglyceridemia-Induced Acute Pancreatitis. *Case Rep Gastroenterol*. 2021 Feb 18;15(1):218–24.
26. Whitcomb DC, E. Lowe M. Chapter 57 - Hereditary, Familial, and Genetic Disorders of the Pancreas and Pancreatic Disorders in Childhood. In: Feldman M, Friedman LS, Brandt LJ, editors. *Sleisenger and Fordtran's Gastrointestinal and Liver Disease (Ninth Edition)* [Internet]. Philadelphia: W.B. Saunders; 2010 [cited 2023 Apr 26]. p. 931-957.e8. Available from: <https://www.sciencedirect.com/science/article/pii/B9781416061892000573>
27. de Pretis N, Amodio A, Frulloni L. Hypertriglyceridemic pancreatitis: Epidemiology, pathophysiology and clinical management. *United European Gastroenterol J*. 2018 Jun;6(5):649–55.
28. Yuan S, Giovannucci EL, Larsson SC. Gallstone disease, diabetes, calcium, triglycerides, smoking and alcohol consumption and pancreatitis risk: Mendelian randomization study. *npj Genom Med*. 2021 Mar 29;6(1):1–7.
29. Valdivielso P, Ramírez-Bueno A, Ewald N. Current knowledge of hypertriglyceridemic pancreatitis. *Eur J Intern Med*. 2014 Oct;25(8):689–94.
30. Ewald N, Hardt PD, Kloer HU. Severe hypertriglyceridemia and pancreatitis: presentation and management. *Curr Opin Lipidol*. 2009 Dec;20(6):497–504.
31. Papachristou GI, Machicado JD, Stevens T, Goenka MK, Ferreira M, Gutierrez SC, et al. Acute pancreatitis patient registry to examine novel therapies in clinical experience (APPRENTICE): an international, multicenter consortium for the study of acute pancreatitis. *Ann Gastroenterol*. 2017;30(1):106–13.
32. Yadav D, Pitchumoni CS. Issues in hyperlipidemic pancreatitis. *J Clin Gastroenterol*. 2003 Jan;36(1):54–62.
33. Koziel D, Gluszek S, Kowalik A, Chlopek M, Pieciak L. Genetic mutations in SPINK1, CFTR, CTRC genes in acute pancreatitis. *BMC Gastroenterol*. 2015 Jun 23;15:70.

34. Raphael KL, Willingham FF. Hereditary pancreatitis: current perspectives. *Clin Exp Gastroenterol*. 2016 Jul 26;9:197–207.
35. Ruhl CE, Everhart JE. Gallstone disease is associated with increased mortality in the United States. *Gastroenterology*. 2011 Feb;140(2):508–16.
36. Sandzén B, Haapamäki MM, Nilsson E, Stenlund HC, Oman M. Cholecystectomy and sphincterotomy in patients with mild acute biliary pancreatitis in Sweden 1988 - 2003: a nationwide register study. *BMC Gastroenterol*. 2009 Oct 23;9:80.
37. Sadr-Azodi O, Orsini N, Andrén-Sandberg Å, Wolk A. Abdominal and total adiposity and the risk of acute pancreatitis: a population-based prospective cohort study. *Am J Gastroenterol*. 2013 Jan;108(1):133–9.
38. Hong S, Qiwen B, Ying J, Wei A, Chaoyang T. Body mass index and the risk and prognosis of acute pancreatitis: a meta-analysis. *Eur J Gastroenterol Hepatol*. 2011 Nov;23(12):1136–43.
39. Girman CJ, Kou TD, Cai B, Alexander CM, O'Neill EA, Williams-Herman DE, et al. Patients with type 2 diabetes mellitus have higher risk for acute pancreatitis compared with those without diabetes. *Diabetes Obes Metab*. 2010 Sep;12(9):766–71.
40. Lai SW, Muo CH, Liao KF, Sung FC, Chen PC. Risk of acute pancreatitis in type 2 diabetes and risk reduction on anti-diabetic drugs: a population-based cohort study in Taiwan. *Am J Gastroenterol*. 2011 Sep;106(9):1697–704.
41. Noel RA, Braun DK, Patterson RE, Bloomgren GL. Increased risk of acute pancreatitis and biliary disease observed in patients with type 2 diabetes: a retrospective cohort study. *Diabetes Care*. 2009 May;32(5):834–8.
42. Prizment AE, Jensen EH, Hopper AM, Virnig BA, Anderson KE. Risk Factors for Pancreatitis in Older Women: The Iowa Women's Health Study. *Ann Epidemiol*. 2015 Jul;25(7):544–8.
43. Lindkvist B, Appelros S, Manjer J, Berglund G, Borgstrom A. A prospective cohort study of smoking in acute pancreatitis. *Pancreatology*. 2008;8(1):63–70.
44. Tolstrup JS, Kristiansen L, Becker U, Grønbaek M. Smoking and risk of acute and chronic pancreatitis among women and men: a population-based cohort study. *Arch Intern Med*. 2009 Mar 23;169(6):603–9.
45. Sadr-Azodi O, Andrén-Sandberg Å, Orsini N, Wolk A. Cigarette smoking, smoking cessation and acute pancreatitis: a prospective population-based study. *Gut*. 2012 Feb;61(2):262–7.
46. Hartwig W, Werner J, Ryschich E, Mayer H, Schmidt J, Gebhard MM, et al. Cigarette Smoke Enhances Ethanol-Induced Pancreatic Injury. *Pancreas*. 2000 Oct;21(3):272.
47. Apte MV, Pirola RC, Wilson JS. Mechanisms of alcoholic pancreatitis. *J Gastroenterol Hepatol*. 2010 Dec;25(12):1816–26.
48. Altomare E, Grattagliano I, Vendemiale G, Palmieri V, Palasciano G. Acute ethanol administration induces oxidative changes in rat pancreatic tissue. *Gut*. 1996 May;38(5):742–6.
49. Casini A, Galli A, Pignatola P, Frulloni L, Grappone C, Milani S, et al. Collagen type I synthesized by pancreatic periacinar stellate cells (PSC) co-localizes with lipid peroxidation-derived aldehydes in chronic alcoholic pancreatitis. *J Pathol*. 2000 Sep;192(1):81–9.
50. Norton ID, Apte MV, Lux O, Haber PS, Pirola RC, Wilson JS. Chronic ethanol administration causes oxidative stress in the rat pancreas. *J Lab Clin Med*. 1998 May;131(5):442–6.
51. Haber PS, Wilson JS, Apte MV, Pirola RC. Fatty acid ethyl esters increase rat pancreatic lysosomal fragility. *J Lab Clin Med*. 1993 Jun;121(6):759–64.
52. Del Castillo-Vaquero A, Salido GM, González A. Increased calcium influx in the presence of ethanol in mouse pancreatic acinar cells. *Int J Exp Pathol*. 2010 Apr;91(2):114–24.
53. Neale lab [Internet]. 2017 [cited 2023 Apr 27]. Heritability 101: What is “heritability”? Available from: <http://www.nealelab.is/blog/2017/9/13/heritability-101-what-is-heritability>

54. Barry CJS, Walker VM, Cheesman R, Davey Smith G, Morris TT, Davies NM. How to estimate heritability: a guide for genetic epidemiologists. *International Journal of Epidemiology*. 2023 Apr 1;52(2):624–32.
55. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet*. 2008 Apr;9(4):255–66.
56. Tang M, Wang T, Zhang X. A review of SNP heritability estimation methods. *Briefings in Bioinformatics*. 2022 May 1;23(3):bbac067.
57. Wright AF. Genetic Variation: Polymorphisms and Mutations. In: *Encyclopedia of Life Sciences* [Internet]. John Wiley & Sons, Ltd; 2005 [cited 2023 May 13]. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1038/npg.els.0005005>
58. Wainschtein P, Jain D, Zheng Z, Cupples LA, Shadyab AH, McKnight B, et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet*. 2022 Mar;54(3):263–73.
59. Goswami C, Chattopadhyay A, Chuang EY. Rare variants: data types and analysis strategies. *Ann Transl Med*. 2021 Jun;9(12):961.
60. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of heritability for human height. *Nat Genet*. 2010 Jul;42(7):565–9.
61. Gorry MC, Ghabbaizedeh D, Furey W, Gates LK, Preston RA, Aston CE, et al. Mutations in the cationic trypsinogen gene are associated with recurrent acute and chronic pancreatitis. *Gastroenterology*. 1997 Oct;113(4):1063–8.
62. Whitcomb DC, LaRusch J, Krasinskas AM, Klei L, Smith JP, Brand RE, et al. Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis. *Nat Genet*. 2012 Dec;44(12):1349–54.
63. Schneider A, Larusch J, Sun X, Aloe A, Lamb J, Hawes R, et al. Combined bicarbonate conductance-impairing variants in CFTR and SPINK1 variants are associated with chronic pancreatitis in patients without cystic fibrosis. *Gastroenterology*. 2011 Jan;140(1):162–71.
64. Sharer N, Schwarz M, Malone G, Howarth A, Painter J, Super M, et al. Mutations of the Cystic Fibrosis Gene in Patients with Chronic Pancreatitis. *N Engl J Med*. 1998 Sep 3;339(10):645–52.
65. Cohn JA, Friedman KJ, Noone PG, Knowles MR, Silverman LM, Jowell PS. Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. *N Engl J Med*. 1998 Sep 3;339(10):653–8.
66. Witt H, Luck W, Hennies HC, Claßen M, Kage A, Laß U, et al. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. *Nat Genet*. 2000 Jun;25(2):213–6.
67. Aoun E, Muddana V, Papachristou GI, Whitcomb DC. SPINK1N34S Is Strongly Associated With Recurrent Acute Pancreatitis but Is Not a Risk Factor for the First or Sentinel Acute Pancreatitis Event. *Official journal of the American College of Gastroenterology | ACG*. 2010 Feb;105(2):446.
68. Schmidt AW, Kühnapfel A, Kirsten H, Grallert H, Hellerbrand C, Kiefer F, et al. Colocalization analysis of pancreas eQTLs with risk loci from alcoholic and novel non-alcoholic chronic pancreatitis GWAS suggests potential disease causing mechanisms. *Pancreatology*. 2022 May 1;22(4):449–56.
69. Rosendahl J, Kirsten H, Hegyi E, Kovacs P, Weiss FU, Laumen H, et al. Genome-wide association study identifies inversion in the CTRB1-CTRB2 locus to modify risk for alcoholic and non-alcoholic chronic pancreatitis. *Gut*. 2018 Oct 1;67(10):1855–63.
70. Hereditary pancreatitis: MedlinePlus Genetics [Internet]. [cited 2023 Apr 27]. Available from: <https://medlineplus.gov/genetics/condition/hereditary-pancreatitis/>

71. Greer PJ, Spagnolo DM, Orlova E, Ohlsen CS, Whitcomb DC. Tu1192: ACUTE PANCREATITIS GWAS: HIGH RISK GENETIC VARIANTS IN THE ABCG8 LOCUS. *Gastroenterology*. 2022 May 1;162(7):S.
72. Shah IA, Prasad H, Banerjee S, Kurien RT, Chowdhury SD, Visweswariah SS. A novel frameshift mutation in TRPV6 is associated with hereditary pancreatitis. *Front Genet*. 2022;13:1058057.
73. Masamune A, Kotani H, Sörgel FL, Chen JM, Hamada S, Sakaguchi R, et al. Variants That Affect Function of Calcium Channel TRPV6 Are Associated With Early-Onset Chronic Pancreatitis. *Gastroenterology*. 2020 May;158(6):1626-1641.e8.
74. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*. 2017 Jul 6;101(1):5–22.
75. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015 Nov;47(11):1236–41.
76. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014 Sep 15;23(R1):R89-98.
77. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003 Feb;32(1):1–22.
78. Zhang Y, Cheng Y, Jiang W, Ye Y, Lu Q, Zhao H. Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. *Brief Bioinform*. 2021 Jan 26;22(5):bbaa442.
79. Ni G, Moser G, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *Am J Hum Genet*. 2018 Jun 7;102(6):1185–94.
80. Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018 Jul 12;362:k601.
81. Smith GD, Ebrahim S. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies. In: *Biosocial Surveys* [Internet]. National Academies Press (US); 2008 [cited 2023 Apr 28]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK62433/>
82. Ference BA, Holmes MV, Smith GD. Using Mendelian Randomization to Improve the Design of Randomized Trials. *Cold Spring Harb Perspect Med*. 2021 Jan 7;11(7):a040980.
83. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ*. 2005 May 7;330(7499):1076–9.
84. Hemani G, Tilling K, Smith GD. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genetics*. 2017 Nov 17;13(11):e1007081.
85. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*. 2008 Apr 15;27(8):1133–63.
86. Taliun SAG, Evans DM. Ten simple rules for conducting a mendelian randomization study. *PLOS Computational Biology*. 2021 Aug 12;17(8):e1009238.
87. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Loos R, editor. *eLife*. 2018 May 30;7:e34408.
88. Sobczyk MK, Zheng J, Smith GD, Gaunt TR. Systematic comparison of Mendelian randomization studies and randomized controlled trials using electronic databases [Internet].

- medRxiv; 2022 [cited 2023 Apr 30]. p. 2022.04.11.22273633. Available from: <https://www.medrxiv.org/content/10.1101/2022.04.11.22273633v2>
89. Roberts R. Mendelian Randomization Studies Promise to Shorten the Journey to FDA Approval. *JACC Basic Transl Sci.* 2018 Nov 12;3(5):690–703.
  90. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology.* 2011 Jun 1;40(3):755–64.
  91. Bowden J, Del Greco M F, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int J Epidemiol.* 2019 Jun;48(3):728–42.
  92. Cho Y, Haycock PC, Sanderson E, Gaunt TR, Zheng J, Morris AP, et al. Exploiting horizontal pleiotropy to search for causal pathways within a Mendelian randomization framework. *Nat Commun.* 2020 Feb 21;11:1010.
  93. Lutz SM, Wu AC, Hokanson JE, Vansteelandt S, Lange C. Caution against examining the role of reverse causality in Mendelian Randomization. *Genetic Epidemiology.* 2021;45(5):445–54.
  94. Burgess S, Swanson SA, Labrecque JA. Are Mendelian randomization investigations immune from bias due to reverse causation? *Eur J Epidemiol.* 2021;36(3):253–7.
  95. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. *Nat Rev Methods Primers.* 2021 Aug 26;1(1):1–21.
  96. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015 Mar;47(3):291–5.
  97. Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLOS Genetics.* 2020 Apr 20;16(4):e1008720.
  98. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell.* 2017 Jun 15;169(7):1177–86.
  99. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am J Hum Genet.* 2022 May 5;109(5):767–82.
  100. Plagnol V, Smyth DJ, Todd JA, Clayton DG. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics.* 2009 Apr;10(2):327–34.
  101. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics.* 2014 May 15;10(5):e1004383.
  102. Wallace C. Statistical testing of shared genetic control for potentially related traits. *Genet Epidemiol.* 2013 Dec;37(8):802–13.
  103. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun.* 2021 Feb 3;12(1):764.
  104. Robinson JW, Hemani G, Babaei MS, Huang Y, Baird DA, Tsai EA, et al. An efficient and robust tool for colocalisation: Pair-wise Conditional and Colocalisation (PWCoCo) [Internet]. *bioRxiv*; 2022 [cited 2023 May 1]. p. 2022.08.08.503158. Available from: <https://www.biorxiv.org/content/10.1101/2022.08.08.503158v1>
  105. Ornitz DM, Legeai-Mallet L. Achondroplasia: Development, Pathogenesis, and Therapy. *Dev Dyn.* 2017 Apr;246(4):291–309.

106. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. *Nature*. 2022 Oct;610(7933):704–12.
107. The nature and identification of quantitative trait loci. *Nat Rev Genet*. 2003 Nov;4(11):911–6.
108. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016 May 15;32(10):1479–85.
109. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012 Sep 7;337(6099):1190–5.
110. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet*. 2010 Apr 1;6(4):e1000895.
111. Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet*. 2016 Dec 1;99(6):1245–60.
112. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016 May;48(5):481–7.
113. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013 Sep 26;501(7468):506–11.
114. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet*. 2012 Mar 25;44(5):502–10.
115. Neavin D, Nguyen Q, Daniszewski MS, Liang HH, Chiu HS, Wee YK, et al. Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biology*. 2021 Mar 5;22(1):76.
116. Zhang K, Hocker JD, Miller M, Hou X, Chiou J, Poirion OB, et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell*. 2021 Nov 24;184(24):5985–6001.e19.
117. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015 May 8;348(6235):648–60.
118. Ongen H, Brown AA, Delaneau O, Panousis NI, Nica AC, GTEx Consortium, et al. Estimating the causal tissues for complex traits and diseases. *Nat Genet*. 2017 Dec;49(12):1676–83.
119. Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Research*. 2020 Jan 8;48(D1):D983–91.
120. Vandiedonck C. Genetic association of molecular traits: A help to identify causative variants in complex diseases. *Clinical Genetics*. 2018;93(3):520–32.
121. Li MJ, Yan B, Sham PC, Wang J. Exploring the function of genetic variants in the non-coding genomic regions: approaches for identifying human regulatory variants affecting gene expression. *Briefings in Bioinformatics*. 2015 May 1;16(3):393–412.
122. Chen XF, Zhu DL, Yang M, Hu WX, Duan YY, Lu BJ, et al. An Osteoporosis Risk SNP at 1p36.12 Acts as an Allele-Specific Enhancer to Modulate LINC00339 Expression via Long-Range Loop Formation. *Am J Hum Genet*. 2018 May 3;102(5):776–93.
123. He X, Fuller CK, Song Y, Meng Q, Zhang B, Yang X, et al. Sherlock: Detecting Gene-Disease Associations by Matching Patterns of Expression QTL and GWAS. *The American Journal of Human Genetics*. 2013 May 2;92(5):667–80.

124. Hunt KA, Zhernakova A, Turner G, Heap GAR, Franke L, Bruinenberg M, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet.* 2008 Apr;40(4):395–402.
125. Porcu E, Rüeger S, Lepik K, Santoni FA, Reymond A, Kutalik Z. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun.* 2019 Jul 24;10(1):3300.
126. Zhao H, Rasheed H, Nøst TH, Cho Y, Liu Y, Bhatta L, et al. Proteome-wide Mendelian randomization in global biobank meta-analysis reveals multi-ancestry drug targets for common diseases [Internet]. medRxiv; 2022 [cited 2023 May 1]. p. 2022.01.09.21268473. Available from: <https://www.medrxiv.org/content/10.1101/2022.01.09.21268473v1>
127. Richardson TG, Zheng J, Davey Smith G, Timpson NJ, Gaunt TR, Relton CL, et al. Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk. *Am J Hum Genet.* 2017 Oct 5;101(4):590–602.
128. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov.* 2002 Sep;1(9):727–30.
129. National Center for Advancing Translational Sciences [Internet]. 2015 [cited 2023 May 14]. Illuminating the Druggable Genome (IDG). Available from: <https://ncats.nih.gov/idg>
130. Drews J. Drug Discovery: A Historical Perspective. *Science.* 2000 Mar 17;287(5460):1960–4.
131. Carvalho-Silva D, Pierleoni A, Pignatelli M, Ong C, Fumis L, Karamanis N, et al. Open Targets Platform: new developments and updates two years on. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D1056–65.
132. Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, et al. Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D985–94.
133. Radoux CJ, Vianello F, McGreig J, Desai N, Bradley AR. The druggable genome: Twenty years later. *Frontiers in Bioinformatics* [Internet]. 2022 [cited 2023 May 14];2. Available from: <https://www.frontiersin.org/articles/10.3389/fbinf.2022.958378>
134. Mitsopoulos C, Schierz AC, Workman P, Al-Lazikani B. Distinctive Behaviors of Druggable Proteins in Cellular Networks. *PLoS Comput Biol.* 2015 Dec;11(12):e1004597.
135. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Science Translational Medicine.* 2017 Mar 29;9(383):eaag1166.
136. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov.* 2010 Mar;9(3):203–14.
137. Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov.* 2009 Dec;8(12):959–68.
138. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov.* 2014 Jun;13(6):419–31.
139. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. *Nat Genet.* 2015 Aug;47(8):856–60.
140. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov.* 2013 Aug;12(8):581–94.
141. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol.* 2012 Apr;30(4):317–20.

142. Peery AF, Dellon ES, Lund J, Crockett SD, McGowan CE, Bulsiewicz WJ, et al. Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology*. 2012 Nov;143(5):1179-1187.e3.
143. Pokras S, Ray M, Zheng S, Ding Y, Chen CC. The Short- and Long-Term Burden of Acute Pancreatitis in the United States: A Retrospective Cohort Study. *Pancreas*. 2021 Mar;50(3):330.
144. Li CL, Jiang M, Pan CQ, Li J, Xu LG. The global, regional, and national burden of acute pancreatitis in 204 countries and territories, 1990-2019. *BMC Gastroenterol*. 2021 Aug 25;21(1):332.
145. Brindise E, Elkhatib I, Kuruvilla A, Silva R. Temporal Trends in Incidence and Outcomes of Acute Pancreatitis in Hospitalized Patients in the United States From 2002 to 2013. *Pancreas*. 2019 Feb;48(2):169–75.
146. Malli A, Li F, Conwell DL, Cruz-Monserrate Z, Hussan H, Krishna SG. The Burden of Systemic Adiposity on Pancreatic Disease: Acute Pancreatitis, Non-Alcoholic Fatty Pancreas Disease, and Pancreatic Cancer. *Journal of the Pancreas*. 18(5):0–0.
147. Papachristou GI, Papachristou DJ, Avula H, Slivka A, Whitcomb DC. Obesity increases the severity of acute pancreatitis: performance of APACHE-O score and correlation with the inflammatory response. *Pancreatology*. 2006;6(4):279–85.
148. Xie J, Xu L, Pan Y, Li P, Liu Y, Pan Y, et al. Impact of visceral adiposity on severity of acute pancreatitis: a propensity score-matched analysis. *BMC Gastroenterol*. 2019 Jun 13;19(1):87.
149. Roberts-Thomson IC. Progression from acute to chronic pancreatitis. *JGH Open: An Open Access Journal of Gastroenterology and Hepatology*. 2021 Dec;5(12):1321.
150. Sankaran SJ, Xiao AY, Wu LM, Windsor JA, Forsmark CE, Petrov MS. Frequency of Progression From Acute to Chronic Pancreatitis and Risk Factors: A Meta-analysis. *Gastroenterology*. 2015 Nov 1;149(6):1490-1500.e1.
151. Sweeny KF, Lin TK, Nathan JD, Denson LA, Husain SZ, Hornung L, et al. Rapid Progression of Acute Pancreatitis to Acute Recurrent Pancreatitis in Children. *Journal of Pediatric Gastroenterology and Nutrition*. 2019 Jan;68(1):104.
152. Ahmed Ali U, Issa Y, Hagenaaers JC, Bakker OJ, van Goor H, Nieuwenhuijs VB, et al. Risk of Recurrent Pancreatitis and Progression to Chronic Pancreatitis After a First Episode of Acute Pancreatitis. *Clin Gastroenterol Hepatol*. 2016 May;14(5):738–46.
153. Lankisch PG, Schirren CA, Kunze E. Undetected fatal acute pancreatitis: why is the disease so frequently overlooked? *Am J Gastroenterol*. 1991 Mar;86(3):322–6.
154. Cappell MS. Acute Pancreatitis: Etiology, Clinical Presentation, Diagnosis, and Therapy. *Medical Clinics of North America*. 2008 Jul;92(4):889–923.
155. Lee DW, Cho CM. Predicting Severity of Acute Pancreatitis. *Medicina (Kaunas)*. 2022 Jun 11;58(6):787.
156. Sarri G, Guo Y, Iheanacho I, Puelles J. Moderately severe and severe acute pancreatitis : a systematic review of the outcomes in the USA and European Union-5. *BMJ Open Gastroenterology*. 2019 Feb 1;6(1):e000248.
157. Christian JB, Arondekar B, Buysman EK, Johnson SL, Seeger JD, Jacobson TA. Clinical and economic benefits observed when follow-up triglyceride levels are less than 500 mg/dL in patients with severe hypertriglyceridemia. *J Clin Lipidol*. 2012;6(5):450–61.
158. Cook ME, Knoph CS, Fjelsted CA, Frøkjær JB, Bilgrau AE, Novovic S, et al. Effects of a peripherally acting  $\mu$ -opioid receptor antagonist for the prevention of recurrent acute pancreatitis: study protocol for an investigator-initiated, randomized, placebo-controlled, double-blind clinical trial (PAMORA-RAP trial). *Trials*. 2023 May 1;24(1):301.

159. Mayerle J, Sendler M, Hegyi E, Beyer G, Lerch MM, Sahin-Tóth M. Genetics, Cell Biology, and Pathophysiology of Pancreatitis. *Gastroenterology*. 2019 May;156(7):1951-1968.e1.
160. Piseddu I, Vielhauer J, Mayerle J. Genetic Testing in Acute and Chronic Pancreatitis. *Curr Treat Options Gastro*. 2022 Sep 1;20(3):429–44.
161. Weiss FU, Laemmerhirt F, Lerch MM. Acute Pancreatitis: Genetic Risk and Clinical Implications. *Journal of Clinical Medicine*. 2021 Jan;10(2):190.
162. LaRusch J, Whitcomb DC. Genetics of pancreatitis. *Curr Opin Gastroenterol*. 2011 Sep;27(5):467–74.
163. Ramamoorthy A, Yee SW, Karnes J. Unveiling the Genetic Architecture of Human Disease for Precision Medicine. *Clinical and Translational Science*. 2019;12(1):3–5.
164. Yang X, Yao L, Fu X, Mukherjee R, Xia Q, Jakubowska MA, et al. Experimental Acute Pancreatitis Models: History, Current Status, and Role in Translational Research. *Frontiers in Physiology* [Internet]. 2020 [cited 2023 May 2];11. Available from: <https://www.frontiersin.org/articles/10.3389/fphys.2020.614591>
165. Mayer J, Rau B, Gansauge F, Beger HG. Inflammatory mediators in human acute pancreatitis: clinical and pathophysiological implications. *Gut*. 2000 Oct;47(4):546–52.
166. Zhang D, Wang T, Dong X, Sun L, Wu Q, Liu J, et al. Systemic Immune-Inflammation Index for Predicting the Prognosis of Critically Ill Patients with Acute Pancreatitis. *Int J Gen Med*. 2021;14:4491–8.
167. Bettaieb A, Xi Y, Hosein E, Coggins N, Bachaalany S, Wiede F, et al. Pancreatic T cell protein-tyrosine phosphatase deficiency ameliorates cerulein-induced acute pancreatitis. *Cell Communication and Signaling*. 2014 Mar 10;12(1):13.
168. Motoo Y, Watanabe H, Yamaguchi Y, Xie MJ, Mouri H, Ohtsubo K, et al. Pancreatitis-associated protein levels in pancreatic juice from patients with pancreatic diseases. *Pancreatology*. 2001;1(1):43–7.
169. Shen A, Kim HJ, Oh GS, Lee SB, Lee S, Pandit A, et al. Pharmacological stimulation of NQO1 decreases NADPH levels and ameliorates acute pancreatitis in mice. *Cell Death Dis*. 2018 Dec 18;10(1):5.
170. Arafa HMM, Hemeida RAM, Hassan MIA, Abdel-Wahab MH, Badary OA, Hamada FMA. Acetyl-L-carnitine ameliorates caerulein-induced acute pancreatitis in rats. *Basic Clin Pharmacol Toxicol*. 2009 Jul;105(1):30–6.
171. Ceyhan GO, Timm AK, Bergmann F, Günther A, Aghdassi AA, Demir IE, et al. Prophylactic glycine administration attenuates pancreatic damage and inflammation in experimental acute pancreatitis. *Pancreatology*. 2011;11(1):57–67.
172. Ferreira L, Pérez-González N, Llanillo M, Calvo JJ, Sánchez-Bernal C. Acute pancreatitis decreases pancreas phospholipid levels and increases susceptibility to lipid peroxidation in rat pancreas. *Lipids*. 2002 Feb;37(2):167–71.
173. Yang C, Fagan AM, Perrin RJ, Rhinn H, Harari O, Cruchaga C. Mendelian randomization and genetic colocalization infer the effects of the multi-tissue proteome on 211 complex disease-related phenotypes. *Genome Med*. 2022 Dec 12;14(1):140.
174. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Wörheide MA, et al. Mapping the proteo-genomic convergence of human diseases. *Science*. 2021 Nov 12;374(6569):eabj1541.
175. Folkersen L, Gustafsson S, Wang Q, Hansen DH, Hedman ÅK, Schork A, et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat Metab*. 2020 Oct;2(10):1135–48.

176. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrismisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021 Dec;53(12):1712–21.
177. Pietzner M, Wheeler E, Carrasco-Zanini J, Kerrison ND, Oerton E, Koprulu M, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun.* 2021 Nov 24;12(1):6822.
178. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, et al. The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Research.* 2022 Nov 18;gkac1046.
179. Hewitt J, Walters M, Padmanabhan S, Dawson J. Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. *BMJ Open.* 2016 Mar 1;6(3):e009161.
180. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology.* 2015 Aug 1;44(4):1137–47.
181. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature.* 2023 Jan;613(7944):508–18.
182. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013 Jan 1;29(1):15–21.
183. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics.* 2012 Jun 1;28(11):1530–2.
184. THE GTEx CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020 Sep 11;369(6509):1318–30.
185. Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet.* 2018 Sep;50(9):1335–41.
186. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010 Sep 1;26(17):2190–1.
187. Alghamdi J, Padmanabhan S. Chapter 12 - Fundamentals of Complex Trait Genetics and Association Studies. In: Padmanabhan S, editor. *Handbook of Pharmacogenomics and Stratified Medicine* [Internet]. San Diego: Academic Press; 2014 [cited 2023 May 4]. p. 235–57. Available from: <https://www.sciencedirect.com/science/article/pii/B9780123868824000128>
188. Bacanu SA, Devlin B, Roeder K. The Power of Genomic Control. *The American Journal of Human Genetics.* 2000 Jun 1;66(6):1933–44.
189. Ehret GB. Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. *Curr Hypertens Rep.* 2010 Feb;12(1):17–25.
190. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software.* 2018 May 19;3(25):731.
191. Staley J. gassocplot [Internet]. 2023 [cited 2023 May 3]. Available from: <https://github.com/jrs95/gassocplot>
192. Lee JJ, McGue M, Iacono WG, Chow CC. The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *Genet Epidemiol.* 2018 Dec;42(8):783–95.
193. Wang L, Balmat TJ, Antonia AL, Constantine FJ, Henao R, Burke TW, et al. An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome Medicine.* 2021 May 17;13(1):83.

194. Kraft P, Chen H, Lindström S. The Use Of Genetic Correlation And Mendelian Randomization Studies To Increase Our Understanding of Relationships Between Complex Traits. *Curr Epidemiol Rep.* 2020 Jun;7(2):104–12.
195. Bender R, Lange S. Adjusting for multiple testing--when and how? *J Clin Epidemiol.* 2001 Apr;54(4):343–9.
196. Newcombe PJ, Conti DV, Richardson S. JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. *Genetic Epidemiology.* 2016;40(3):188–201.
197. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020 Oct;52(10):1122–31.
198. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. *Nat Commun.* 2017 May 18;8(1):15452.
199. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc.* 2012 Feb 16;7(3):500–7.
200. Bartosek N, Vartanian M, Khalid Z, Shah B. Paliperidone-Induced Hypertriglyceridemia Resulting in Acute Interstitial Pancreatitis. *Prim Care Companion CNS Disord.* 2023 Apr 20;25(2):46709.
201. Möller K, Jenssen C, Braden B, Hocke M, Hollerbach S, Ignee A, et al. Pancreatic changes with lifestyle and age: What is normal and what is concerning? *Endoscopic Ultrasound.* 2023 Apr;12(2):213.
202. Möller K, Jenssen C, Ignee A, Hocke M, Faiss S, Iglesias-Garcia J, et al. Pancreatic duct imaging during aging. *Endoscopic Ultrasound.* 2023 Apr;12(2):200.
203. Monaghan B, Monaghan A, UI Ain Q, Duggan SN, Conlon KC, Gormley J. Significantly lower physical activity participation in individuals with chronic pancreatitis compared to controls: An exploratory study of objectively assessed physical activity levels. *Dig Surg.* 2023 Apr 21;
204. Khatua B, El-Kurdi B, Singh VP. Obesity and pancreatitis. *Curr Opin Gastroenterol.* 2017 Sep;33(5):374–82.
205. Ravi Kanth V, Nageshwar Reddy D. Genetics of acute and chronic pancreatitis: An update. *World J Gastrointest Pathophysiol.* 2014 Nov 15;5(4):427–37.
206. Lee PJ, Lahooti A, Culp S, Boutsicaris A, Holovach P, Wozniak K, et al. Obesity and alcoholic etiology as risk factors for multisystem organ failure in acute pancreatitis: Multinational study. *United European Gastroenterol J.* 2023 May;11(4):383–91.
207. Varrey P, RamanaMurthy SV. Acute Pancreatitis: Presenting in Different Clinical Settings. *J Assoc Physicians India.* 2023 Jan;71(1):1.
208. Waddell H, Stevenson TJ, Mole DJ. The role of the circadian rhythms in critical illness with a focus on acute pancreatitis. *Heliyon.* 2023 Apr;9(4):e15335.
209. Ghazanfar MA, Ke L, Ramsay G, Smith M, Giovinazzo F, Mohamed M, et al. Management of Splanchnic Vein Thrombosis in Patients With Acute Pancreatitis: An International Survey of Current Practice. *Pancreas.* 2022 Oct 1;51(9):1211–6.
210. Garg R, Mohammed A, Singh A, Siddiki H, Bhatt A, Sanaka MR, et al. Mortality Trends, Outcomes, and Predictors of Portal Vein Thrombosis in Acute Pancreatitis Patients: A Propensity-Matched National Study. *Dig Dis Sci.* 2023 Apr 25;
211. Hart PA, Bradley D, Conwell DL, Dungan K, Krishna SG, Wyne K, et al. Diabetes Mellitus following Acute Pancreatitis. *Lancet Gastroenterol Hepatol.* 2021 Aug;6(8):668–75.

212. Schepis T, Tringali A, Spada C, Costamagna G, Boškoski I. Intrapancreatic Fat Deposition: Cause or Consequence of First Acute Pancreatitis Attack? *Official journal of the American College of Gastroenterology | ACG*. 2023 May;118(5):910.
213. Petrov MS. Fatty change of the pancreas: the Pandora's box of pancreatology. *The Lancet Gastroenterology & Hepatology* [Internet]. 2023 Apr 21 [cited 2023 May 8];0(0). Available from: [https://www.thelancet-com.acces.bibl.ulaval.ca/journals/langas/article/PIIS2468-1253\(23\)00064-X/fulltext](https://www.thelancet-com.acces.bibl.ulaval.ca/journals/langas/article/PIIS2468-1253(23)00064-X/fulltext)
214. Maetzel H, Rutkowski W, Panic N, Mari A, Hedström A, Kulinski P, et al. Non-alcoholic fatty pancreas disease and pancreatic exocrine insufficiency: pilot study and systematic review. *Scand J Gastroenterol*. 2023 Apr 23;1–8.
215. Buch S, Schafmayer C, Völzke H, Becker C, Franke A, von Eller-Eberstein H, et al. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet*. 2007 Aug;39(8):995–9.
216. Liang KW, Huang HH, Wang L, Lu WY, Chou YH, Tantoh DM, et al. Risk of gallstones based on ABCG8 rs11887534 single nucleotide polymorphism among Taiwanese men and women. *BMC Gastroenterol*. 2021 Dec 14;21:468.
217. Berge KE, Tian H, Graf GA, Yu L, Grishin NV, Schultz J, et al. Accumulation of Dietary Cholesterol in Sitosterolemia Caused by Mutations in Adjacent ABC Transporters. *Science* [Internet]. 2000 Dec 1 [cited 2022 Feb 3]; Available from: <https://www.science.org/doi/abs/10.1126/science.290.5497.1771>
218. Jouini R, Vergès B. La sitostérolémie. *Médecine des Maladies Métaboliques*. 2016 Nov 1;10(7):650–3.
219. Liu MT, Yang HX. Neonatal hyperinsulinism with an ABCC8 mutation: A case report. *World J Clin Cases*. 2023 Apr 6;11(10):2254–9.
220. Scholz M, Horn K, Pott J, Gross A, Kleber ME, Delgado GE, et al. Genome-wide meta-analysis of phytosterols reveals five novel loci and a detrimental effect on coronary atherosclerosis. *Nat Commun*. 2022 Jan 10;13(1):143.
221. Nattagh-Eshtivani E, Barghchi H, Pahlavani N, Barati M, Amiri Y, Fadel A, et al. Biological and pharmacological effects and nutritional impact of phytosterols: A comprehensive review. *Phytotherapy Research*. 2022;36(1):299–322.
222. Salehi-Sahlabadi A, Varkaneh HK, Shahdadian F, Ghaedi E, Nouri M, Singh A, et al. Effects of Phytosterols supplementation on blood glucose, glycosylated hemoglobin (HbA1c) and insulin levels in humans: a systematic review and meta-analysis of randomized controlled trials. *J Diabetes Metab Disord*. 2020 Apr 19;19(1):625–32.
223. Prasad M, Jayaraman S, Eladi MA, El-Sherbiny M, Abdelrahman MAE, Veeraraghavan VP, et al. A Comprehensive Review on Therapeutic Perspectives of Phytosterols in Insulin Resistance: A Mechanistic Approach. *Molecules*. 2022 Feb 28;27(5):1595.
224. Gupta R, Sharma AK, Dobhal MP, Sharma MC, Gupta RS. Antidiabetic and antioxidant potential of  $\beta$ -sitosterol in streptozotocin-induced experimental hyperglycemia. *J Diabetes*. 2011 Mar;3(1):29–37.
225. Yoo EG. Sitosterolemia: a review and update of pathophysiology, clinical spectrum, diagnosis, and management. *Ann Pediatr Endocrinol Metab*. 2016 Mar;21(1):7–14.
226. Tada H, Okada H, Nomura A, Takamura M, Kawashiri M aki. Beneficial effect of ezetimibe-atorvastatin combination therapy in patients with a mutation in ABCG5 or ABCG8 gene. *Lipids Health Dis*. 2020 Jan 4;19:3.

227. Feraille E, Krempf M, Charbonnel B, Bouhour JB, Nicolas G. Hypertension artérielle de l'obèse. Rôle de l'hyperinsulinisme et de l'insulinorésistance. *La Revue de Médecine Interne*. 1990 Jul 1;11(4):293–6.
228. Masson E. EM-Consulte. [cited 2023 May 9]. L'hypertension artérielle au cours de l'insulinothérapie du diabétique de type 2. Available from: <https://www.em-consulte.com/article/130276/l-hypertension-arterielle-au-cours-de-l-insulinoth>
229. Yoshida S, Takakuwa S, Kihira H, Nishio Y, Haneda M. Malignant Hypertension Complicated with Necrotizing Pancreatitis After Starting Treatment: A Case Report. *Am J Case Rep*. 2022 Feb 28;23:e935271-1-e935271-6.
230. Wang Z, Wang Q, Song J, Li M, Wang F, Chen B, et al. Treatment of acute pancreatitis with pancreatic duct decompression via ERCP: A case report series. *Experimental and Therapeutic Medicine*. 2020 Sep 1;20(3):2593–8.
231. Méndez-González J, Julve J, Rotllan N, Llaverias G, Blanco-Vaca F, Escolà-Gil JC. ATP-binding cassette G5/G8 deficiency causes hypertriglyceridemia by affecting multiple metabolic pathways. *Biochim Biophys Acta*. 2011 Dec;1811(12):1186–93.
232. Chen JM, Mercier B, Audrezet MP, Ferec C. Mutational analysis of the human pancreatic secretory trypsin inhibitor (PSTI) gene in hereditary and sporadic chronic pancreatitis. *Journal of Medical Genetics*. 2000 Jan 1;37(1):67–9.
233. Witt H, Sahin-Tóth M, Landt O, Chen JM, Kähne T, Drenth JP, et al. A degradation-sensitive anionic trypsinogen (PRSS2) variant protects against chronic pancreatitis. *Nat Genet*. 2006 Jun;38(6):668–73.
234. Shelton C, Solomon S, LaRusch J, Whitcomb DC. PRSS1-Related Hereditary Pancreatitis. In: Adam MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJ, Gripp KW, et al., editors. *GeneReviews®* [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cited 2023 May 8]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK84399/>
235. Wan J, Haddock A, Edenfield B, Ji B, Bi Y. Transgenic expression of human PRSS2 exacerbates pancreatitis in mice. *Gut*. 2020 Nov 1;69(11):2051–2.
236. Zhou J, Sahin-Tóth M. Chymotrypsin C (CTRC) mutations in chronic pancreatitis. *J Gastroenterol Hepatol*. 2011 Aug;26(8):1238–46.
237. Chen JM, Férec C. Chapter 577 - Cationic Trypsin (Human). In: Rawlings ND, Salvesen G, editors. *Handbook of Proteolytic Enzymes (Third Edition)* [Internet]. Academic Press; 2013 [cited 2023 May 9]. p. 2609–14. Available from: <https://www.sciencedirect.com/science/article/pii/B9780123822192005779>
238. Oracz G, Zaród M, Ewers M, Laumen H, Gambin T, Kamiński P, et al. Loss of function TRPV6 variants are associated with chronic pancreatitis in nonalcoholic early-onset Polish and German patients. *Pancreatology*. 2021 Dec 1;21(8):1434–42.
239. Goma M, Hagiwara SI, Wada T, Maeyama T, Okamoto N, Ishii S, et al. A case of early-onset idiopathic chronic pancreatitis associated with a loss-of-function TRPV6 p.R483Q variant successfully treated by pancreatic duct stenting. *Clin J Gastroenterol*. 2023 Apr 29;
240. Peng JB, Suzuki Y, Gyimesi G, Hediger MA. TRPV5 and TRPV6 Calcium-Selective Channels. In: Kozak JA, Putney JW, editors. *Calcium Entry Channels in Non-Excitable Cells* [Internet]. Boca Raton (FL): CRC Press/Taylor & Francis; 2018 [cited 2022 Jan 17]. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK531440/>
241. Woudenberg-Vrenken TE, Lameris AL, Weißgerber P, Olausson J, Flockerzi V, Bindels RJM, et al. Functional TRPV6 channels are crucial for transepithelial Ca<sup>2+</sup> absorption. *Am J Physiol Gastrointest Liver Physiol*. 2012 Oct;303(7):G879-885.

242. Gerasimenko JV, Gerasimenko OV, Petersen OH. The role of Ca<sup>2+</sup> in the pathophysiology of pancreatitis. *J Physiol*. 2014 Jan 15;592(2):269–80.
243. Gerasimenko JV, Peng S, Tsugorka T, Gerasimenko OV. Ca<sup>2+</sup> signalling underlying pancreatitis. *Cell Calcium*. 2018 Mar;70:95–101.
244. Krüger B, Albrecht E, Lerch MM. The role of intracellular calcium signaling in premature protease activation and the onset of pancreatitis. *Am J Pathol*. 2000 Jul;157(1):43–50.
245. Fang X, Cai Y, Liu J, Wang Z, Wu Q, Zhang Z, et al. Twist2 contributes to breast cancer progression by promoting an epithelial-mesenchymal transition and cancer stem-like cell self-renewal. *Oncogene*. 2011 Nov 24;30(47):4707–20.
246. Zhang Y, Guo F, Li S, Wang F, Meng Z, Zhao J, et al. Decreased high density lipoprotein cholesterol is an independent predictor for persistent organ failure, pancreatic necrosis and mortality in acute pancreatitis. *Sci Rep*. 2017 Aug 14;7:8064.
247. Mudry JM, Massart J, Szekeres FLM, Krook A. TWIST1 and TWIST2 regulate glycogen storage and inflammatory genes in skeletal muscle. *J Endocrinol*. 2015 Mar;224(3):303–13.
248. Yang CC, Dai CY, Luo KH, Lee KW, Wu CH, Hung CH, et al. Single Nucleotide Polymorphism of TWIST2 May Be a Modifier for the Association between High-Density Lipoprotein Cholesterol and Blood Lead (Pb) Level. *Int J Environ Res Public Health*. 2022 Jan 26;19(3):1352.
249. Zhang Y, Ge X, Li Y, Zhang B, Wang P, Hao M, et al. TWIST2 and the PPAR signaling pathway are important in the progression of nonalcoholic steatohepatitis. *Lipids Health Dis*. 2021 Apr 20;20(1):39.
250. Song Y, Zhang W, Zhang J, You Z, Hu T, Shao G, et al. TWIST2 inhibits EMT and induces oxidative stress in lung cancer cells by regulating the FGF21-mediated AMPK/mTOR pathway. *Exp Cell Res*. 2021 Aug 1;405(1):112661.
251. Lanthier N. Les nouveaux traitements de la stéatohépatite non-alcoolique. *Nutrition Clinique et Métabolisme*. 2020 Oct;34(3):216–22.
252. Coate KC, Hernandez G, Thorne CA, Sun S, Le TDV, Vale K, et al. FGF21 Is an Exocrine Pancreas Secretagogue. *Cell Metab*. 2017 Feb 7;25(2):472–80.
253. Antonucci L, Fagman JB, Kim JY, Todoric J, Gukovsky I, Mackey M, et al. Basal autophagy maintains pancreatic acinar cell homeostasis and protein synthesis and prevents ER stress. *Proc Natl Acad Sci USA* [Internet]. 2015 Nov 10 [cited 2023 May 8];112(45). Available from: <https://pnas.org/doi/full/10.1073/pnas.1519384112>
254. Giri AK, Midha S, Banerjee P, Agrawal A, Mehdi SJ, Dhingra R, et al. Common Variants in CLDN2 and MORC4 Genes Confer Disease Susceptibility in Patients with Chronic Pancreatitis. *PLoS One*. 2016 Jan 28;11(1):e0147345.
255. Tencer AH, Cox KL, Wright GM, Zhang Y, Petell CJ, Klein BJ, et al. Molecular mechanism of the MORC4 ATPase activation. *Nat Commun*. 2020 Oct 29;11(1):5466.
256. Seltsam K, Pentner C, Weigl F, Sutedjo S, Zimmer C, Beer S, et al. Sequencing of the complex CTRB1-CTRB2 locus in chronic pancreatitis. *Pancreatol*. 2020 Dec 1;20(8):1598–603.
257. Gold-Smith FD, Singh RG, Petrov MS. Elevated Circulating Levels of Motilin are Associated with Diabetes in Individuals after Acute Pancreatitis. *Exp Clin Endocrinol Diabetes*. 2020 Jan;128(01):43–51.
258. OWYANG C, FUNAKOSHI A, VINIK AI. Evidence for Modulation of Motilin Secretion by Pancreatico-Biliary Juice in Health and in Chronic Pancreatitis\*. *The Journal of Clinical Endocrinology & Metabolism*. 1983 Nov 1;57(5):1015–20.
259. Portincasa P, Moschetta A, Palasciano G. Cholesterol gallstone disease. *The Lancet*. 2006 Jul 15;368(9531):230–9.

260. Vitek L, Carey MC. New pathophysiological concepts underlying pathogenesis of pigment gallstones. *Clin Res Hepatol Gastroenterol*. 2012 Apr;36(2):122–9.
261. Buch S, Schafmayer C, Völzke H, Seeger M, Miquel JF, Sookoian SC, et al. Loci from a genome-wide analysis of bilirubin levels are associated with gallstone risk and composition. *Gastroenterology*. 2010 Dec;139(6):1942-1951.e2.
262. Stender S, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Extreme Bilirubin Levels as a Causal Risk Factor for Symptomatic Gallstone Disease. *JAMA Internal Medicine*. 2013 Jul 8;173(13):1222–8.
263. Lin JP, O'Donnell CJ, Schwaiger JP, Cupples LA, Lingenhel A, Hunt SC, et al. Association Between the UGT1A1\*28 Allele, Bilirubin Levels, and Coronary Heart Disease in the Framingham Heart Study. *Circulation*. 2006 Oct 3;114(14):1476–81.
264. Wu TW, Fung KP, Wu J, Yang CC, Weisel RD. Antioxidation of human low density lipoprotein by unconjugated and conjugated bilirubins. *Biochemical Pharmacology*. 1996 Mar 22;51(6):859–62.
265. Mortensen MB, Dzaye O, Bøtker HE, Jensen JM, Maeng M, Bentzon JF, et al. Low-Density Lipoprotein Cholesterol Is Predominantly Associated With Atherosclerotic Cardiovascular Disease Events in Patients With Evidence of Coronary Atherosclerosis: The Western Denmark Heart Registry. *Circulation*. 2023 Apr 4;147(14):1053–63.
266. Loh NY, Minchin JEN, Pinnick KE, Verma M, Todorčević M, Denton N, et al. RSPO3 impacts body fat distribution and regulates adipose cell biology in vitro. *Nat Commun*. 2020 Jun 3;11(1):2797.
267. Nagano K. R-spondin signaling as a pivotal regulator of tissue development and homeostasis. *Jpn Dent Sci Rev*. 2019 Nov;55(1):80–7.
268. Yang RB, Ng CKD, Wasserman SM, Colman SD, Shenoy S, Mehraban F, et al. Identification of a novel family of cell-surface proteins expressed in human vascular endothelium. *J Biol Chem*. 2002 Nov 29;277(48):46364–73.
269. Martínez-Hernández A, Córdova EJ, Rosillo-Salazar O, García-Ortíz H, Contreras-Cubas C, Islas-Andrade S, et al. Association of HMOX1 and NQO1 Polymorphisms with Metabolic Syndrome Components. *PLOS ONE*. 2015 May 1;10(5):e0123313.
270. Ahmed MH, Ali A. Nonalcoholic fatty liver disease and cholesterol gallstones: which comes first? *Scand J Gastroenterol*. 2014 May;49(5):521–7.
271. O'Dea LSL, MacDougall J, Alexander VJ, Digenio A, Hubbard B, Arca M, et al. Differentiating Familial Chylomicronemia Syndrome From Multifactorial Severe Hypertriglyceridemia by Clinical Profiles. *J Endocr Soc*. 2019 Oct 11;3(12):2397–410.
272. Volanesorsen and Triglyceride Levels in Familial Chylomicronemia Syndrome | *NEJM* [Internet]. 2022 [cited 2022 Jul 22]. Available from: [https://www.nejm.org/doi/10.1056/NEJMoa1715944?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%20pubmed](https://www.nejm.org/doi/10.1056/NEJMoa1715944?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed)
273. Ahmad Z, Banerjee P, Hamon S, Chan KC, Bouzelmat A, Sasiela WJ, et al. Inhibition of Angiopoietin-Like Protein 3 With a Monoclonal Antibody Reduces Triglycerides in Hypertriglyceridemia. *Circulation*. 2019 Aug 6;140(6):470–86.
274. Chiang JYL, Ferrell JM. Up to date on cholesterol 7 alpha-hydroxylase (CYP7A1) in bile acid synthesis. *Liver Res*. 2020 Jun;4(2):47–63.
275. Sanyal A, Charles ED, Neuschwander-Tetri BA, Loomba R, Harrison SA, Abdelmalek MF, et al. Pegbelfermin (BMS-986036), a PEGylated fibroblast growth factor 21 analogue, in patients with non-alcoholic steatohepatitis: a randomised, double-blind, placebo-controlled, phase 2a trial. *Lancet*. 2019 Dec 22;392(10165):2705–17.

276. Hernandez G, Luo T, Javed TA, Wen L, Kalwat MA, Vale K, et al. Pancreatitis is an FGF21-deficient state that is corrected by replacement therapy. *Sci Transl Med*. 2020 Jan 8;12(525):eaay5186.
277. Poh W, Wong W, Ong H, Aung MO, Lim SG, Chua BT, et al. Klotho-beta overexpression as a novel target for suppressing proliferation and fibroblast growth factor receptor-4 signaling in hepatocellular carcinoma. *Molecular Cancer*. 2012 Mar 23;11(1):14.
278. Zhang M, Haughey M, Wang NY, Blease K, Kapoun AM, Couto S, et al. Targeting the Wnt signaling pathway through R-spondin 3 identifies an anti-fibrosis treatment strategy for multiple organs. *PLOS ONE*. 2020 Mar 11;15(3):e0229445.
279. Klöppel G, Detlefsen S, Feyerabend B. Fibrosis of the pancreas: the initial tissue damage and the resulting pattern. *Virchows Arch*. 2004 Jul;445(1):1–8.
280. Huang C, Iovanna J, Santofimia-Castaño P. Targeting Fibrosis: The Bridge That Connects Pancreatitis and Pancreatic Cancer. *Int J Mol Sci*. 2021 May 7;22(9):4970.
281. Pérez S, Pereda J, Sabater L, Sastre J. Redox signaling in acute pancreatitis. *Redox Biology*. 2015 Aug 1;5:1–14.
282. Zhang J, Bai J, Zhou Q, Hu Y, Wang Q, Yang L, et al. Glutathione prevents high glucose-induced pancreatic fibrosis by suppressing pancreatic stellate cell activation via the ROS/TGFβ/SMAD pathway. *Cell Death Dis*. 2022 May 6;13(5):1–13.
283. NCATS Inxight Drugs — VATIQUINONE [Internet]. [cited 2023 May 10]. Available from: <https://drugs.ncats.io/drug/6O85FK9I0X>
284. Solinas C, Aiello M, Rozali E, Lambertini M, Willard-Gallo K, Migliori E. Programmed Cell Death-Ligand 2: A Neglected But Important Target in the Immune Response to Cancer? *Transl Oncol*. 2020 Jul 1;13(10):100811.
285. Definition of B7-DC cross-linking antibody rHIgM12B7 - NCI Drug Dictionary - NCI [Internet]. 2011 [cited 2023 May 10]. Available from: <https://www.cancer.gov/publications/dictionaries/cancer-drug/def/b7-dc-crosslinking-antibody-rhigm12b7>
286. Giorda CB, Nada E, Tartaglino B, Marafetti L, Gnani R. A systematic review of acute pancreatitis as an adverse event of type 2 diabetes drugs: from hard facts to a balanced position. *Diabetes, Obesity and Metabolism*. 2014 Nov;16(11):1041–7.