From the Department of Medicine, Huddinge
Karolinska Institutet, Stockholm, Sweden

GENETIC AND EPIGENETIC DIAGNOSTICS OF VIRAL SUSCEPTIBILITY AND INBORN ERRORS OF IMMUNITY

Laura Covill

Karolinska Institutet

Stockholm 2024

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2024

ISBN 978-91-8017-817-4
DOI https://doi.org/10.69622/27152478
Cover illustration: Finding the idiomatic needle in a haystack.

Genetic and Epigenetic Diagnostics of Viral Susceptibility and Inborn Errors of Immunity

Thesis for Doctoral Degree (Ph.D.)

By

Laura Elaine Covill

The thesis will be defended in public at Erna Möllersalen, NEO, Karolinska Institute

November 29th 2024 at 1pm

Principal Supervisor:
Professor Yenan Bryceson
Karolinska Institute
Department of Medicine, Huddinge
Center for Hematology and Regenerative Medicine

Co-supervisors:
Associate Professor Mikael Sundin
Karolinska Institute
Department of Clinical Science, Intervention and Technology

Professor Anna Wedell
Karolinska Institute
Department of Molecular Medicine and Surgery

Dr Beatrice Zitti
University of Geneva
Department of Pathology and Immunology

Opponent:
Professor Sophie Hambleton
Newcastle University
Department of Translational & Clinical Resarch Institute
Faculty of Medical Sciences

Examination Board:
Professor Karin Loré
Karolinska Institute
Department of Medicine, Solna

Associate Professor Linda Holmfeldt
Uppsala University
Department of Immunology, Genetics & Pathology

Associate Professor Carsten Daub
Karolinska Institute
Department of Medicine, Huddinge
Division of Biosciences and Nutrition

"In the game of life and evolution, there are three players at the table: human beings, nature, and machines. I am firmly on the side of nature. But nature, I suspect, is on the side of the machines."

- George Dyson, *Darwin Among the Machines*

Popular science summary of the thesis

A surprisingly high proportion of the general public has a rare genetic disease – aren't they supposed to be rare? But although each individual disease has very few people affected, all the diseases taken together affect around 5% of all people – about 350 million patients. With so many people needing care, it's important to make sure diagnosis happens quickly and efficiently. Unfortunately, this is usually not the case. Instead, getting the disease diagnosis which will lead to the best treatment never happens for most people with a rare disease. If one does materialize, it is an arduous process often taking years and many hospital visits. The journey to a diagnosis is so long, it has frequently been termed the 'diagnostic odyssey'.

One reason the process takes so long is that looking through the genome to find a single mutation causing a genetic disease, out of millions of differences which naturally exist between humans, is like trying to find a needle in a haystack. Our current technologies give results that may not always show which mutation is causing the problem.

The recent COVID-19 pandemic showed some of the problems which can result from having an undiagnosed rare disease. Many people who would easily shrug off an infection with most viruses were extremely ill or died. We know now that some of these cases were due to underlying genetic problems with their immune systems – but which ones? Understanding this properly gives healthcare systems the best chance of giving the best treatment to those people.

This thesis looks at three different ways to diagnose rare disease patients. First, we examined patients extremely ill with COVID-19 as a group, to find out via statistics whether mutations in particular parts of the immune system might influence whether someone is more likely to get sick. Secondly, for patients with very dangerous levels of inflammation, we show that a new method of looking cells from patients or healthy people can pick out the patients who have a genetic mutation in a set of genes known to cause this condition. Thirdly, the thesis outlines a way to tackle the situation where the disease-causing mutation is somewhere in the list, but it's a kind of mutation which has unknown significance. It might be in a genomic region where the function is uncertain, or changes between different cell types. The mutation could be far away from the gene it affects, so the connection is never made. We can prioritise these kinds of mutation by using extra data gained from looking at cells taken from the patient, and

examining different stages of the process by which proteins are made in the cells using instructions encoded in DNA. The additional biological data of knowing when, how, and from where the instructions are being given, gives the extra layer of functional interpretation needed to properly rank these mutations, which were difficult to interpret from 3 billion letters of code and no other information. Armed with the correct information about the activity of these mutations in patient cells, it will be much easier for doctors to find the right diagnosis for each patient.

Abstract

In inborn errors of immunity (IEI), as in other monogenic rare diseases, rate of diagnosis has failed to improve beyond 35% despite accessibility of whole genome sequencing for many patients. Here, efforts to improve diagnostics of patients with severe inflammatory diseases in different contexts are presented.

We performed three studies on groups of patients who suffered critical COVID-19 between 2019-2022. In **Paper I**, we described several patients with an IEI of type I interferon (IFN) which presented unexpectedly late in life after infection with SARS-CoV-2. We diagnosed two patients with biallelic deficiency of IRF7, and characterized the clinical and immunologic phenotypes of seven patients in total. This marks the first time this has been possible in more than an orphan case for this IEI. We further suggested that, based on investigations in two patients, enhanced T cell responses may offer some compensatory immunity for the reduction of type I interferon response. In **Paper II**, we evaluated a cohort from a single intensive care unit who suffered from critical COVID-19 despite relative health and youth prior to infection. We identified two cases caused by presence of autoantibodies to type I IFN, preventing IFN signalling. We investigated incidence of rare variants in the type I IFN production and signalling pathways compared to controls, as well as screening for rare variants in other IEI genes. We followed up with immunological assays to functionally validate identified variants, and noting a trend towards lower responses in type I IFN signalling in three of six individuals carrying heterozygous very rare, or homozygous rare variants. **Paper III** investigated genetics of hyperinflammation in COVID-19 patients. Haemophagocytic lymphohistiocytosis (HLH) is a hyperinflammatory syndrome with some clinical features similar to the cytokine storm experienced by many critical COVID-19 patients, suggesting that variation in cytotoxicity genes associated with HLH may predispose to cytokine storm if triggered by viral infection. Here, we took advantage of large datasets of critical patients and asymptomatic infected provided by the COVID Human Genetic Effort. With thousands of genomes available, we established that variants in genes associated with HLH were not more frequent in critical disease, arguing against a strong pathogenetic overlap.

Paper IV described the increased efficacy of a set of diagnostic flow cytometry assays piloted by our lab and reported previously. By evaluating our improved assays in almost 500 donors, including 92 with genetically diagnosed primary HLH,

we confirmed that our assays reliably separate data generated from patient samples from control healthy donors or patients with other IEIs not affecting lymphocyte cytotoxicity. Our assays also showed improvement in separating degranulation defects from controls compared to the assays which are the current standard in most laboratories.

Finally, **Paper V** details MAGNET, a novel algorithm for approaching patients who have a predicted rare disease, but for whom whole genome sequencing in a clinical context has been unsuccessful. Using integrated ATAC-seq with patient and parental genomes, and supported by RNA-seq, we analysed two proof-of-concept patients previously diagnosed with HLH and were able to prioritise the variants outside of protein coding regions which cause their disease from genome-wide screening. We optimised this workflow with studies of allelic chromatin accessibility in patients and controls, and identified a strong candidate variant in an IEI patient still unexplained in the clinic.

Taken together, these studies have increased our understanding of the genetic aetiology of IEIs, and the speed and rate with which we can diagnose patients suffering these conditions.

List of scientific papers

I. **Respiratory viral infections in otherwise healthy humans with inherited IRF7 deficiency**
Campbell TM, Liu Z, Zhang Q, Moncada-Velez M, **Covill LE**, Zhang P, Darazam IA, Bastard P, Bizien L, Bucciol G, Lind Enoksson S, Jouanguy E, Karabela SN, Khan T, Kendir-Demirkol Y, Arias AA, Mansouri D, Marits P, Marr N, Migeotte I, Moens L, Ozcelik T, Pellier I, Sendel A, Şenoğlu S, Shahrooei M, Smith CIE, Vandernoot I, Willekens K, Yaşar KK, Bergman P, Abel L, Cobat A, Casanova JL, Meyts I, Bryceson YT
J Exp Med., 219(7):e20220202 (2022)

II. **Evaluation of genetic or cellular impairments in Type I IFN immunity in a cohort of young adults with critical COVID-19**
Covill LE*, Sendel A*, Campbell TM*, Piiroinen I, Lind Enoksson S, Wahren Borgström E, Hansen S, Ma K, Marits P, Norlin AC, Smith CIE, Kåhlin J, Eriksson LI, Bergman P§, Bryceson YT§
J Clin Immunol. 44(2):50 (2024)

III. **No association between HLH-associated gene variants and life-threatening COVID-19**
Covill LE, Cobat A, COVID Human Genetic Effort, Zhang Q, Bryceson YT
Manuscript

IV. **Efficacy of T cell exocytosis assays for the diagnosis of primary defects in cytotoxic lymphocyte exocytosis**
Chiang SCC, **Covill LE***, Tesi B*, Campbell TM, Schlums H, Nejati-Zendegani J, Mördrup K, Wood S, Theorell J, Sekine T, Al-Herz W, Akar HH, Belen FB, Chan MY, Devecioglu O, Aksu T, Ifversen M, Malinowska I, Sabel M, Unal E, Unal S, Introne W, Krzewski K, Gilmour KC, Ehl S, Ljunggren HG, Nordenskjöld M, Horne AC, Henter JI, Meeths M, Bryceson YT
Blood. 144(8):873-887 (2024)

V. **ATAC-seq from affected cell types robustly prioritizes pathogenic non-coding variants in patients with monogenic disease**
Covill LE, Campbell TM, Holmes TD, Frengen NS, Gustafsson C, Winroth A, Hauenstein J, Vinay Pandey R, Erichsen HC, Redzic D, Yoke CM, Vonlanthen S, Månsson R, Wedell A, Meeths M, Sundin M, Bryceson YT
Manuscript

Related scientific papers not included in the thesis

I. **X-linked recessive TLR7 deficiency in ~1% of men under 60 years old with life-threatening COVID-19**
Asano T, Boisson B, Onodi F, Matuozzo D, Moncada-Velez M, Renkilaraj MRLM, Zhang P, Meertens L, Bolze A, Materna M, Korniotis S, Gervais A, Talouarn E, Bigio B, Seeleuthner Y, Bilguvar K, Zhang Y, Neehus AL, Ogishi M, Pelham SJ, Le Voyer T, Rosain J, Philippot Q, Soler-Palacin P, Colobran R, Martin-Nalda A, Riviere JG, Tandjaoui-Lambiotte Y, Chaibi K, Shahrooei M, Darazam IA, Olyaei NA, Mansouri D, Hatipoglu N, Palabiyik F, Ozcelik T, Novelli G, Novalli A, Casari G, Aiuti A, Carrera P, Bondesan S, Barzaghi F, Rovere-Querini P, Tresoldi C, Franco JL, Rojas J, Reyes LF, Bustos IG, Arias AA, Morelle G, Kyheng K, Troya J, Planas-Serra L, Schlüter A, Gut M, Pujol A, Allende LM, Rodriguez-Gallego C, Flores C, Cabrera-Marante O, Pleguezuelo DE, Perez de Diego R, Keles S, Aytekin G, Akcan OM, Bryceson YT, Bergman P, Brodin P, Smole D, Smith CIE, Norlin AC, Campbell TM, **Covill LE**, Hammarström L, Pan-Hammarström Q, Abolhassani H, Mane S, Marr N, Ata M, al Ali F, Khan T, Spaan AN, Dalgard CL, Bonfanti P, Biondi A, Tubiana S, Burdet C, Nussbaum R, Kahn-Kirby A, Snow AL, COVID Human Genetic Effort, COVID-STORM Clinicians, COVID Clinicians, Imagine COVID Group, French COVID Cohort Study Group, CoV-Contact Group, Amsterdam UMC Covid-19 Biobank, NIAID-USUHS COVID Study Group, Bustamante J, Puel A, Boisson-Dupuis S, Zhang SY, Beziat V, Lifton RP, Bastard P, Notarangelo LD, Abel L, Su HC, Jouanguy E, Amara A, Soumelis V, Cobat A, Zhang Q, Casanova JL
Sci Immunol., 6:eabl4348 (2021)

II. **Rare predicted loss-of-function variants of type I IFN immunity genes are associated with life-threatening COVID-19**
Matuozzo D, Talouarn E, Marchal A, Zhang P, Manry J, Seeleuthner Y, Zhang Y, Bolze A, Chaldebas M, Milisavljevic B, Gervais A, Bastard P, Asano T, Bizien L, Barzaghi F, Abolhassani H, Abou Tayoun A, Aiuti A, Darazam AI, Allende LM, Alonso-Arias R, Arias AA, Aytekin G, Bergman P, Bondesan S, Bryceson YT, Bustos IG, Cabrera-Marante O, Carcel S, Carrera P, Casari G, Chaibi K, Colobran R, Condino-Neto A, **Covill LE**, Delmonte OM, El Zein L, Flores C, Gregersen PK, Gut M, Haerynck F, Halwani R, Hancerli S, Hammarström L, Hatipoglu N, Karbuz A, Keles S, Kyheng C, Leon-Lopez R, Franco JL, Mansouri D, Martinez-Picado J, Akcan OM, Migeotte I, Morange PE, Morelle G, Martin-Nalda A, Novelli G, Novelli A, Ozcelik T, Palabiyik F, Pan-Hammarström Q, Perez de Diego R, Planas-Serra L, Pleguezuelo DE, Prando C, Pujol A, Reyes LF, Riviere JG, Rodriguez-Gallego C, Rojas J, Rovere-Querini P, Schlüter

A, Shahrooei M, Sobh A, Soler-Palacin P, Tandjaoui-Lambiotte Y, Tipu I, Tresoldi C, Troya J, van de Beek D, Zatz M, Zawadzki P, Zaid al-Muhsen S, Alosaimi MF, Alsohime FM, Baris-Feldman H, Butte MJ, Constantinescu SN, Cooper MA, Dalgard CL, Fellay J, Heath JR, Lau YL, Lifton RP, Maniatis T, Mogensen TH, von Bernuth H, Lermine A, Vidaud M, Boland A, Deleuze JF, Nussbaum R, Kahn-Kirby A, Mentre F, Tubiana S, Gorochov G, Tubach F, Hausfater P, COVID-STORM Clinicians, Meyts I, Zhang SY, Puel A, Notarangelo LD, Boisson-Dupuis S, Su HC, Boisson B, Jouanguy E, Casanova JL, Zhang Q, Abel L, Cobat A
Genome Med., 15(1):22 (2023).

Additional scientific papers

I. **Hobit identifies tissue-resident memory T cell precursors that are regulated by Eomes**
Parga-Vidal L, Behr FM, Kragten NAM, Nota B, Wesselink TH, Kavazović I, **Covill LE**, Schuller MBP, Bryceson YT, Wensveen FM, van Lier RAW, van Dam TJP, Stark R, van Gisbergen KPJM
Science Immunology 6(62)eabg3533 (2021)

II. **Deficiency in the SNARE protein SYNTAXIN-11 causes a secondary B cell defect**
Kögl T, Chang HF, Staniek J, Chiang SCC, Thoulass G, Lao J, Weißert K, Dettmer-Monaco V, Geiger K, Manna PT, Beziat V, Momenilandi M, Tu SM, Keppler SJ, Pattu V, Wolf P, Kupferschmid L, Tholen S, **Covill LE**, Ebert K, Straub T, Groß M, Gather R, Engel H, Salzer U, Schell C, Maier S, Lehmberg K, Cornu T, Pircher H, Shahrooei M, Parvaneh N, Elling R, Rizzi M, Bryceson YT, Ehl S, Aichele P, Ammann S
J Exp Med 221(7):e20221122 (2024)

III. **NK cell and Monocyte Dysfunction in Multisystem Inflammatory Syndrome in Children**
Dick JK, Sangala JA, Venkatramana KD, Khaimraj A, Hamel L, Erickson SM, Hicks D, Soigner Y, **Covill LE**, Johnson A, Ehrhardt MJ, Ernste K, Brodin P, Koup RA, Khaitan A, Baehr C, Thielen BK, Henzler C, Skipper C, Miller JS, Bryceson YT, Wu J, John CC, Panoskaltsis-Mortari A, Orioles A, Steiner ME, Cheeran MCJ, Pravetoni M, Hart GT
J Immunol ji2400395 (2024)

Contents

1 Background 1
1.1 Diagnostics in the age of high-throughput sequencing 1
1.2 Inborn errors of immunity 3
1.2.1 Impairments of cytotoxic lymphocytes 4
1.2.2 Genetics of familial HLH 4
1.2.3 HLH manifestations in other IEIs 6
1.2.4 Molecular assays evaluating lymphocyte cytotoxicity 7
1.2.5 Personalised medicine in HLH treatment 7
1.3 COVID-19 pandemic 8
1.3.1 Presence of monogenic disease in critical COVID-19 9
1.3.2 Contribution of common risk loci to severe viral infection 10
1.4 Functional genomics 11
1.4.1 Epigenetic regulation and non-coding variation in disease: approaches for diagnosing the 'missing' 60% 11
1.4.2 Open-access resources 12
1.4.3 Enhancers and promoters 12
1.4.4 Alterations to intron boundaries 14
1.4.5 Disruption or introduction of intronic branchpoints 15
1.4.6 Regulation of additional alternative splicing machinery 17
1.4.7 Inclusion of a poison exon 18
1.4.8 Deletion of TAD boundaries 18
1.4.9 Untranslated region variation 20
1.4.10 Non-coding repeat expansions 20
1.4.11 Variation in non-coding transcripts 21
1.5 Monoallelic expression 22
1.5.1 Imprinting 22
1.5.2 Imprinting disorders 23
1.5.3 Random monoallelic expression 23
2 Research aims 25
3 Materials and methods 27
3.1 Cohorts 27
3.1.1 CovPID20 cohort 27
3.1.2 Type I IFN deficiency patients 27
3.1.3 COVID Human Genetic Effort cohort 27
3.1.4 Cytotoxic lymphocyte deficiency patients 27

3.1.5 'Proof-of-concept' FHL3 patients 28
3.1.6 Patients with unknown IEIs 28
3.2 Molecular assays 28
3.3 Computational analyses 29
3.3.1 Public datasets 29
3.3.2 Private datasets 30
3.3.3 Computing resources 30
3.3.4 Ancestry principal component analysis 31
3.3.5 Polygenic risk score calculation 31
3.3.6 Odds ratio calculations 31
3.3.7 Variant co-occurrence analysis 32
3.3.8 ROC curve generation 32
3.3.9 NGS data preprocessing 32
3.3.10 MAGNET diagnostic workflow 34
3.4 Ethical considerations 35
4 Results 39
4.1 Rare inborn errors of immunity cause life-threatening COVID-19 39
4.2 Contribution of Type I IFN variation to critical COVID-19 susceptibility 41
4.3 Contribution of hyperinflammatory gene variation to critical COVID-19 susceptibility 44
4.4 Stimulation with anti-CD3 and anti-CD16 is an effective diagnostic platform for defective cytotoxic lymphocyte degranulation 46
4.5 Integration of Omics data improves diagnostic rate in IEIs 48
5 Conclusions 51
6 Points of perspective 53
7 Acknowledgements 55
8 Declaration about the use of generative AI 59
9 References 61

List of abbreviations

1kGP	1000 Genomes Project
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
AR	Autosomal recessive
ARDS	Acute respiratory distress syndrome
ATAC-seq	Assay for transposase-accessible chromatin sequencing
AUROC	Area under the ROC curve
bam	Binary alignment map
bp	Base pair
BP	Branchpoint
CHGE	COVID Human Genetic Effort
CHS	Chediak-Higashi syndrome
ChIP-seq	Chromatin immunoprecipitation sequencing
CMV	Cytomegalovirus
CNV	Copy number variant
CRISPR	Clustered regularly interspaced short palindromic repeats
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
ENCODE	Encyclopaedia of DNA Elements
eQTL	Expression quantitative trait loci
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencers
FANTOM5	Functional annotation of the mammalian genome 5
FHL	Familial haemophagocytic lymphohistiocytosis

GATK	Genome analysis tool-kit
GERP	Genomic evolutionary rate profiling
GM-CSF	Granulocyte-macrophage colony stimulating factor
GnomAD	Genome Aggregation Database
GoF	Gain-of-function
GS2	Griscelli syndrome type 2
GTEx	Genotype-tissue expression
GWAS	Genome-wide association study
H3K4me1	Histone-3 lysine-4 monomethylation
H3K4me3	Histone-3 lysine-4 trimethylation
HLH	Haemophagocytic lymphohistiocytosis
hnRNP	Heterogenous nuclear ribonucleoproteins
HPC	High performance computing
HPO	Human phenotype ontology
HPS2	Hermansky-Pudlak Syndrome type 2
HSCT	Haematopoietic stem cell transplant
ICU	Intensive care unit
IEI	Inborn error of immunity
IFN	Interferon
IL	Interleukin
ISG	Interferon-stimulated gene
kb	Kilobase
LLM	Large language model
lncRNA	Long non-coding RNA
LoF	Loss-of-function
MAF	Minor allele frequency
MAS	Macrophage activation syndrome

MIS-C	Multi-system inflammatory syndrome in children
NAISS	National academic infrastructure for supercomputing in Sweden
NAS	Nonsense-associated altered splicing
NBIS	National Bioinformatics Institute of Sweden
NK	Natural killer
NMD	Nonsense-mediated decay
OMIM	Online Mendelian Inheritance in Man
PBMC	Peripheral blood mononuclear cell
PC	Principal component
PCR	Polymerase chain reaction
pDC	Plasmacytoid dendritic cell
PhastCons	Phylogenetic analysis with space/time models conserved elements
PhyloP	Phylogenetic p-values
PID	Primary immunodeficiency
PRS	Polygenic risk score
pSTAT	Phosphorylated STAT
QC	Quality control
RME	Random monoallelic expression
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
ROC	Receiver operating characteristic
RT-PCR	Real-time polymerase chain reaction
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
S-CAP	Splicing Clinically Applicable Pathogenicity
SCID	Severe combined immunodeficiency

sftp	ssh file transfer protocol
SiPhy	Site-specific phylogenetic analysis
snRNA	Small nuclear RNA
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SQUIRLS	Super Quick Information-content Random-forest Learning of Splice variants
SR	Serine/arginine-rich
ssh	Secure shell
SV	Structural variant
TAD	Topologically associated domain
TBE	Tick-born encephalitis
TCR	T cell receptor
TLR	Toll-like receptor
TSS	Transcription start site
uORF	Upstream open reading frame
UTR	Untranslated region
VCF	Variant call format
VEP	Variant effector prediction
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
XLP	X-linked lymphoproliferative
XR	X-linked recessive

Introduction

In 1964, a letter to the editor of Nature claimed that from the size and assumed triplet code of haemoglobin, and the molecular weight of a chromosome, the rational extrapolation could be made that the haploid human genome contains 6.7 million genes (deemed 'disturbingly high' by the astute author) (1). The scientific belief commonly held at the beginning of the Human Genome Project in 1990, that the human genome must contain around 100,000 genes was derived from a different assumption: that natural selection would favour cells which jettison most non-protein-coding DNA, because it is not useful to the organism (2). The findings from the Project concluded that the true number was barely a quarter of this estimate, and those 20,000 genes occupied only 1% of the total cell DNA (3), changing the landscape of genomics.

The sheer scale and complexity of non-coding genomic regulation which was unrecognised prior to the Human Genome Project is also one of many reasons why diagnostics of Mendelian diseases is a continued challenge 20 years after the complete human genome sequence has been known, and many years after next-generation sequencing became a routine diagnostic tool. Genes without an intrinsic variant can be the cause of disease due to a region many kilobases away; other times, a coding variant which any clinical geneticist would have picked out is spliced out or degraded by the quality control checks of the cell, removing the threat. I have detailed some of the mechanisms we are aware of, and how this may inform variant prioritisation; doubtless many more mechanisms remain.

If this were not enough to be going on with, two months into my PhD heralded the arrival of COVID-19. Whilst a catastrophe for communities and healthcare, for immunologists and geneticists this was also an incentive to act quickly to understand the underlying susceptibilities to infection. The cohesion of my work on non-coding variation, and on the search for IEI cases in COVID-19 and otherwise, lies in diagnostics, and the hope of providing improved prognosis and care for the affected by understanding them a little better.

1 Background

1.1 Diagnostics in the age of high-throughput sequencing

Diseases occurring in fewer than 1 in 2000 people are defined as rare, and whilst this represents few people per disease, the estimated overall rate of the compendium of rare diseases amounts to around 7000 diseases affecting up to 5% of the global population (4). Many of these diseases are single-gene ('monogenic'), caused by one or two variants affecting normal gene sequence or expression. The faster and more accurately a genetic diagnosis can be provided, the better the patient's prognosis is likely to be (5). Whilst the causes of some rare diseases are understood and may be diagnosed easily in new cases, most pose significant problems for healthcare systems, and for patients in getting the support and treatment required (the 'diagnostic odyssey', (6)). As well as possibly painful or life-threatening physiological symptoms, the psychological impact of living with an undiagnosed rare disease on the patient and family is often devastating. Narratives reported from undiagnosed rare disease cohorts described anxiety, uncertainty, and fear (6–8). Thus, improvements to available diagnostic platforms may have far-reaching impact.

The advent of high-throughput sequencing technologies has brought a set of novel challenges to the field of disease diagnostics, even whilst improving general speed and efficiency of data acquisition for individual patients and kindreds. Whole exome sequencing (WES) provides sequence for all the exons in protein-coding genes of an individual (9). It has enabled a molecular diagnosis in 20% of monogenic disease cases (10), up from 5-10% diagnosis rate using earlier tools such as karyotyping, microarrays, or Sanger sequencing (11–13). However, WES relies on the cause of disease being traceable from exonic loci, *e.g.* single nucleotide variants (SNVs) in protein-coding regions, or structural variants (SVs) with breakpoints explicitly in exonic loci. Copy number variant (CNV) identification, deletion or addition of a gene copy possibly causing issues with gene dosage, from WES data is possible but with the risk of high error rates (14,15). Whole genome sequencing (WGS) has expanded the range of disease-causing variants routinely found in patients to include some SVs and non-coding variants in regulatory regions outside of exons, but interpretation is still limited by our understanding of genome architecture and regulation when performing variant identification. Short-read sequencing may also fail to identify SVs or CNVs in repetitive genomic regions. Thus, despite the increased coverage (from 1-2% of an

individual's DNA sequence to 100%) provided by WGS over WES, often the same SNVs are pulled out during analysis by clinical genetics departments, as demonstrated by the struggle to increase diagnostic yield from WES to WGS (16). Despite the integration of WGS into most advanced clinical settings for diagnosis of rare, monogenic diseases, current diagnostic yields have not improved beyond ~35% of predicted cases. With either WES or WGS, the sheer volume of data generated represents a difficulty in successfully narrowing down, from a starting point of hundreds of thousands or millions of variants, which variants are most likely to be the cause of the patient's disease (17). Many population databases have amassed exome and genome data from healthy controls (18–21). It can be particularly useful for diagnostic purposes to gauge the frequency of a variant of interest in the ethnic group matching the carrier, and many countries have local genome databanks (22–24). As an additional difficulty, likelihood of incidental findings is increased using genomic data, and careful consideration must be given to if and how these will be logged and reported (25,26).

Referral for WES or WGS in the clinic has greatly increased in the past 10 years (Figure 1), and standardised protocols for variant interpretation have become necessary to deal with case load and improve reproducibility. For variant interpretation, the American College of Medical Genetics and Genomics (ACMG) has released a guide using 28 criteria to classify variants into five groups: pathogenic, likely pathogenic, likely benign, benign, and variant of uncertain significance (VUS) (25). To group patients more effectively, the human phenotype ontology (HPO) provides a vernacular to describe phenotypic abnormalities in disease uniformly, also allowing optimised literature searches and improving diagnostic capabilities of large language models (LLMs) (27). Tools to convert clinical descriptions to HPO terms and codes are now freely available (28,29). Further serving this purpose, institutes in the US, UK, Canada, Japan, and much of the EU now operate the MatchMaker exchange, where patient HPO terms or candidate gene can be shared and compared (30).

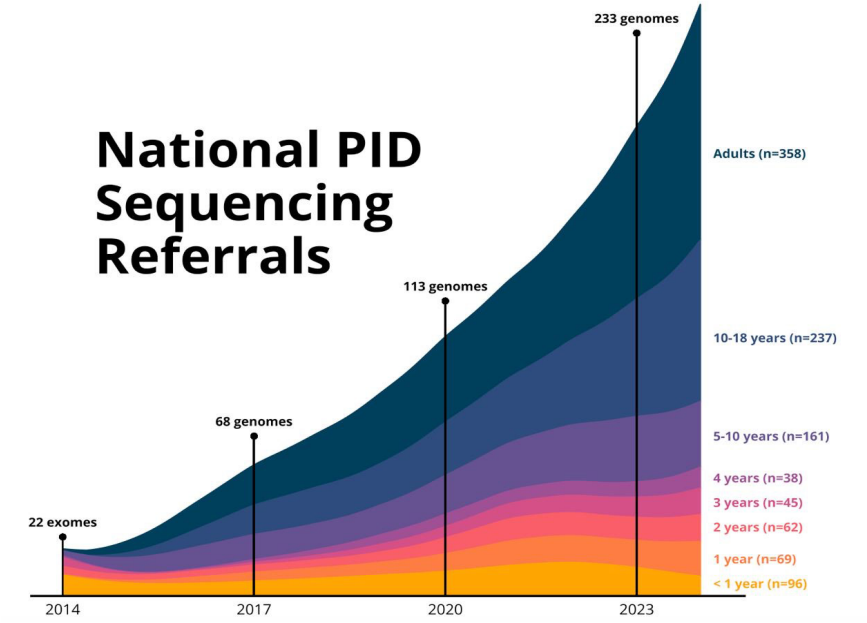

Figure 1. Referrals for to the central facility for Clinical Immunology, at Karolinska Hospital (Huddinge), for high-throughput sequencing between 2014 and September 2024. From 2015, only genomes were sequenced. Trend extrapolation was used to extend the figure until the end of 2024.

1.2 Inborn errors of immunity

Rare monogenic diseases affecting immune system components are classified as inborn errors of immunity (IEIs). As of the 2022 update from the International Union of Immunological Societies (IUIS) expert committee, 485 IEIs have this official status and HPO terms have been expanded to accommodate the wide range of potential manifestations (31,32). IEIs can be life-threatening from an early age. Untreated severe combined immunodeficiencies (SCID) have such a high case fatality rate when left untreated that they are now often screened for in newborns to provide early intervention (33). As newborn screening and other cohort studies have demonstrated, swift diagnosis and treatment can drastically improve clinical outcomes in IEI (34). However, as with other monogenic rare diseases, the rate of diagnosis with WGS is approximately 35%, with work remaining to be done to find novel disease genes and regulatory mechanisms to diagnose the remaining 65% (35,36). Immune cells are easily accessible through

blood draws, allowing functional analyses to be performed. Overall, the disease severity and suitability for detailed phenotyping of IEIs makes them an attractive target for diagnostic efforts.

1.2.1 Impairments of cytotoxic lymphocytes

Cytotoxic T lymphocytes and natural killer (NK) cells are key early responders in the immune system, possessing the ability to recognise infected or neoplastic cells, and release lytic granules which destroy them (37). The machinery required for target cell recognition and lysis, as well as for proliferation and differentiation of the $CD8^+$ T cells and NK cells, is intolerant to loss of function, which causes susceptibility to infections and cancer. Depending on the affected protein, this may cause many different IEIs, which can be challenging to distinguish from each other by clinical phenotype. Investigations into patients with deficiencies in cytotoxic lymphocyte function can provide not only improved care for these patients, but also a better understanding of lymphocyte biology which can be harnessed for treatment of infections and cancers in patients without a rare disease.

Haemophagocytic lymphohistiocytosis (HLH) is a fulminant hyperinflammatory condition clinically diagnosed by the presence of at least 5 out of 8 criteria: episodic fever; splenomegaly; cytopenias affecting at least two of three lineages in the peripheral blood; hypertriglyceridemia (>3mmol/L) or hypofibrinogenemia (<1.5g/L); haemophagocytosis in the bone marrow, spleen, or lymph nodes; low or absent NK-cell activity; hyperferritenaemia; and high levels of soluble IL-2 receptor (38). On a molecular level, defects in NK-cell and T cell cytotoxicity typically result in an exaggerated immunological response to a pathogenic trigger, including overabundance of proinflammatory cytokines (39). HLH may be familial, with a genetic cause of disease, or secondary to another condition such as rheumatic disease, malignancy, or viral infection (40). Non-genetic HLH is referred to generally as secondary HLH (sHLH), or as macrophage activation syndrome (MAS) when in the presence of rheumatic phenotypes. The distinction is further complicated by whether underlying genetic variants predisposing to disease but of lower penetrance should be classed as familial or sHLH.

1.2.2 Genetics of familial HLH

As of 2018, genetic variants are typically identified in 18-40% of familial haemophagocytic lymphohistiocytosis (FHL) patients depending on sequencing approach, typically in an autosomal recessive (AR) inheritance pattern (41).

Several key components of exocytosis are vulnerable to deficiency if a deleterious variant is present on both alleles.

PRF1 encodes perforin, the glycoprotein involved in boring into target cell membranes to facilitate entry of granzymes, of which Granzyme A and B are generally most abundant (42). However, granzyme-based deficiencies have not been reported, as there is functional redundancy between the granzymes which protects against this. Loss-of-function *Prf1* variants in mice have resulted in high levels of lymphoma due to the importance of the perforin-dependent pathway in cancer surveillance (43). Biallelic *PRF1* deficiency is associated with FHL2 (44), whilst biallelic variants in exocytosis genes *UNC13D* (coding for Munc13-4), *STX11* (coding for Syntaxin 11) and *STXBP2* (coding for Munc18-2) are responsible for FHL3, FHL4, and FHL5 (45–48). Munc13-4 binds to the cytotoxic granule (containing perforin and granzymes) for docking and fusion at the immune synapse; Syntaxin 11 and Munc18-2 are part of the SNARE complex with exocytotic function located in the cell membrane (46,49). All three are required for optimal degranulation by T cells and NK cells, and the conditions resultant of deficiency present as phenotypically indistinct from each other. Haploinsufficiency of *UNC13D* has also been associated with cancer susceptibility, although with extremely variable penetrance and presenting at a much later age than biallelic cases of HLH (50).

Although these four genes are regarded as the only canonical FHL genes, additional genes have been uncovered as responsible for more HLH cases. Characterisation of a patient with *RHOG* deficiency represented a significant breakthrough into the molecular mechanism by which Munc13-4 can participate in granule docking with the cell membrane despite lacking a C1 lipid binding domain (51). Variation in further genes encoding proteins involved in vesicle trafficking and membrane docking for exocytosis can also contribute to similarly-presenting hyperinflammatory syndromes, such as *RAB27A*, which is associated with Griscelli syndrome type 2 (GS2); *LYST*, associated with Chediak-Higashi syndrome (CHS); and *AP3B1*, associated with Hermansky-Pudlak syndrome type 2 (HPS2) (52–54). Notably, GS2, CHS, and HPS2 all also typically present with partial depigmentation due to defects in melanosome transportation, which can greatly simplify diagnosis. For a confident molecular diagnosis however, biallelic genetic variants must be coupled with observed functional readouts, such as protein ablation or deficit in cytotoxic lymphocyte degranulation when challenged. Importantly, many cases of HLH with non-coding or structural variants

in the aforementioned genes have been reported, with the *UNC13D* deleterious variants c.118-308C>T in intron 1, and a 253kb inversion at the 3' end of the gene, accounting for half of all FHL3 cases in Sweden (55). The c.118-308C>T variant was also found in 38% of FHL3 patients in South Korea (56). Further variants disrupting regulatory regions of HLH genes have been found affecting *UNC13D* (57), and *RAB27A* (58).

1.2.3 HLH manifestations in other IEIs

Dysregulation of various immune and metabolic molecular pathways have been accompanied by development of HLH in patients without a cytotoxicity defect being directly responsible (59). Of the many HLH-associated disease genes which have been identified, those included in Papers III and IV are addressed here. These can be roughly divided into three categories.

Large multi-protein complexes known as inflammasomes assemble in response to detection of pathogen components (60). They activate proteases including caspase-1, which targets proinflammatory cytokines such as IL-18 and IL-1β for cleavage into their mature form (61). Gain-of-function (GoF) variants in *NLRC4* constitutively activate the inflammasome, causing an autoinflammatory disease sometimes diagnosed as MAS (62). GoF *NLRC4* variants are autosomal dominant (AD), and are often *de novo* for this reason. CDC42 and NCKAP1L are regulators of the actin cytoskeleton, suggested to play a major role in inflammasome assembly. All three genes have been described in instances of disease without viral trigger (63–66).

Intrinsic failure to control infection can also result in hyperinflammation. Some of these conditions are X-linked recessive (XR) and explain some male bias in HLH burden. X-linked lymphoproliferative syndrome (XLP) types 1 and 2 are hallmarked by susceptibility to Epstein-Barr virus (EBV) infection, the eponymous lymphoproliferation, and HLH. XLP-1 is caused by aberration in *SH2D1A* (67), coding for signalling lymphocytic activation molecule-associated protein (SAP), and XLP-2 by defects in *XIAP* (also associated with inflammasome regulation (68)). Though both XLP-1 and XLP-2 are associated with HLH, they may be distinguished from canonical FHLs by assaying lymphocyte cytotoxicity (69). *MAGT1* encodes a ubiquitous magnesium ion (Mg^{2+}) transporter, which however has functional redundancy with TUSC3 reported in some tissues (70). Thus, the consequence associated with MAGT1-deficiency is combined immunodeficiency (71), since T cell receptor (TCR)-triggered Mg^{2+} influx is prevented in lymphocytes but possible

effects in other cells are mitigated by TUSC3. Rarely, MAGT1-deficient patients may also experience HLH. Several paediatric patients with variants in genes in the type I IFN signalling pathway (*IFNAR1*, *IFNAR2*, *STAT1*, *STAT2*, *IRF9*) or genes triggered by type I IFNs (*ZNFX1*) have been recently described with severe hyperinflammatory presentation, fulfilling HLH criteria (72–77).

On the side of metabolic diseases, a spectrum with varying clinical severity of autoinflammatory diseases are associated with variants in *MVK*, coding for Mevalonate kinase (78). Many more metabolic genes have been associated with HLH (79).

1.2.4 Molecular assays evaluating lymphocyte cytotoxicity

Impaired cytotoxic lymphocyte activity in suspected HLH patients has historically been confirmed by the chromium (^{51}Cr) release assay, in which ^{51}Cr -loaded target K562 erythroleukaemia cells are exposed to cytotoxic T or NK cells (80). The volume of ^{51}Cr released back into the supernatant after cell lysis is quantified and used to judge efficacy of target cell killing by the lymphocytes. This assay requires large numbers of cells, a challenge in patients who are often too young to draw sufficient volume of blood and who are often lymphopaenic. Furthermore, use of radioactive ^{51}Cr poses a safety concern for lab workers. Alternative functional assays have been adopted by many labs. Intracellular perforin expression can be quantified by flow cytometry, effectively diagnosing FHL2 in cases where expression is lost. To evaluate defects in exocytosis, quantifying the proportion of patient and control lymphocytes expressing extracellular CD107a (a marker for degranulation) after challenge with the K562 cell line is more practical than ^{51}Cr release (81). An evolution of this assay, where CD107a was instead measured after challenge to Fc receptors, improved efficacy when trialled on 14 healthy donors and 19 HLH patients (82). However, the sensitivity of these assays for diagnosing presence of a primary degranulation defect from patient lymphocytes has not yet been shown in a large cohort.

1.2.5 Personalised medicine in HLH treatment

The distinction between familial and secondary HLH is crucial for determining treatment. The HLH-2004 therapeutic guidelines offer a blanket treatment for all patients meeting HLH criteria, regardless of diagnosis, employing immunosuppressants and chemotherapy (notably etoposide) (38). In patients with FHL, this currently offers a 3-year survival rate of 77% (83). However, personalised treatment may be even more effective.

Autoinflammatory diseases including those caused by NLRC4 and CDC42 can be treated with specific anti-inflammatory drugs such as anakinra (64), which targets IL-1, although not all autoinflammatory phenotypes have been responsive. Inhibition of IL-18 has been trialled in a Phase 3 study on patients with variants in *NLRC4* or *XIAP* (NCT03113760; (84)), and may represent a more effective first-line treatment of these patients (85).

Ultimately, the only current curative treatment for FHL is a haematopoietic stem cell transplant (HSCT). For secondary forms of disease without an underlying genetic defect, or for patients with autoinflammatory disease, this would be ineffective (86). This establishes the need for molecular diagnostics prior to HSCT, an inherently risky procedure. In 2003, hospital mortality stood at around 6% following HSCT (87).

New precision medicine therapies able to target specific pathogenic variants *in vivo* are becoming available, and are an extremely attractive alternative to HSCT. Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas technology uses a specific guide RNA to target sequence of choice and induce a double strand break (88). The repair of this break can be manipulated to replace a pathogenic nucleotide with a harmless one. Genome editing has been trialled in human diseases (89,90), and more recently animal studies have begun to evaluate the safety and efficacy of CRISPR-based editing in HLH (91). These technologies promise an exciting new avenue for patient health, but rely on knowledge of the precise pathogenic variant or variants to design a guide RNA, making diagnostic efforts even more important.

1.3 COVID-19 pandemic

In December 2019, cases of flu-like disease were first reported resulting from infection with a novel coronavirus, SARS-CoV-2 (92,93). Not everybody is equally affected by infection: disease severity ranges from asymptomatic (94–96) to complications including severe cytokine storm, respiratory failure with acute respiratory distress syndrome (ARDS) (97), and multi-organ failure (98–102). The possible severe consequence of contracting COVID-19 is illustrated by the number of deaths, which has exceeded 6 million as of 2023 (103).

The case fatality rate has been estimated to be 2.3%–5% (104,105); however, considering the significant number of asymptomatic infections, the infection fatality rate was estimated as 0.15–0.82% in the pre-vaccination era (106–108).

Clinical risk factors identified for severe illness include age above 50 years, male sex, and comorbidities such as obesity, hypertension, diabetes, and severe asthma (109–117).

Given the diverse presentations, identifying genetic factors which may give some clue to the likely progression of COVID-19 became a critical goal.

1.3.1 Presence of monogenic disease in critical COVID-19

During the emergence of the COVID-19 pandemic, there was an assumption that patients already diagnosed with an IEI might be at particular risk from the virus. In particular, the COVID Human Genetic Effort (CHGE) aimed to understand the genetic determinants of critical disease (118,119). Pre-existing IEI was not generally found to be a co-morbidity conveying especial mortality risk upon infection. An exception lies with presence of autoantibodies to type 1 IFN. A small number of these are associated with an IEI, for example patients with APS-1 (120–122). Autoantibodies have also been identified in 5.2-10.0% of severe cases where no IEI is present, with an increase in prevalence commensurate to increased age (123–128). Type I IFNs are produced large-scale by plasmacytoid dendritic cells (pDCs) amongst other cell types (129), in response to viral infection. The entire pathway comprises viral protein detection, production of signalling molecules IFN-α and IFN-β, and downstream signalling to initiate transcription of anti-viral interferon stimulated genes (ISGs (130)). Several studies performed on young, previously healthy individuals identified monogenic IEIs, including biallelic deleterious variants in AR Type I IFN genes such as *TLR3*, *IRF7*, *IRF9*, *IFNAR1* and *IFNAR2* (131–135). X-linked disorders were also discovered, with *TLR7* deficiency described in familial clusters and purported to be causative in up to 1% of critical cases in young men, possibly explaining some of the sex bias in critical COVID-19 (136–138).

Comparisons may be drawn between the similar presentations of cytokine storm in critical COVID-19 and the hyperinflammation of HLH and MAS, including efficacy of immunosuppressive therapies (139) and the possible contribution of type I IFN signalling genes. Contribution of heterozygous loss-of-function variants in canonical primary HLH genes to acute disease is not out of the question, given the previous association of heterozygous variants with hyperinflammatory response to viral infection late in life (140). Inversely, it is also possible that even if heterozygous loss-of-function variants in genes required for lymphocyte cytotoxicity are not enriched in the most severe cases of COVID-19 resulting in

hospitalization and treatment for ARDS, they may be less compatible with asymptomatic COVID-19, since individuals carrying them will typically experience some inflammatory response after virulent viral infection. In this instance, underrepresentation in a cohort of asymptomatic individuals would be expected. This has been investigated to an extent (141,142), and has also been considered in connection with the paediatric disorder Multisystem Inflammatory Syndrome in Children (MIS-C) (143), but small sample sizes and variability in ethnicity and disease severity suggest that further studies are needed.

1.3.2 Contribution of common risk loci to severe viral infection

Aside from instances of monogenic conditions in COVID-19, many cases of critical disease in formerly healthy people remain to be explained. Several genome-wide association studies (GWAS) have been performed (144,145), across patient groups with different levels of disease severity. Numerous risk loci were identified in these studies, with the strongest signal from an intronic polymorphism in *LZTFL1* on chromosome 3p21.31 (146), and further associations with the ABO blood group locus and several type I IFN loci or IFN-induced genes. A recent GWAS found that risk loci for critical COVID-19 and Influenza A were not shared (147). Whilst some patients with critical COVID-19 had a history of other severe viral infections, the extent to which this is a genetic phenomenon remains unclear. Dissecting the genetic aetiology of susceptibility to specific variants will also need to discover the mechanisms by which these variants confer risk. Uncovering further contributing factors to the missing heritability will help elucidate the overall risk for these individuals. However, interpreting results of these studies in the clinic also faces the challenge of applying risk loci results from homogenously white cohorts to ICU patients of different ethnic backgrounds.

Offering another avenue for investigation with regards to COVID-19, is a common variant in *PRF1*, p.Ala91Val. This variant was originally assumed to be a neutral substitution due to its common minor allele frequency (MAF) and the conservation of the amino acid change. After it was detected in FHL2 patients (148), it was investigated more thoroughly (149), and it is now treated as a hypomorphic variant with around 50% of normal function, with the possibility to predispose to cancers or infections, or to cause late-onset HLH if it occurs in trans with a more disruptive variant.

1.4 Functional genomics

1.4.1 Epigenetic regulation and non-coding variation in disease: approaches for diagnosing the 'missing' 60%

WGS has current diagnostic yield of up to 40% when applied to monogenic rare disease, with an average of 35% (16,150,151). Thus, a majority of cases considered likely to be monogenic go undiagnosed. In a large pilot study by Genomics England, it was found that only 4% of rare disease diagnoses were by identification of disease-causing non-coding variants (16), i.e. 1.4% of all patients in the study were diagnosed with a disease-causing non-coding variant (Figure 2). When considering all the possible mechanisms by which non-coding variants may cause disease, the proportion of coding to non-coding variation, and the lack of investigation into these regions in a clinical setting, it may be deduced that the full range of disease being caused by such variation has not yet been uncovered. It is therefore critical that regulatory elements in non-coding regions of genomic DNA be thoroughly understood and considered to improve molecular diagnostics of rare disease.

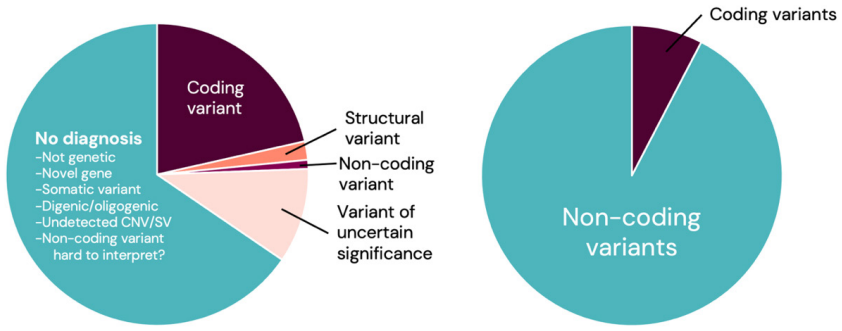

Figure 2. Distribution of coding and non-coding variants of significance for human disease. (A) Results of monogenic disease diagnostics with whole genome sequencing. 35% of patients are diagnosed through WGS; the remaining 65% are likely to be heterogenous in reason that a molecular diagnosis was not achieved. **(B)** SNPs identified after fine-mapping GWAS associations to coding and non-coding regions (data from the GWAS catalog).

The mechanism by which most pathogenic coding variants cause disease can be summarised thusly. A variant miscodes for an erroneous amino acid substitution,

deletion or inclusion, or truncation; the protein is not expressed, or gains or loses function. At the protein level the precise impact may be complex, but at the most fundamental level the genetics is simple. When it comes to mechanisms of functional non-coding variation however, the diversity of regulatory roles creates a plethora of distinct possibilities for disruption of genome regulation, transcription, or architecture, whilst the possibility of cell type-specific roles for each of these elements may make them difficult to identify.

1.4.2 Open-access resources

The Encyclopaedia of DNA Elements (ENCODE) project has used computational techniques to inferentially map non-coding elements on a genome-wide scale (152). Important regulatory non-coding regions are likely to be subject to constraint, and an abundance of tools are available to score or predict nucleotide conservation. Initially, PhastCons software predicted conserved elements over five vertebrate and four invertebrate species (153). The Genomic Evolutionary Rate Profiling (GERP) score estimates expected and observed rate of evolutionarily neutral substitution upon multiple sequence alignments from mammals (154), also included as one of four components of a later tool PhyloP (155). Development of SiPhy included amino acid codon redundancy and nucleotide-specific constraint data in the model from multiple sequence alignments across 21 mammals (156); all of these tools ultimate give similar though not identical readouts and may be more or less appropriate for different purposes.

Experimental validation is necessary for reliable identification of functional non-coding regions, and this has been conducted to an extent; however, the variety of data collection techniques and cell types required has rendered this work incomplete. A further asset in development of lineage-specific element libraries which can be employed for annotation is the FANTOM5 selection of enhancer and promoters which are active in many cell populations, using primary cells from healthy donors albeit in small donor numbers (157). Expression data from primary tissue is further available from the Genotype-Tissue Expression (GTEx) consortium (158,159). With such databases freely available, many different regulatory elements have been characterised in human health and disease.

1.4.3 Enhancers and promoters

Two of the most constrained regulatory regions are promoter and enhancer regions (55,160–162). Proximal promoters contain motifs which specific transcription factors and RNA polymerase may bind to upstream of the

transcription start site (TSS), whilst distal enhancers may be much more difficult to identify due to the long distance between the regulatory sequence and the gene upon which it acts. An extreme example is MFCS1, an enhancer of *SHH* located 978kb away on the same chromosome (and in an intron of *LMBR1*, an unrelated gene), causing preaxial polydactyly typically associated with *SHH* when the region has been ectopically rearranged (163). Enhancers may fall either side or within the target gene and are instrumental in creating the correct 3D chromatin architecture for transcription to initiate once a critical mass of pioneer factors, transcription factors and co-activators is reached; there may also be many of them per gene, sometimes conferring resilience to gene expression if one enhancer is modified (164–166). Where effects of coding variation on the protein can be resolved with established experimental tools such as western blot, substantial effort is often required to identify and validate the functional impact of disruptions to enhancer or promoter regions. Initial difficulty may be with successfully annotating the region as an enhancer or promoter in the relevant cell type for disease. Physical interaction data obtained through Hi-C and ChIP-seq, which use cross-linking or immunoprecipitation to pull down sequences of interest for sequencing (167,168), can link a distal enhancer region to the actual gene with affected expression and the transcription factors involved. Some efforts have incorporated these data into machine learning models predicting further enhancer regions (169–171). Alternatively, expression data comparisons between individuals with enhancer sequence variation can provide candidate genes if the cohort is of sufficient size for statistically significant conclusions to be drawn. In addition to sequence specificity, epigenetic markers distinguishing promoter and enhancer regions have been identified: a histone H3K4me3 signature is found close to promoter regions and is associated with gene expression, whilst H3K4me1 is enriched close to active or primed enhancers (172,173). Thus, epigenetic screens can elucidate the function of non-coding regions with suspected disease involvement, as performed to identify the interaction of a non-coding region featuring a common cancer risk variant (rs6983267) with the *MYC* gene locus, despite the gene being 335kb away (174). Alternatively, other variants in coding regions may obscure further functional regulatory variants (175). During screening of the *RET* locus in patients affected with Hirschsprung disease, two SNVs were identified as candidates, in exon 2 and 5 nucleotides upstream of the TSS, before the correct, intronic, variant was finally identified in linkage disequilibrium with the other considered variants (176).

Expression quantitative trait loci (eQTLs) is a broad term referring to regions affecting variance in gene expression, without specific criteria dictating how the aforestated change is brought about biologically; in fact, many eQTL associations remain poorly understood. Variation in enhancer and promoter regions as described above has been highly associated with eQTLs in GWAS fine-mapping. Specifically, as much as 92% of significant single nucleotide polymorphisms (SNPs) across 5901 GWAS with data submitted to the NHGRI-EBI GWAS Catalog as of March 2024 (177), have been fine-mapped to non-coding variants (Figure 2). Gate *et al.* examined the effect of common single nucleotide polymorphisms on both chromatin accessibility and expression, with attention paid not only to differential expression, but also to the allelic imbalance resulting from a variant in a *cis*-regulatory region (178). Chromatin accessibility was estimated by the assay for transposase-accessible chromatin sequencing (ATAC-seq (179)). This technique relies on the enzyme Tn5 transposase, which can be used to cleave open regions of chromatin, and then insert sequencing adapters into the fragments to tag them for sequencing. The number of reads mapping to a particular region are then treated as a 'peak' if they pass a threshold number, determining how open and accessible the chromatin is genome-wide. Although the focus of the work performed by Gate *et al.* is on predicting the impact of common variants from autoimmune disease GWAS, the same principle can be easily applied to monogenic disease when rare variants have a greater effect size on a single allele, detectable by variant calling of ATAC-seq.

1.4.4 Alterations to intron boundaries

Introns are found in almost every human gene and are spliced out of pre-mRNA prior to translation. Classical splice site sequences are present at the 3' (splice acceptor site; $(y)_n$ag / G in major introns) and 5' (splice donor site; AG / guragu in major introns) intron boundaries, allowing the spliceosome to bind and excise the non-coding intronic sequence. Precision is required to prevent a frameshift in the mRNA, which would be translated into a functionless peptide. Nonetheless, there are many examples in disease of splicing dysregulation caused by introduction or deletion of a critical splice site. Splice site deletion may cause exon skipping or retention of the intron (180). In an instance of in-frame splicing error, mutations in the donor splice site in intron 9 of *WT1* removes the ability to generate a specific isoform including three additional amino acids, and are characteristic of patients with Frasier syndrome (181).

Creation of an unnatural splice site may be within an exon or regulatory region, causing further disruption; or simply introduce a premature stop codon into the frame-shifted mRNA, usually leading to transcript degradation by nonsense-mediated decay (NMD) and loss of expression (LoE). In a case of two siblings with hypergonadotropism and androgen insensitivity, real-time polymerase chain reaction (RT-PCR) in lymphoblastoid cell lines revealed that an exonic variant, which had been originally annotated as a nonsense variant, actually created a new splice donor site causing an in-frame 39 amino acid deletion in *MAP3K1*, inherited through the maternal line (182). Although dominant GoF in male offspring, no associated phenotype had been observed in at least two generations of female carriers. Creation of novel splice donor or acceptor sites has also been observed in *ATM*, associated with ataxia-telangiectasia (183); in intron 4 of *PDS*, leading to a 6bp insertion in the gene product and resulting hereditary deafness (184); and in many further instances (185,186). Gathering examples of deleterious variation has provided a training dataset for tools like SpliceAI, a deep residual neural network with a current best area under the receiver operating curve (AUROC) of 0.91, demonstrating good separation of deleterious and benign variation relating to splicing using only the genome sequence and no RNA data (187). The Super Quick Information-content Random-forest Learning of Splice variants (SQUIRLS) algorithm was trained on a manually curated dataset from literature of over 8000 variants, and outperforms SpliceAI in speed but is similar or slightly lower in AUROC (188). Meanwhile further tools are designed to interrogate a specific aspect of splice variants, such as CryptSplice, a training algorithm to call activating cryptic splice sites (189); S-CAP, which includes gene constraint and other scores to predict pathogenicity rather than splice effect (190); and many others, any of which may be used to call potential further disease-causing variants in a fresh patient genome (191–193).

1.4.5 Disruption or introduction of intronic branchpoints

Pre-mRNA splicing requires a spliceosome snRNA to recognise specific and conserved ribonucleoside motifs within major and minor introns. Major (U2-dependent) and minor (U12-dependent) spliceosomes bind first to the branchpoint (BP) consensus sequence in the middle of the intron, always containing an adenosine and following the sequence YTNAY in humans (194). The adenosine mounts a nucleophilic attack on the 5′ splice site, creating a lariat structure once the 3′ splice site is severed from the pre-mRNA as well. The lariat intron is removed for debranching into a linear component and subsequent

degradation (195,196). Major and minor introns can be differentiated by the nucleosides at the intron-exon boundary, which are GT at the 5′ site and AG at the 3′ site in around 99% of major introns (197), and GT-AG or AT-AC in cumulatively 95% of minor introns (198,199), although this does not necessarily contribute to recruitment of a particular spliceosome (200). The BP is typically 20-40 nucleotides away from the 3′ splice site (201–205), although protective mechanisms have evolved in enhancer and splice sites regions such that BP sites in introns often have 2 or 3 distally located backups. In particularly large introns, there may even be stratified intron removal, using many of the available BPs to anchor the spliceosome, known as recursive splicing (206). It is thus important not to eliminate variants affecting additional recursive BPs during diagnostic WGS analysis.

BPs have been previously characterised in many gene introns by three distinct approaches. Unlike the majority of intronic sequence, BPs typically possess high evolutionary conservation, which when combined with the necessary adherence to the YTNAY motif can be easily scored using bioinformatic approaches found in Branch Point Prediction (BPP (207)), sometimes combined with machine learning algorithms as in Branchpointer (208), or deep learning as in LaBranchoR or RNABPS (209,210). Tools are also available to call branchpoint predictions based on RNA-seq data (211,212). A benchmark study suggested that Branchpointer had the best AUROC when set to detect BPs amongst Ensembl sequences (213). The second approach is simply collation of disrupted BPs contributing to monogenic disease from literature and variant databases. The final contribution to annotation of BPs is from select patients who have disease caused by loss-of-function (LoF) in the human debranching enzyme, DBR1; lack of ability to debranch the removed introns from their lariat configuration necessarily results in lariat accumulation. RNA extraction and sequencing from patients with hypomorphic variants provided the opportunity to evaluate the branching of these lariats in higher frequencies, which in primary fibroblasts was three times the number found in healthy controls (214). Some of this data, along with previous publicly available data, was used to train the BPHunter tool (215).

Notably, identification of a variant which is suspected to disrupt or introduce a new BP still requires functional validation to provide a firm diagnosis, as observed in several cases where such a variant was not conclusively disease-causing (216,217). In fresh analysis of 38,688 members of the 100K genome rare diseases pilot study, where 258 de novo SNVs close to BPs or splice sites were evaluated,

only six participants received a new diagnosis in consequence, as in the majority of cases material for validation was not available or functional analyses did not support the bioinformatic prediction (218).

1.4.6 Regulation of additional alternative splicing machinery

Cis-regulatory elements modulating exon skipping or inclusion are known as exon-splicing enhancers (ESEs) or exonic splicing silencers (ESSs) respectively. ESEs are binding sites for serine/arginine-rich (SR) proteins able to aid recruitment of the spliceosome (219–223). At the other end of the regulatory spectrum, ESSs are binding sites for proteins negatively associated with exon inclusion, for example in the heterogenous nuclear ribonucleoprotein (hnRNP) family. A possible mechanism for mediation of exon skipping in pre-mRNA by hnRNPs is for the proteins to bind flanking ESSs and then dimerise, with the intervening exon looped out of the pre-mRNA transcript (224,225). ESE sequences are distinguishable to a degree by predictive technologies (226), while ESS sequences are less defined. Incidence of mutations in both these elements has been reported in disease.

Loss of *SMN1*, but not *SMN2*, is associated with spinal muscular atrophy (227). Functional characterisation of the loci involved in disease revealed that *SMN2* has sequence homology with *SMN1*, but transcripts lack exon 7 (228). This led to the discovery of a disease-causing variant in *SMN1* which acts by disrupting an ESE (229,230). Fewer ESSs have been identified in instances of disease but nevertheless do occur, as evinced in the description by Aznarez et al. of the creation of an ESS in *CFTR*, aberrantly splicing out exon 11 (231).

Occasionally, nonsense-associated altered splicing (NAS) may be induced as a corrective mechanism, skipping exons when a premature stop codon affects necessary splicing motifs (232). In a family with non-ocular Stickler syndrome, who carried a variant which caused exon 57 to be skipped entirely, an in-frame deletion of 18 amino acids was observed adhering to this putative mechanism. This was significant as a disease mechanism as nonsense variants in the relevant gene, *COL11A2*, are not associated with disease (233). Variable penetrance of NAS activation in disease are exemplified in cases of a *PTCH* nonsense variant identified in a patient with Gorlin syndrome (234), and a *CEP290* variant causing exon skipping, resulting in mild familial retinal dystrophy (235).

1.4.7 Inclusion of a poison exon

Alternative splicing is a naturally occurring phenomenon between different cell types. When aberrant alternative splicing results in inclusion or exclusion of ectopic sequence, this can nonetheless be a mechanism of disease. Erroneous inclusion of a poison exon can manifest in a range of physiological pathologies. Bearing resemblance to the situations triggering NAS, poison exons are inherent in some genes as exons which have a premature stop codon and are spliced out of pre-mRNA transcripts to avoid NMD (236). Poison exons are typically highly conserved and are likely to serve a biological function to justify their maintenance in the genome since prior to speciation from invertebrates (237), suggested by Pervouchine *et al* to be autoregulation of spliceosome components (238), but this is somewhat speculative in nature. Inclusion of poison exons can be caused by trans-regulatory element malfunction, as observed in mouse studies (239), or by variance in *cis* essential splice sites. A well-characterised kindred hosts variants altering the splicing of *FLNA*, a gene associated with periventricular heterotopia (240). Notably, the disease pathology in carriers without standard poison exon suppression in FLNA (and thus a premature stop codon) was less extreme than those patients with a complete LoE. Further variants in *SCN1A* have been studied across multiple kindreds with Dravet syndrome, a congenital early-onset epilepsy (241). Initial cases described with poison exon inclusion and bioinformatic evidence from GENCODE detailing lack of exon translation in canonical transcripts prompted identification of an additional patient during a re-screening of remaining undiagnosed patients with developmental and epileptic encephalopathies (242). This successful cold case diagnosis illustrates the importance of having a procedure in place for re-annotating previous patients when new disease mechanisms and variant discovery tools become available, and may be recapitulated with poison exon splice variant detection in more genes and integrated into clinical workflows (243).

1.4.8 Deletion of TAD boundaries

Development of new chromosome conformation capture (3C) and its successor technologies led to the discovery that many of the *cis*-regulatory elements modulating gene expression are confined to a genomic compartment with strict boundaries to maintain regulatory exclusivity for the target gene or genes (Figure 3,(244–246). These boundaries are enriched for repeat elements such as Alu (244), and for CTCF binding motifs, a shared property with insulator regions proximal to enhancers and likely to mediate higher order chromatin architecture

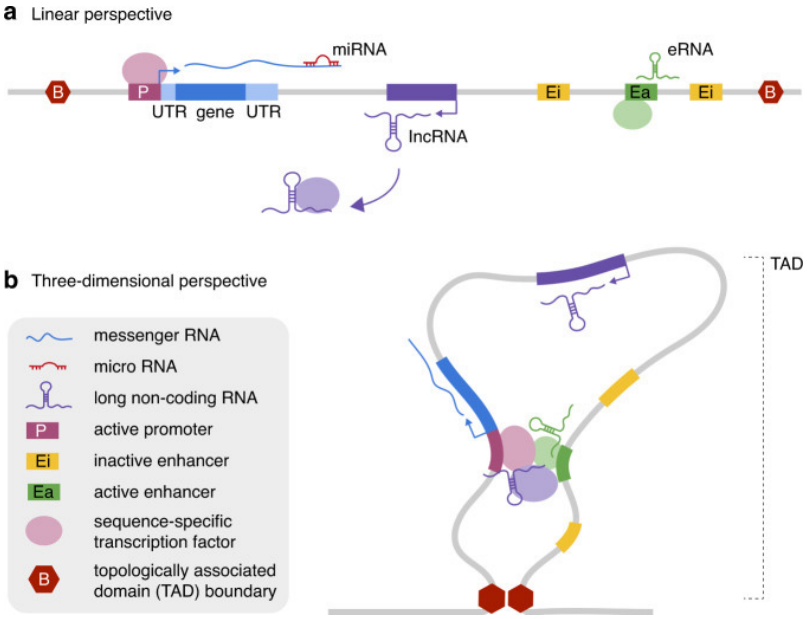

Figure 3. Locus map showing a protein-coding gene and its regulatory elements in (A) linear and (B) higher-order conformation. Reproduced from D'haene and Vergult. (257)

by looping intervening regions for interaction (247,248). Attempts to compare between 3D structural maps lack consensus on whether these topologically associated domains (TADs) are generally static, or differ across cell types (249,250); however, the more CTCF binding, the greater the conservation of the boundary and the stronger the insulation of the TAD (251). Recently, similarity analysis software have become available to compute variability in TADs between cell populations and conditions (252,253); data collation has also been published in a searchable form in the 3D Genome Browser (254), and an imputation of TAD architecture in cell types for which experimental data has not been generated is also available (255). Many TAD callers have been developed, giving options for integration into pipelines in different languages and with different requirements (256).

From the definition of TADs containing much of the expression rate-determining machinery and the inchoation of TAD mapping, it was a short time until the identification of human disease caused by TAD boundary disruption. Study of

three patients each with a heterozygous structural variant in a locus containing four genes implicated gene misexpression due to the removal of TAD boundaries. The misexpression contributed to a severe skeletal malformation, corroborated by 4C analysis on mouse models showing interaction of the *PAX3* promoter with the *EPHA4* locus (258). Thus, although consideration of SVs is typically restricted to coding regions in a clinical setting, the availability of TAD architecture data and the discovery of patients in whom this disease mechanism has been described, the introduction of routine annotation of rare, non-coding SVs with TAD boundary overlap may well prove to be beneficial.

1.4.9 Untranslated region variation

Untranslated regions (UTRs) prior to the start codon (the 5'UTR) and after the stop codon (the 3'UTR) are regulatory regions with several described biological functions. The 5'UTR is necessary for recruitment of the small ribosomal subunit to the transcribed mRNA, whereupon the rest of the ribosomal complex will be recruited and translation will be initiated at an AUG start codon (259). AUG selection and translation efficiency are also dependent on the adjacent sequence (260), and the various possible configurations of the 5'UTR structure; a stem-loop formation by a self-complementary structure for example may impede the ribosome (261). Stem-loops stabilise by protein interaction, and alteration of the protein binding motifs at this locus in the *FTL* gene increased L-ferritin production resulting in hyperferritinaemia in affected patients (262). Upstream open reading frames (uORF) in the 5'UTR are present in almost half of genes, and reduce protein expression by 30-80% (263). A further possible element providing transcriptional regulation are known as start-stop elements, where an uORF is immediately followed by a stop codon (264). The ribosome struggles to dissociate from the mRNA under these circumstances and is held in position, preventing translation of the remaining 5'UTR sequence before re-initiation at the canonical start codon, though more work is required to understand what regulatory activity this may have. Creation of an uORF, or disruption of a stop codon belonging to an uORF, have both been implicated mechanisms in instances of monogenic disease (265–269). These variants can be annotated in WGS data for diagnostics using the UTRannotator software (270).

1.4.10 Non-coding repeat expansions

Repeat expansions in non-coding regions may cause mRNA gain-of-function, whereby RNA transcripts aggregate prior to splicing and the resultant foci cannot

be removed from the nucleus, possibly interfering with splicing machinery perturbing splicing of other mRNA transcripts (271). This is particularly prevalent in progressive neurological disorders. Notably, there is a larger average intron size in neuronally expressed genes (206), increasing the probability of a replication error. Myriad possible examples include intronic penta- or hexanucleotide repeats in *C9orf72* (causing amyotrophic lateral sclerosis [ALS]), *TK2* or *BEAN* (causing spinocerebellar ataxia type 31 [SCA31]); 3'UTR repeats in *DMPK* (causing myotonic dystrophy type 1 [DM1]); trinucleotide expansions in the *HTT* 5'UTR (causing Huntington's Disease [HD]); and trinucleotide repeats in the *FMR* 5'UTR (causing Fragile X-associated tremor ataxia syndrome [FXTAS]) (272–279). Screening for repeat expansions in monogenic disease is currently standard when WGS is performed, necessitating tools which can detect repetitive patient genomic sequence not conforming to the reference alignment (280).

1.4.11 Variation in non-coding transcripts

Non-coding disease genes historically have been extremely difficult to identify and characterize. The lack of protein product is an impediment to identifying when alterations to transcript volume and functionality have occurred at all, let alone when there may be wider impact to the cell. There are many subtypes of non-coding RNAs, not all of which have been associated with disease as of yet (281). An easy mechanism to characterise is *cis*-acting lncRNAs, which regulate a proximal protein coding gene. SV displacement of these lncRNAs will thus prevent their function. Notably, this basic mechanism can also impact lncRNAs situated further away from the gene they act upon: one of the first reported instances described lncRNA disease gene *Maenli*, which is 300kb upstream of the site it regulates: protein-coding gene *EN1*. *Maenli* homozygous deletions are a phenocopy of the limb malformation caused by *EN1* loss in mice and human patients (282).

Until recently, no haploinsufficiency of a lncRNA had been identified as causative of monogenic disease. *De novo* heterozygous deletions of *CHASERR*, a lncRNA interacting with the *CHD2* gene locus, have been found to elicit a phenotype distinct from patients with CHD2 LoF or deficiency (283,284). Reduction in *CHASERR* transcript abundance correlated with increased *CHD2* expression from the allele in *cis* with the deletion. Thus, loss of the inhibitory lncRNA caused a deleterious overexpression of CHD2 in patients. Knowledge of this potential

disease mechanism, whereby dosage-sensitive protein-coding genes may be dysregulated by disturbance to proximal lncRNA genes, may lead to further diagnoses if such variants can be properly identified and prioritised from patient WGS.

1.5 Monoallelic expression

The assumption is that having two copies of each gene, they are likely to be transcribed equally. Other than the obvious exception of the X and Y chromosomes (where either only one copy is present, or a second has undergone random deactivation), there are naturally occurring exceptions where epigenetic control supersedes biallelic expression in mammalian genomes.

Recently, methylation analysis has been employed as a diagnostic tool, searching for disease-associated episignatures which can resolve VUS or SVs with unclear impact. However, a pilot study of 582 cases reported a yield of only 2% diagnosis with this technique (285).

1.5.1 Imprinting

Genomic imprinting refers to the specific expression of a gene or cluster of genes, by parent of origin (286). Although two intact copies of the gene and regulatory elements are present, differentially methylated regions (DMRs) are established in the germline cells to determine which copy will be expressed, and maintained through development (287). Clustering of imprinted genes is common, and specific sequences have responsibility for the initiation of differential methylation over the entire gene subset; these are known as imprinting control regions. Strict maintenance of parent-of-origin expression discriminates imprinting from other modes of monoallelic expression.

Early discovery of endogenous imprinted regions utilised mouse embryos, which may have been a confounding factor for the first imprinting genes being associated with development (288–291). Furthermore, the translation of this work from mouse to interpretation of human cellular mechanisms is imperfect. Subsequent works have been able to rely on detection of allele-specific transcripts, which allows the screening of many tissues and possibly detect more elusive or tissue-specific regions of genomic imprinting. Long-read WGS with base modification data is now also available, providing detailed maps of methylated DNA (292).

1.5.2 Imprinting disorders

The nature of constitutively monoallelic expression in imprinted regions means that even if the silenced copy of the gene is completely healthy, any damage to the active copy can result in total loss of protein activity. Well-defined DMRs have been identified by human imprinted disease, notably chr11p15.5 (Beckwith-Wiedemann/Russell-Silver syndromes, (293,294)), and chr15q11.2 (Prader-Willi/Angelman syndromes, (295)). Imprinting diseases can be paired by the resulting condition if the maternally or paternally expressed genes from the region are damaged.

If the gene is intact, but two copies of the region have been inherited from the same parent, the same disease will result since the methylation pattern will not allow gene expression (296). This phenomenon, where one parental chromosome is duplicated during meiosis, is known as uniparental isodisomy. In WGS it may appear that there are large runs of homozygosity in the patient, similar to what one might observe in a consanguineous patient (297).

Disorders caused not by direct impairment to a DMR, but to the machinery responsible for maintenance of methylation, have also been studied. Methyltransferase DNMT1 ensures that methylation patterns survive DNA replication. Homozygous deleterious variants in *DNMT1* are embryonic lethal, but heterozygous AD cases have been reported (298). Affected individuals have global hypomethylation including loss of parent-of-origin specific regulation (299), demonstrating the utility of genome-wide methylation analysis in diagnosing conditions affecting global methylation maintenance as well as gene-specific expression.

1.5.3 Random monoallelic expression

Monoallelic expression seemingly not strictly determined by parent-of-origin is instead termed random monoallelic expression (RME). Although this phenomenon has been observed in up to 10% of genes in both humans and mice (300,301), the reason remains unclear. It may thus be possible that variants on a single allele of a gene undergoing RME may be sufficient to cause disease in genes which would otherwise be assumed to be AR in disease inheritance. Alternatively, RME may be a mechanism of disease prevention, if an allele carrying a damaging variant is silenced.

2 Research aims

The overall aim of this thesis is to increase diagnostic yield in monogenic rare disease, focusing on cases where inborn errors of immunity or susceptibility to viral infection are suspected. Specific aims were as follows:

1. During the COVID-19 pandemic, many young and previously healthy individuals were stricken with severe or critical disease. We aimed to identify how much of this disease burden could be attributed to monogenic IEI, or to rare variants carried heterozygously which could predispose to critical disease after viral trigger whilst not qualifying as a full IEI. We describe a family in whom two cases of monogenic IRF7 deficiency were diagnosed in **Paper I**. **Paper II** and **Paper III** then explore the extent to which defects in Type I IFN and hyperinflammation-associated genes were represented in larger cohorts of critical COVID-19 patients.
2. Paediatric hyperinflammatory syndromes may be genetic or secondary to a range of conditions including infection or cancer. Haematopoietic stem cell transplant is the only effective cure for familial HLH, but it is drastic and an accurate molecular diagnosis must be in place before transplantation is carried out. **Paper IV** aims to improve the proportion of patients diagnosed, and at greater speed and accuracy compared to previous workflows.
3. In **Paper V**, we developed and implemented a pipeline for molecular diagnosis of individual patients for whom whole genome sequencing alone has been ineffective, by interrogating the impact of non-coding variants on chromatin accessibility for prioritization of variants. Two proof-of-concept patients with FHL3 caused partly by an intronic variant were used at the beginning of this study. Identification of the known variants through genome-wide analyses was successful in the two proof-of-concept patients. Three patients with undiagnosed conditions were subsequently trialed with this method, and one candidate variant is being investigated further using validation with RNA-seq and downstream, targeted bioinformatic analyses.

3 Materials and methods

3.1 Cohorts

3.1.1 CovPID20 cohort

Patients with critical COVID-19 were recruited to a study investigating incidence of IEI among such cases. Of 90 patients treated in the ICU at Karolinska sjukhuset, Huddinge, 24 declined to participate and 28 were excluded due to risk factors, mortality, or an inciting reason beyond critical COVID-19 for their ICU stay. The remaining 38 patients were reported in **Paper II**.

3.1.2 Type I IFN deficiency patients

Familial clusters of young, healthy adults suffering from severe or critical COVID-19 were referred for WGS on suspicion of IEI. Analyses of WGS revealed truncating variants in *IRF7* (homozygous state) and *TLR7* (hemizygous state) in the two kindreds evaluated. The *IRF7*-deficient patients were reported in **Paper I**, and both the *IRF7*-deficient and *TLR7*-deficient patients were used as controls in type I IFN production and signalling assays in **Paper II**.

3.1.3 COVID Human Genetic Effort cohort

High-throughput sequencing was performed on 3269 patients with critical COVID-19, and 1373 individuals who had been infected with SARS-CoV-2 but remained asymptomatic throughout, as previously reported (302). The patients were collected by the labs listed in Appendix I ('List of CHGE authors') for **Paper III**, and mostly sequenced disparately. The asymptomatic sequencing data was mostly collected by the Casanova lab in Paris, and has a bias towards white European ethnicity. To maintain consistency across centres, tool versions approved for preprocessing were: BWA 0.7.12, GATK 3.4-46, PICARD 1.92, and alignment to hg19 or GRCh37 was required for all samples. GATK HaplotypeCaller was used for variant calling. All sequencing was then transferred as variant call format (VCF) or binary alignment map (bam) files to the administrative centre of the CHGE at Rockefeller University, with appropriate data transfer agreements in place.

3.1.4 Cytotoxic lymphocyte deficiency patients

Electronic health records and/or blood samples were procured for 1761 patients with suspected IEI or other relevant immunological phenotype over the course of 12 years in the Bryceson lab. Where possible, cellular phenotyping was performed

consisting of immune cell counting, cytotoxic lymphocyte granule and exocytosis profiling (82). Often parents, siblings, or other family members were also collected. In addition to internal controls, transport controls were also assayed where appropriate.

3.1.5 'Proof-of-concept' FHL3 patients

Two patients with FHL3 were subjected to phenotypic profiling, cellular analyses, and Omics profiling of sorted T cells and NK cells as described in Section 3.2. The first patient was an 8 year old male of Chinese ethnicity, who died before transplant could be attempted. He carried the intron 1 variant c.118-307G>A on his maternally inherited allele and c.1388A>C (p.Gln463Pro) on his paternally inherited allele. The second patient was a 3 year old female of Serbian origin, who was received HSCT after diagnosis and is currently alive and in a stable condition. She carries the intron 1 variant c.118-308C>T on her maternally inherited allele and c.2346_2349del (p.Arg782SerfsX12) on her paternally inherited allele.

3.1.6 Patients with unknown IEIs

Three patients who had been referred to Karolinska Hospital (Huddinge) or to the Bryceson group directly with suspected IEI, but who were WGS-negative, were selected for prospective trial of a novel diagnostics pipeline. Patients were selected on the basis of available WGS for the whole trio, early and severe disease onset, and availability of cells for omics and follow-up functional analyses. Further descriptions are available in **Paper V**.

3.2 Molecular assays

Whole genome sequencing
Genomic DNA was extracted from whole blood using the QIAamp DNA Mini kit (QIAgen cat no. 51304). Sequencing libraries were prepared using a PCR-free paired-end protocol and sequenced on the Illumina NovaSeq 6000 by Clinical Genomics, Stockholm. Reads were preprocessed and mapped to GRCh37 as previously described (150).

ATAC-sequencing
Frozen peripheral blood mononuclear cells (PBMCs) from patients and their parents were resuscitated and sorted into populations previously observed to be affected by disease for each distinct patient, as described in 3.1.5. Following re-suspension, 5000 cells were delivered into Lo-bind Eppendorf tubes and libraries were prepared according to the Omni-ATAC protocol (303). Sequencing was

performed on the NextSeq 2000 platform by the Bioinformatics and Expression analysis core facility.

RNA-sequencing
Frozen PBMCs from patients and their parents were resuscitated and sorted as for ATAC-seq. 1000 cells from each subset were delivered into PCR tubes and mini-bulk libraries were prepared according to the T-RHEX-RNAseq protocol (304). Sequencing was performed on the NextSeq 2000 platform by the Bioinformatics and Expression analysis core facility.

Sanger sequencing
Variants identified in *TLR7* or *IRF7* through WGS were validated by Sanger sequencing. Briefly, genomic DNA was extracted from whole blood using the QIAamp DNA Mini kit (QIAgen cat no. 51304). Primers were designed using Primer3 to interrogate 400-600bp surrounding the expected variant (305). Polymerase chain reactions (PCRs) were performed and products were separated by gel electrophoresis and extracted using the QIAquick gel extraction kit (QIAgen cat no. 28704). DNA fragments were sent to Eurofins Genomics for Sanger sequencing.

3.3 Computational analyses

3.3.1 Public datasets

GnomAD
The Genome Aggregation Database (GnomAD) combines exomes and genomes from diverse sources to provide a variant frequency database alongside several gene- and variant-specific metrics (19,20). In GnomAD v4.0, data from 730,947 exomes and 76,215 genomes are available. Notably, gene-level details such as probability of loss-of-function intolerance (pLI) and distribution of ClinVar variants are tailored to aid diagnostics of Mendelian disease (306). However, no phenotypic information is available beyond ethnicity.

1000 Genomes Project
Whole genomes are available from the 1000 Genomes Project (1kGP) for 2504 individuals, of deliberately diverse ethnicity (18). Ethnicity and biological sex is available for all individuals. Access to some phenotypic data may be applied for on a research basis.

SweGen
Variation in WGS from 1000 Swedes is available either through the SweFreq

access portal, where individual variants may be viewed, or in VCFs for WGS on application to the National Bioinformatics Institute of Sweden (NBIS). Individuals were selected via a nation-wide screening of array data to represent a cross-section of Swedish ethnicity as described (23).

GENOMICC SNPs

Summary statistics were available for a GWAS performed on WGS data from 7491 critically ill COVID-19 patients, against 48400 controls by the Genetics of Mortality in Critical Care (GenOMICC) investigators (https://genomicc.org/data/r2/) as previously described (307). Patients descended from a variety of ethnic backgrounds, and had been sourced from 224 ICUs in the UK. This data was used to perform polygenic risk score (PRS) analysis in **Paper II**.

NHGRI GWAS Catalog

The NHGRI-EBI catalog of human genome-wide association studies contains datasets from published and unpublished GWAS based on a minimum of 100,000 germline SNPs pre-QC. Eligible datasets are identified by curators from the EBI (published data) or submission (unpublished data). Thus, all known trait associations for any given SNPs may be accessed at a single point. Figure 2 was generated from the catalog downloaded in February 2024.

3.3.2 Private datasets

Clinical Immunology patients

Demographic data for patients referred to Clinical Immunology, Karolinska Hospital (Huddinge) between 2014 and 2024 for WES or WGS was kindly provided by Sofie Vonlanthen.

Control subsets

For control epigenetic datasets, matched specific cell compartments were sorted and ATAC-seq and RNA-seq libraries were prepared. For each subset, 4-8 healthy donors were sequenced. NK and T cell subsets were reported previously (308). Sequencing data for B-cells from healthy controls was kindly provided by the Månsson lab.

3.3.3 Computing resources

Scout

Preliminary analyses of patient and trio WGS were performed using the VCF graphical browser Scout, developed and maintained by Clinical Genomics

Stockholm. Where appropriate, candidate variants were validated by Sanger sequencing.

UPPMAX
Further computations were enabled by resources in projects sens2017581 and sens2017582, provided by the high performance computing (HPC) made available by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX. UPPMAX is funded by the Swedish Research Council through grant agreement no. 2022-06725.

3.3.4 Ancestry principal component analysis

Assessment of ancestry by principal components (PCs) aims to group individuals quickly by shared variation. Subsequently, we decided to employ the calculated components to identify which of our patients came from ethnicities which are lowly represented in population databases. Common variants across all critical COVID-19 patient and control genomes were merged using bcftools. Plink 2.0 was used to filter out variants which had low Hardy-Weinberg equilibrium exact test p-value below 0.00001, or had missing data; then to calculate eigenvectors across the cohorts. The first two principal components were plotted using ggplot2 in R.

3.3.5 Polygenic risk score calculation

To assess whether patients with critical COVID-19 may have a predisposition caused by impact of many variants with small effect size, polygenic risk scores were evaluated. Quality controls were performed on downloaded summary statistics from the GENOMICC GWAS to ensure no instances of mismatched, duplicate, or ambiguous SNPs were included in the calculation. Merged common variants from the critical COVID-19 patients included in Paper II and the 1kGP were pruned for regions of linkage disequilibrium in tiles of 200 variants with plink 2. The final PRS calculation sums β-coefficients (effect size and direction) from the summary statistics were for the SNPs from each WGS in our cohort or the 1kGP. This was performed using PRSice-2 using the first five of the previously generated ancestry PCs as covariates. PRS distributions for the two groups were compared.

3.3.6 Odds ratio calculations

The increase or decrease of event likelihood in a target group may be quantified as an odds ratio (OR), in simple form calculated as $n_1 \times k_2 / n_2 \times k_1$ where n represents group size and k represents number of events within the group. An OR

of 1 indicates no difference in event probability; events are less likely in groups with ORs closer to 0, or more likely with an OR greater than 1. ORs were used to assess incidence of rare variants in genes associated with IEI in **Papers II** and **III**. The actual computation in both papers used Firth's bias-reduced logistic regression with the logistf R package, for which the coefficient exponential was equivalent to an OR. This method was chosen to mitigate bias due to small sample size (which runs the risk of an extreme maximum likelihood estimate), and to allow easy adjustment for ancestry covariates.

3.3.7 Variant co-occurrence analysis

Variants in the same individual in the same gene may be on the same, or opposite allele. GnomAD offers a tool (for version 2 only) allowing look-up of inferred phasing for variant pairs, in coding and UTR regions. Variants frequently occurring together are expected to be on the same haplotype.

3.3.8 ROC curve generation

Mapping the true positive rate ('sensitivity') against the false positive rate ('specificity') in a binary classifier model generates a plot known as the receiver operating characteristic (ROC) curve. A diagonal from the (0,0) to (1,1) coordinates indicates the success rate of a random classifier, while the further towards the (0,1) coordinate in the upper left corner (representing no false negatives and no false positives) the plotted line approaches, the better the performance of the classifier. In **Paper IV**, we compared the sensitivity and specificity of three different flow cytometry assays for identifying patients with a defect in exocytosis. ROC curves were plotted for each assay comparing patients with exocytosis defects to: patients with hyperinflammation but no genetic cause; patients with an IEI unrelated to exocytotic machinery; healthy adult controls from blood bank donations; and healthy samples sent with patient samples to control for transport. The pROC package in R (v3.0.2) was used for all ROC curve generation.

3.3.9 NGS data preprocessing

Demultiplexed ATAC-seq and RNA-seq fastq files were trimmed and duplicates removed using PICARD (v2.12.0). Filtration and mapping to GRCh37 used Samtools (v1.9). Variant calling from WGS, ATAC-seq, or RNA-seq read data to gVCFs was performed with GATK v.4.1.4.1. Importantly, since some variants may not be called from ATAC-seq or RNA-seq alone due to the measurable material from specific

cells, VCFs were called jointly from patient or trio WGS data and ATAC-seq or RNA-seq. VCF files were hard filtered using recommended GATK filtration parameters for SNPs and indels indicated in Table 1 (309). Genotyped VCFs were phased using WhatsHap software to reconstruct haplotypes from physical and pedigree data (310).

Table 1. Recommended post-variant calling GATK filters

Filter	Target
QUAL < 30.0	Low-confidence site annotation for SNVs and small insertions or deletions (whilst GQ would refer to the confidence in the specific call).
QD < 2.0	SNVs and small insertions or deletions with quality of call artificially inflated by read depth. Calculated by variant quality score / allele depth
SOR > 3.0	Strand bias of forward and reverse strands between reference and alternative calls of SNVs. Calculated by symmetric odds ratio test.
FS > 60.0 FS > 200.0	Strand bias of forward and reverse strands between reference and alternative calls of SNVs; and of small insertions and deletions. Calculated by Fisher's exact test.
MQ < 40.0	False-positive variants due to an inaccurate read mapping (particularly in regions with repeat regions or sequence complexity) Calculated using the root mean square of the mapping quality generated by the mapping software.
MQRankSum < -12.5	Heterozygous SNVs with bias in mapping quality for reference vs alternative reads.

	Calculated by Mann-Whitney-Wilcoxon rank sum test
ReadPosRankSum < -8.0 ReadPosRankSum < -20.0	Heterozygous SNVs with bias in relative positioning of reference vs alternative alleles in reads Heterozygous small insertions or deletions with bias in relative positioning of reference vs alternative alleles in reads. Calculated by Mann-Whitney-Wilcoxon rank sum test.

3.3.10 MAGNET diagnostic workflow

Variant annotation

WGS variants were annotated with: MAF from GnomAD genomes; Genomic Evolutionary Rate Profiling (GERP) scores; scores from the UTRannotator; eQTL loci from the Database of Immune Cell Expression (DICE (311)). The GERP score obtained was scaled with the rescale function in R.

Differential accessibility analysis

Peaks were called from patient and control data using Homer(312). Raw read inputs were normalized and differential accessibility (DA) analysis was performed using the Homer DESeq2 wrapper.

Differential expression analysis

Gene count matrices were produced from mapped reads for patient and control data using Homer, and differential expression (DE) analysis was performed using the Homer DESeq2 wrapper.

Allelic imbalance in chromatin accessibility

For heterozygous variants in WGS, significant allelic imbalance in ATAC-seq was modelled by the distribution X~Bin(n, 0.5). The number of reads from the alternative allele, *k*, was used to find the two-tailed p-value. Thus, each read is effectively treated as a Bernoulli trial (a test with two possible, equally likely outcomes).

Variant prioritisation

The full algorithm designed for analysis of variants covered by ATAC-seq is as follows: $2\sum_{x=k}^{n}\binom{n}{x}p_0^x(1-p_0)^{n-x} + MAF + \left(1-\sqrt{\frac{\sum GERP^2}{n-1}}\right) + DAp$

where n = total number of reads at the locus and k = number of reads called from the alternative allele. Variants were ordered by largest to smallest scores, and assigned rankings accordingly.

3.4 Ethical considerations

Papers I and II. For the second study, forty-four patients from the Karolinska Hospital (Huddinge) ICU donated samples and gave informed consent to allow their biological material and information obtained by questionnaires, clinical read-outs during disease, and histories to be used in the study. We did not include any patients under the age of 18, or who had died in the ICU. Furthermore, it was made clear to the patients when they gave consent that this may be withdrawn at any time and their data would be erased as far as possible. It was determined that unless a cause of disease could be confidently determined, genetic variants uncovered during the study would not be returned to the patients; however two patients who were found to have autoantibodies to Type 1 IFN were followed up and discussed their condition with their physician.

In this project, genetic data from the patients was uploaded to the HPC available through UPPMAX. Under GDPR, genomic data qualifies as sensitive personal data. To ensure the requisite levels of data protection were met, we used the Bianca cluster, which was purpose-built for sensitive jobs and has no direct internet access. Log-in is only possible from university networks or VPNs, and through a secure shell protocol (ssh). Instead, data transfers are made via an ssh file transfer protocol (sftp) server to reduce possibility of a data leak.

Additionally, for Studies I and II, blood was obtained from four severely or critically ill patients with COVID-19 aged over 18 who had familial clusters. All four gave informed consent to have their blood drawn, and for functional and genetic assays. After a candidate variant was identified through whole genome sequencing and validated by Sanger sequencing, we reported the disease-causing variants back to the clinicians. I compiled reports on the WGS for each of the four patients, aimed at helping the clinicians in question to deliver the information each patient

might require about their variant, in addition to our preliminary diagnosis. One concern was that third parties in the patients' families would inevitably be affected by the delivery of this information. The patients and their families were keen to continue in the study, and several related adults to the two kindreds also provided samples and consent for genetic testing, with the understanding that they were engaging with a research study and not a formal healthcare service. As such, after an adult female individual (cousin to the probands) was found to be a symptomless carrier of the X-linked *TLR7* variant, it was decided by all concerned that the family should be offered genetic counselling before deciding whether to proceed with testing her two sons, who were both minors and would likely be symptomatic if carriers.

The projects were carried out in accordance with ethical application Dnr 2020-01911. The main ethical considerations we actually grappled with were: the taking and storage of patient samples, which were anonymised in our biobanking system in compliance with GDPR; the genomic sequencing, likewise anonymised in our lab and with the caveat that this was being performed for research purposes and secondary findings would not be reported back to the patients (although IEI-gene variants were reported to the clinicians); and the complication once children were found to be potential carriers. Since investigation is ongoing into how actionable *TLR7* variants are, this part of the study had to be handled extremely carefully with the involvement of clinicians and counselling services close to the family in question.

The 48 genomes sequenced were further shared with collaborators in the COVID Human Genetic Effort, as they performed gene burden testing on all the probands. The real names and personal details of the patients were not shared with collaborators to protect the privacy of these individuals. Data sharing was carried out through Globus, a secure file manager. The genomes are securely stored on Clinical Genomics servers for five years and will eventually be archived. A second copy is kept by our lab, on securely stored hard drives not accessible by any lab members not directly involved in this research project, and not connected to the internet, where the data security could be compromised. However, the full implications of data sharing across borders are complex, as researchers from many countries are involved in the CHGE and laws in other countries regarding data processing, storage, and safety may be more lax than the stringent rules in the EU. The responsibility of each party must also be questioned in instances where there are several controllers. We ensured that only the central labs, in

Rockefeller and Paris, would be granted access to the platform on which the raw data is kept in this collaboration, and that the platform would be an HPC equipped for sensitive data storage.

Paper III. This project was performed under the umbrella of the CHGE, and took place during my visiting studentship at Rockefeller University. All data was anonymised, and genetic data has been stored securely on either password protected hard drives, or on a sensitive data HPC. No samples or other materials were used during this study, and as such no further ethics approval was required than the CHGE constituent members' ethics approvals.

Paper IV. Patients were referred to the Bryceson lab over a 12 year period with suspected defects in lymphocyte cytotoxicity. After informed patient or parental consent had been obtained by the primary care physician, blood samples were shipped to Stockholm and subjected to flow cytometry assays. Any extra material was frozen down and stored. Our sample storage was ethically approved and the conditions adhered to the restrictions in Biobankslagen, the Swedish Biobank Act. Precise logging of our biobank samples into a laboratory information management system (LIMS) preserved the privacy of patient personal data by using IDs and keeping the original patient information secure to only personnel directly involved with patient work.

Where reduction in numbers of lymphocytes, or degranulation was considered consistent with a differential diagnosis of primary immunodeficiency, we attempted to establish the molecular cause of disease. Until 2013, this was mostly by Sanger sequencing, or by protein quantification via Western blot. Since 2013, we have instead used exome or genome sequencing for diagnostics. Data storage used the same safety precautions as described the cohorts sequenced for Papers I and II. Genome analyses were performed on a research basis. During my tenure as analyst, we would return a report including any rare variants in IUIS genes to the clinician, which contained the results of bioinformatic predictive tools, variant inheritance, and my interpretation of the likelihood that each variant could be disease-causing.

Paper V. This paper is concerned with precision medicine; using highly detailed clinical phenotyping, genomic and epigenomic data to obtain a molecular diagnosis for individual patients. As such, nearly all of the material gathered in relation with this project is sensitive. Patients are typically selected for the project after they have already been enrolled in the wider IEI characterisation study

carried out in the lab, and collection of samples and informed consent for immune phenotyping and genetic assays is covered under Dnr 2013/1723-31/4.

A particular ethical challenge represented in generating diagnoses and using patient genetic data in a research setting is the difficulty in balancing the different interests of the clinic, where patient data sensitivity and privacy is prioritised, with the needs of contemporary research, which relies heavily on data exchange and open access to maintain reliability and reproducibility. Without demonstration of these attributes, the value of the research for the community is diminished, even whilst a diagnosis may have improved care of the patient. Adherence to data protection laws helps balance these interests, but it is important to note that whilst we attempt to anonymise data as far as possible, as well as process and store it securely, the use of WGS means that were we to release this data full anonymisation could not be achieved. Even declaring variants which cause rare disease may be identifying, as in the instances of the IRF7-deficient patients (6 known living cases worldwide) and the TLR7-deficient patients (15 known living cases worldwide). However, since our patients in both these cases had significant media involvement of their own instigation, any identification resulting from these publications could not be credited entirely to the impact of the articles. For the patient cases we have studied in this project to date, several factors had to be considered. Firstly, diagnostics have been carried out on a research basis, and although validation is scrupulously performed and only actionable variants are returned to the clinic, the burden of quality assurance ultimately lies with the clinic when it comes to using this information in patient treatment; thus, communication between the research project and the clinic is critically important for this study. Secondly, the patient cases have so far been mostly early onset cases, and publishing of sensitive data will be handled with caution and due consideration for protection of the minors involved. Finally, although public interest in their own genome sequence has increased over the past few years, as exemplified by the popularity of companies such as 23andMe, we have declined to provide the opportunity to patients to obtain their own raw genomic data and do not include an SOP for data transference.

4 Results

4.1 Rare inborn errors of immunity cause life-threatening COVID-19

In **Paper I**, we performed WGS on a 38-year-old Swedish man with healthy BMI and no history of severe infections, who suffered respiratory failure after infection with SARS-CoV-2 and required treatment in the intensive care unit (ICU) for 15 days (P5 in **Paper I**). A homozygous frameshift variant in *IRF7* (p.Ala280GlyfsX12) was detected and validated with Sanger sequencing (Figure 1). Prior to the SARS-CoV-2 pandemic in 2020, only one patient had ever been reported with AR IRF7 deficiency (313). The patient originally reported by Ciancanelli et al (P1 in **Paper I**) had presented with severe influenza A infection at 2.5 years, and remained healthy since with a strict vaccination schedule coupled with serologic tests to ensure vaccine efficiency. Nonetheless, the early presentation of P1 gave a contrast with our patient, who had remained healthy until his fourth decade.

The 32-year-old younger brother (P6 in **Paper I**) of the index Swedish patient had also been hospitalised with severe COVID-19. Although he did not require ICU treatment during his COVID-19 episode, he had been hospitalised with infections twice as a teenager, once with Influenza A and once with a streptococcal infection. After his recovery from COVID-19, he was hospitalised with severe tick-borne encephalitis (TBE), despite having completed a three-dose course of vaccination. After discovering the homozygous *IRF7* variant in P5, we also performed WGS on P6, and found that he too carried the variant in a homozygous state. Genotyping by Sanger sequencing showed that the mother of the two patients was a heterozygous carrier. The father was already deceased, but it was presumed that he was also likely to have been a heterozygous carrier. No consanguinity had been reported between the mother and father, leading us to speculate on how the dynamics of the regional population genetics might lead to a predicted damaging variant being found in unrelated individuals. At the time of reporting, p.Ala280GlyfsX12 was the most common pLoF variant in GnomAD (MAF = 0.00009), although the release of GnomAD v4 has provided data on four variants with higher MAFs. Incidence of p.Ala280GlyfsX12 in a Swedish population database was also unexpectedly high (MAF = 0.0015 in SweGen). We postulated that this variant may be a founder mutation in some Swedish/Finnish regions: specifically, a rare variant more common in a group which has been isolated for geographical or cultural reasons, producing genetic drift from the general human population. Since the drift has occurred for reasons independent of natural selection, the

founder mutation may not improve fitness, and thus pathogenic variants may be maintained at higher frequency than would be expected. Furhtermore, all the pLoF variants in GnomAD v4 with MAF greater than 0.00001 were most prevalent in the European or Ashkenazi Jewish populations, and most were not present at all in Asian or African populations. This distribution suggests that constraint on *IRF7* has been more severe in some ethnic groups, possibly influenced by regional pathogen exposure.

Through patient matching made possible by the CHGE consortium, we found that three further patients (P2, P3, and P4) had been identified by other teams following bouts of critical COVID-19. Two of these patients (P2 and P3 in **Paper I**) were first reported by Zhang et al in the seminal CHGE paper of 2020. A final, seventh patient was diagnosed following a severe course of respiratory syncytial virus (RSV) aged 6 months.

Reduced protein expression had already been noted in Western blots performed on IFN-β-stimulated cells from P1, P2, and P3 (313,314). Despite p.Ala280GlyfsX12 having a CADD score of only 10.1 (a truncating variant would normally be expected to have a CADD score in the 25.0-40.0 range), quantification by Western blot on IFN-β-stimulated cells showed no detectable protein in P6 (Figure 3), and it was assumed that the frameshift variant caused NMD of *IRF7* RNA transcripts. Similarly, no protein was detectable in stimulated cells from P7. Primary cells from P4 were not available, as the patient had died of COVID-19. However, HEK293T cells were transfected with the variants from all seven patients in addition to an IFN-β luciferase reporter, and assessed for expression and function of IRF7. Cells induced with all seven patient genotypes had significantly reduced production of IFN-β, as quantified by the luciferase reporter (Figure 2 and Zhang et al.). Notably, though there was around 25% of wildtype IFN-β production retained in HEK293T cells transfected with p.Ala280GlyfsX12, this may be different to the pathology in the two patients, as an IRF7 fragment was detectable by Western blot in transfected HEK293T cells where no protein at all could be found in primary material.

Follow-up immunological interrogation of patient pDCs showed ablation of IFN-α production after stimulation with TLR7 or TLR9 agonists in P5, P6 and P7 (Figure 3), although residual IFN-β production was retained presumably allowing some downstream effects of type I IFN signalling to continue. Another possible compensatory mechanism was observed in $CD4^+$ and $CD8^+$ T cells. T cell subsets

from P5 and P6 stimulated with different pooled viral peptides had significantly increased frequency of IFN-γ response compared to healthy controls when the stimulant was derived from H1N1 influenza or SARS-CoV-2, but not from cytomegalovirus (CMV) or EBV (Figure 4). This demonstrates that there is an increased frequency of memory T cells to the viruses IRF7 deficient patients are most vulnerable for.

Overall, we were able to explain cases of severe and critical COVID-19 in two Swedish patients at the molecular level, and recommended a vaccine plan based on this information. Recently it has been reported that autoantibodies to type I IFN (effectively a phenocopy of monogenic deficiencies in type I interferon) are detectable in up to 10% of patients with severe TBE (315), suggesting that P6's hospitalisation with TBE, could indeed be attributable to his IRF7-deficiency. We reported these patients with two other novel IRF7 deficiency diagnoses, and supplementary details and updates on the phenotypes of three previously reported cases of IRF7 deficiency. At the point of publication, these seven patients together represented every reported case of AR IRF7 deficiency globally. We further investigated the possibility of a compensatory mechanism in T cells, which along with retained IFN-β production, may partially explain why the patients, despite abrogation of IFN-α and IFN-λ production, had remarkable resistance to a wide range of pathogens and had already lived to a mean age of 29 years.

4.2 Contribution of Type I IFN variation to critical COVID-19 susceptibility

To elucidate the immunologic and genetic contributions to critical COVID-19 in the previously young and healthy, genomes were sequenced for 38 patients enrolled in the CovPID20 study. All patients were treated in the ICU of Karolinska Hospital in Huddinge prior to vaccine availability. Previous work on young critically ill patients, particularly when several family members were severely affected, had revealed monogenic deficits in type I IFN-associated genes. We previously diagnosed patients with AR *IRF7* deficiency (described in **Paper I**), and XR *TLR7* deficiency, and these patients were used as controls for immunological assays in **Paper II**. Autoantibodies to type I IFN were detected in two of the 38 patients, who subsequently could also be considered controls in other immunological assays since no additional defects were expected or seen (Figure 2).

In GWAS performed on cohorts of patients critically ill with COVID-19, a risk locus on chr3p21.31 consistently conferred the strongest association with disease

severity (316). Linkage disequilibrium was highest for an intronic polymorphism in *LZTFL1* (rs17713054G>A), which encodes a protein involved in cilia function (317,318). We detected the variant rs17713054G>A in 17 of our 38-strong cohort with critical COVID-19, a 4.0-fold enrichment compared to the 0.055 minor allele frequency (MAF) in GnomAD. Whilst this could be biased by which ethnic populations are included in the respective cohorts, we nonetheless investigated whether other common risk variants were also disproportionately represented in our cohort. Polygenic risk score (PRS) analyses were calculated on summary statistics for the 10,000 SNPs most associated with critical disease in the GENOMICC screen (145), adjusted by ancestry PCs, and evaluated by ROC curve. We used the 2504 individuals in the 1000 Genomes Project (1kGP) population database in lieu of larger datasets, since in the 1kGP whole genomes (rather than lists of variants detected per gene) were available. The AUROC was 0.53, indicating limited separation between our cohort and the 1kGP (Figure 3). Thus, PRS analysis of common risk variants was not able to predict disease severity reliably. We noted that whether the donors included in the 1KGP had been infected with SARS-CoV-2, and their relative disease severity, was unknown and thus we may be comparing with a control cohort containing critically ill patients – possibly around 75 cases of critical disease if the 1kGP adheres to the statistic that 3% of those infected with SARS-CoV-2 will develop critical COVID-19, although a lower rate could be expected given the cohort nominally represents a healthier sample than the total population. Unfortunately there was no way to determine this information, and compiling genome data for sufficient numbers of exclusively people asymptomatic after infection was unrealistic.

We also investigated type I IFN genes, as several instances of monogenic disease had already been identified in patients with critical COVID-19, including with a putative AD disease mechanism (314). All variants in coding or splicing regions of type I IFN genes previously associated with human monogenic IEI were considered, namely *DDX58*, *IKBKB*, *IFIH1*, *IFNAR1*, *IFNAR2*, *IRF3*, *IRF5*, *IRF7*, *IRF9*, *JAK1*, *MYD88*, *STAT1*, *STAT2*, *TBK1*, *TICAM1*, *TRAF3*, *TYK2*, and *UNC93B1*. These variants were filtered by combined annotation dependent depletion (CADD) score greater than the 99% mutation significance cut-off for each gene (the CADD threshold which 99% of pathogenic variants are predicted to be above), and MAF <0.001, the cut-off for a variant to be considered very rare. We also examined the loci for structural variants, and did not find any overlapping either coding or regulatory regions (Supp Figure 3). Six patients carried either a heterozygous very rare

variant, or a homozygous rare variant (<0.01), in the Type I IFN signalling pathway. Overall, a small enrichment was observed of very rare variants in type I IFN signalling gene variants in our cohort compared to the 1kGP, after adjustment with ethnicity PCs (Figure 4). No enrichment was present for variants in any individual genes in type I IFN production.

To validate whether the rare and very rare variants observed in *IFNAR1*, *IFNAR2*, *TYK2*, *JAK1*, *STAT1*, *STAT2*, and *IRF9* were damaging to stability or function of the proteins, we performed an immunological interrogation of the type I IFN signalling pathway. First, patient and healthy donor $CD4^+$ T cells were stimulated with IFN-α. Phosphorylated STAT1 (pSTAT1) and STAT2 (pSTAT2) was quantified, and a reduction in pSTATs was observed in three of the very rare or rare variant carriers compared to healthy controls (P29, P30, and P35), although this was not significant (Figure 5). Since effects of decreased function may be more quantifiable downstream, we also assessed induction of three interferon-stimulated genes (ISGs) by flow cytometry after $CD4^+$ T cells were stimulated with IFN-α. The same three patients in whom lower STAT phosphorylation had been noted, also had low induction of MX1, IRF7, and IFIT1 compared to healthy controls and to critical COVID-19 patients who did not carry any type I IFN signalling pathway variant (Figure 6). The remaining three patients of the six carriers displayed normal responses in the assays to measure pSTAT and ISG levels, demonstrating the importance of functional validation to accompany genetic data.

Further genome analyses aimed to identify any biallelic rare variants in the gene list circulated and curated by the International Union of Immunological Societies (IUIS; (32)). We investigated a possible defect of *CSF2RA* in a patient with a very severe lung phenotype during her COVID-19. *CSF2RA* codes for a subunit of the receptor for granulocyte-macrophage colony-stimulating factor (GM-CSF), and is associated with hereditary pulmonary surfactant metabolism dysfunction manifesting as respiratory distress during infection (OMIM #300770). Although normally an X-linked IEI found in males, two *CSF2RA* variants were identified in the patient which had not been observed on the same haplotype in GnomAD, allowing the possibility of compound heterozygous recessive-X linked disease. However, stimulation with GM-CSF did not show impaired pSTAT5 induction in the patient monocytes (Figure 7).

In summary, in a cohort from an ICU of 38 young, formerly healthy individuals, we found that two possessed autoantibodies to type I IFN, effectively explaining the severity of their disease. Of the 36 remaining patients, 6 were carriers for a very rare heterozygous or rare homozygous variant in the type I IFN signalling pathway, and in three of these cases there was functional evidence of reduced type I IFN signalling compared to healthy controls. Although a further case of possible *CSF2RA* deficiency was investigated, no defect was seen in patient monocytes. Thus, we suggest that defects in type I IFN signalling may have contributed to the disease of five patients, constituting 13.2% of the cohort. Screening for autoantibodies to type I IFN is likely to yield more actionable results than genome analyses, which require larger cohorts and stringent functional follow-up analyses in most unrelated cases. In the remaining patients, other factors such as common variants predisposing to disease, viral load, or environmental factors may have played a part too challenging to dissect in a small cohort.

4.3 Contribution of hyperinflammatory gene variation to critical COVID-19 susceptibility

Severe SARS-CoV-2 infections typically result in acute respiratory distress syndrome (ARDS). COVID-19 mortality is often accompanied excessive proinflammatory cytokine release, labelled a "cytokine storm". The resulting inflammation exacerbates ARDS, and can leave survivors with severe organ damage and trauma (319).

HLH shares some clinical features with COVID-19 cytokine storm (320,321). The majority of familial HLH cases are caused by mutations in genes required for lymphocyte cytotoxicity (*PRF1*, *UNC13D*, *STX11*, *STXBP2*, *RAB27A*, *LYST*, *AP3B1*, *RHOG* (51,322,323)). Individuals with autosomal recessive loss-of-function variants in these genes typically present with fulminant hyperinflammatory disease in infancy or childhood. Variants in genes involved in inflammasome activation (*NCKAP1L*, *CDC42*, *NLRC4)* may also be causative of HLH. In contrast to primary HLH, secondary forms of HLH lacking strong genetic components occur more often in adulthood. Viruses are frequent triggers of primary as well as secondary HLH (324). Notably, heterozygous genetic variants in familial HLH genes may be a contributing factor in secondary HLH (325). It is therefore worth considering that the pathological mechanism of these genes may extend beyond biallelic early-onset disease to haploinsufficiency increasing susceptibility to

severe hyperinflammation and immune dysregulation upon viral infection (Thesis Figure 4). Providing further parallels to HLH, immunosuppressive glucocorticoid therapies have been efficacious in treating severe COVID-19 patients (139).

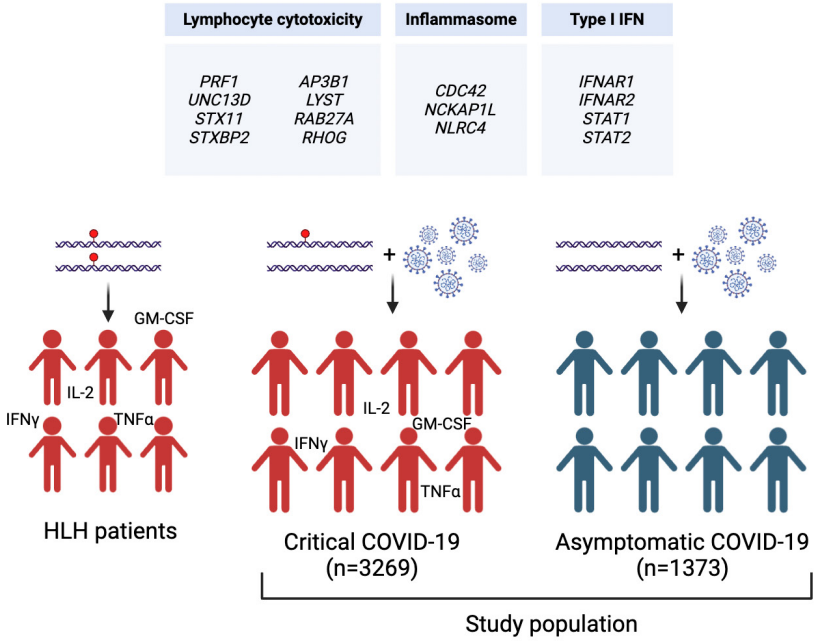

Figure 4. We investigated whether the similarity in molecular presentation and effective therapies against HLH and cytokine storm in critical COVID-19 shared genetic aetiology. Our proposed model of disease was that whilst biallelic damaging variants in genes associated with lymphocyte cytotoxicity, inflammasomes and type I IFN signalling causes early-onset HLH, a variant in a single allele may result in worse control over the immune response triggered by SARS-CoV-2 in an adult with otherwise good health. However, after analysis we found no enrichment in rare variants in these genes in critical patients compared to individuals who were asymptomatic after infection.

We assessed rare variants in *PRF1*, *UNC13D*, *STX11*, *STXBP2*, *RAB27A*, *LYST*, *AP3B1*, *RHOG*, *NCKAP1L*, *CDC42*, *NLRC4*, *IFNAR1*, *IFNAR2*, *STAT1*, and *STAT2* in 4642 genomes and exomes collected by the COVID Human Genetics Effort (CHGE), belonging to people who were infected with SARS-CoV-2 and were either asymptomatic, or became critically ill. No study participant carried a rare variant

in *CDC42*, one of two genes known to have an AD mode of inheritance. In all other genes, no enrichment in either cohort was noted after adjustment for ancestry PCs. This included the other gene with known AD inheritance (*NLRC4*), and *IFNAR1*, in which patients with predicted AD disease were reported by the CHGE. Surprisingly given the rate of inborn errors of type I IFN immunity in critical COVID-19 patients among the CHGE, none of the genes involved in type I IFN signalling had any enrichment in critically ill patients. Where two rare variants appeared in the same gene and same individual, we used the GnomAD co-occurrence tool to establish likelihood that these variants could be on the same haplotype. A slightly greater proportion of the critical COVID-19 patients were predicted to carry biallelic variants (2.6%) than the asymptomatic infected (2.2%) but again this difference was not significant.

We also selected a common yet hypomorphic variant in *PRF1*, p.Ala91Val (c.272C>T), to quantify in the CHGE cohorts. This variant had similar frequency amongst the critically ill (4.0%) and asymptomatic (5.0%), although the distribution of carriership varied. A larger proportion of critically ill patients were homozygous for p.Ala91Val, but this difference was not statistically significant.

Overall, contrary to the findings of Luo et al., we were not able to confirm any enrichment of rare variants in any genes associated with HLH, nor of a common hypomorphic *PRF1* variant conferring around 50% of PRF1 function. Based on our findings and the lack of reports of primary HLH patients suffering critical COVID-19, we predict that SARS-CoV-2 is unlikely to serve as a viral trigger for HLH. Thus, genetic screening of HLH genes is unlikely to garner results which may aid diagnosis or treatment of critical COVID-19 patients in the clinic.

4.4 Stimulation with anti-CD3 and anti-CD16 is an effective diagnostic platform for defective cytotoxic lymphocyte degranulation

The clinical utility of WGS in diagnosis of HLH and other hyperinflammatory syndromes is currently undercut by the expense and timeframe for a conclusion to be reached. Given the fulminant nature of these IEIs, many centres prefer to systematically test affected cells in patients, namely CD8+ T cells and NK cells. The Bryceson lab previously published protocols for effective assays designed to separate patients with defects in lymphocyte cytotoxicity not only from healthy controls, but also from patients with other genetic hyperinflammatory syndromes, secondary HLH, or MAS (82). The improvement of these assays on previously described techniques is due to the activation of T cells or NK cells by anti-CD3 or

anti-CD16 monoclonal antibodies respectively, rather than by K562 target cells. The dependent variable, of surface CD107a detectable by flow cytometry, remains consistent (Thesis Figure 5). T cells and NK cells are more effectively triggered, and thus the proportion of degranulating cells with detectable surface CD107a is increased in samples from healthy donors. In addition, potential biases due to overrepresentation of NK-cell subsets with reduced responsiveness to K562 cells (i.e. adaptive NK cells) are eliminated as anti-CD16 antibodies more uniformly engage and trigger $CD56^{dim}$ NK cells.

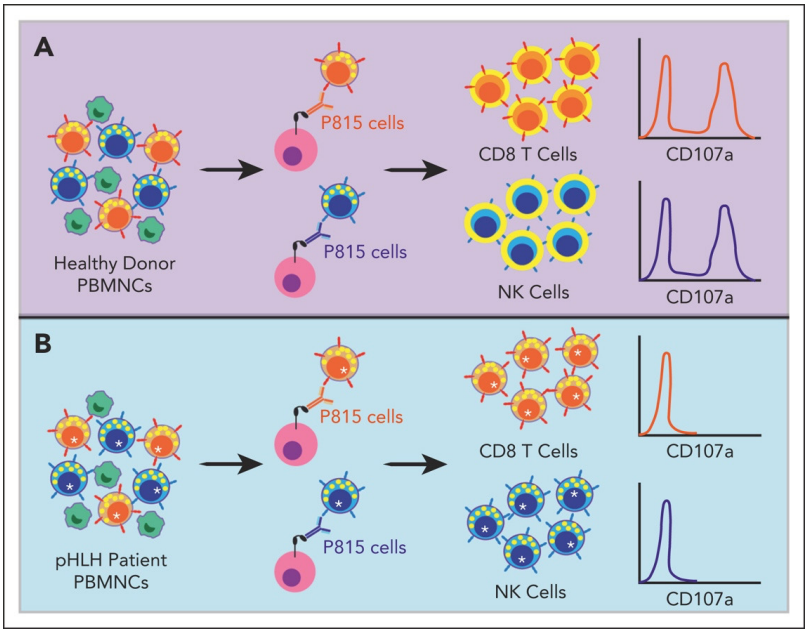

Figure 5. Stimulation of cytotoxic lymphocytes with P815 + anti-CD3 or anti-CD16. The lymphocytes will try to kill the P815 cells by release of cytotoxic granules. In healthy donors **(A)** this is measurable by presence of surface CD107a, indicating successful exocytosis. The absence of CD107a in HLH patients **(B)** indicates a defect in the required exocytotic machinery. Reproduced from Meyer and Nichols, 2024 (326).

We validated these assays by including them in a standard panel performed on patients referred to the Bryceson lab and fulfilling HLH criteria over a ten year period. Stimulation of NK cells with K562s was also performed on all samples

during this time. Ninety-two patients were ultimately diagnosed with biallelic LoF in one of *UNC13D*, *STX11*, *STXBP2*, *LYST*, *RAB27A*, *AP3B1*, or *RHOG*, and these comprised the group in whom an exocytosis defect would be anticipated. Other patients were divided into those with a genetic diagnosis (in *PRF1*, *SH2D1A*, *XIAP*, *CD27*, *GATA2*, *MAGT1*, *ZNFX1*, *CYBA*, or *ITK*; n=58) and those without after a genetic evaluation had been performed (n=63). Transport controls (n=84), and volunteer donors (n=198) were also included for an overall study population of 495 individuals. As well as prioritising patients for genetic work-up, this allowed comprehensive analysis comparing the three protocols. All three assays (NK cell stimulation with K562s, NK cell stimulation with P815 cells coated in anti-CD16, T cell stimulation with P815 cells coated in anti-CD3) were able to separate patients with a predicted exocytosis defect from all other groups (Figure 3A-C). However, the higher rates of exocytosis in healthy donors prompted by stimulation with anti-CD3 or anti-CD16 of T cells or NK cells were able to separate groups with greater sensitivity (Figure 4A-D). Youden's index was used as a summary statistic, and consistently found that stimulating NK cells with anti-CD16 was the best performing diagnostic test. Combination analyses on any two of the three tests found that the greatest accuracy (of 99.3%) could be achieved by combining anti-CD16 NK cell assays with anti-CD3 T cell assays. We also noted that out of all the assays, the T cell assay was the most robust when analyses must be performed on samples which had been under transportation stress.

Overall, we were able to demonstrate an effective and expedient workflow to stratify patients with similar hyperinflammatory phenotypes, but different ultimate diagnoses.

4.5 Integration of Omics data improves diagnostic rate in IEIs

From 2014 to 2024, the number of patients referred to the Karolinska Institute for analysis with high throughput sequencing technologies on suspicion of IEI has increased by an order of magnitude (as shown in the first figure of this thesis). Yet, as in other rare disease fields, the actual proportional yield of diagnoses has stagnated. We hypothesised that some of these patients carried at least one non-coding variant which had not been identified as pathogenic during a clinical evaluation.

Specificity of transcription and the regulatory regions which maintain this divergence in different cell types has been demonstrated conclusively, but the effect of individual variation in this context is more challenging to analyse at scale

whilst maintaining subject privacy. The GTEx database has featured in some studies tackling this (327–330), but even so, sample sizes and number of cell types assayed remains small enough that new techniques may be required to properly assist variant annotation for diagnostics (where data a single or very small number of individuals must be interpreted). Transcriptome and methylome analyses have already been trialled as diagnostic tools paired with WGS, with additional yields of 2-15% (285,331–334). These methods display sensitivity to different variant types, suggesting that patient phenotypes should be carefully considered when hunting for a diagnosis beyond the coding genome. RNA-seq excels at detecting or confirming aberrant splicing, which can occur in any gene with more than a single exon. However, differential transcript volumes and allelic bias are currently more difficult to analyse consistently, depending on the control samples from which RNA-seq is available, and whether the relevant gene contains a heterozygous coding variant which can offer a datapoint for allelic imbalance. Methylome analysis can give very relevant data in cases linked to imprinting control or methylation maintenance, but many diseases simply do not have a characteristic signature which can be interpreted easily.

To increase interpretability of non-coding variants identified in rare disease patient genomes, we devised a workflow which would provide additional relevant datapoints and the means to rank them genome-wide (Figure 2). Our starting point was to select patients who had early onset and severe phenotypes, and for whom WGS was also available for parents, both criteria which have indicated better chances of a diagnosis (335,336), as well as increasing the likelihood of a true IEI. We also selected patients for whom repeat sample collection was possible. Using ATAC-seq and RNA-seq, we decided on four parameters to be analysed for each genomic variant: differential chromatin accessibility and allelic imbalance, and differential transcript expression and allelic bias. We sorted cells from patients which were affected by their disease, ensuring that all chromatin accessibility and transcript data would be relevant to the correct cell types.

Combined with population frequency and evolutionary conservation, these factors generate a score predicting variant functionality. Notably, the direction of functionality regarding increase or decrease of accessibility and transcription is not currently reported. For the manuscript as included, RNA-seq was only available for one of the proof-of-concept patients, so results the ATAC-seq modules are reported. RNA-seq is ongoing for additional patients.

Our workflow, the Mendelian disease ATAC-seq and Genome NETwork analysis pipeline (MAGNET), successfully prioritized variants in *UNC13D* intron 1 of two FHL3 patients in the top five over almost all cell types (Tables 1 and 2). Having established the initial algorithm which would be employed, we also optimized the workflow by examining regions of natural biological imbalance in chromatin accessibility, so that interpretation of monoallelic accessibility of variants called in these regions could be refined. We curated a list of validated imprinted genes based on a literature review, some of which correlated with regions of imbalance in chromatin accessibility in our data (Figure 4). Although these variants are not filtered out during analysis, they are flagged as being on our curated list of imprinted genes in the pipeline output and a list of only imprinted variants is also available if an imprinting disorder is suspected.

However, some regions of imbalance not previously reported were striking. The largest of these was on chr9, and we did not find an enrichment of genes involved in *e.g.* development, which might indicate an imprinted region according to existing reporting patterns. Reasons for this could be that this is a gene-sparse region, which may not have been picked up by investigations using transcriptomics or structural variation; that monoallelic expression has been under-researched in many cell types including lymphocytes, so any more lineage-specific DMRs may not have been investigated; or errors of sequencing or data processing. As an example, choice of genome build has been suggested to impact ability to successfully interpret omics data for rare disease diagnostics (337). Re-evaluating this region after mapping to the most recent genome build, T2T-CHM13v2.0 (3), would resolve this concern.

We tested MAGNET on three further patients who remained unexplained in the clinic after WGS (both at the time of sequencing and after recent re-evaluation). Of the top-ranked variants from the three patients, we identified a variant in a MYB promoter region. We judged this variant to be of interest due to apparent *de novo* genotype, the AD inheritance of *MYB*, previous identified cases of IEI caused by *MYB* variation, and the general conservation of the promoter locus. No alternative TSSs were found by looking at FANTOM5-CAGE data (Figure 6), either in lymphocytes or other examined cell types. We plan to assess this variant further functionally, first by generating RNA-seq on the patient cells; then by follow-up experiments to identify which TFs can bind to this locus, and whether *MYB* transcription is impacted by the variant in cell lines.

5 Conclusions

In this thesis, I have detailed several molecular and bioinformatic strategies for improving diagnostics of patients with IEIs. These run the gamut of clinical diagnostic work, from establishing the speed and efficacy of a flow cytometry assay in diagnostic work which has the potential to become routine in suspected HLH cases; to presentation of a novel functional genomics workflow designed to identify candidate non-coding variants in challenging undiagnosed patients. A summary for each paper is detailed below.

Paper I reported rare biallelic variants in IRF7 in two Swedish patients, and summarised what is known about presentation of IRF7 deficiency to date. Several of the cases presented late in life, only after patients were exposed to a novel coronavirus.

Paper II attempted to characterise, from the patients in a single ICU most likely to have an underlying genetic aetiology of the severity of their infection, what proportion may have an IEI, and which detection methods should have the highest efficacy. This contrasted to other strategies we and others have used, such as in the IRF7 and TLR7 cases where patients were referred for analysis on the grounds of familial presentation. We found a possible small contribution between very rare heterozygous/rare homozygous variants in Type I IFN signalling pathway and reduced ISG signature, but ultimately the effect size was difficult to quantify. For directing critical COVID-19 treatment, screening patients for autoantibodies to type I IFN is the most likely to yield useful diagnostic information.

Paper III determined that there was no evidence for enrichment of very rare variants in genes associated with primary HLH in a large cohort of patients with critical COVID-19, compared to asymptomatic infected individuals. A common but hypomorphic variant in *PRF1* was also examined, and likewise no enrichment could be identified.

Paper IV details the efficacy of two previously described protocols for rapid diagnostics of patients with an impediment in the exocytosis of cytolytic granules in T cells and NK cells. The assays had routinely high sensitivity and specificity in the detection of samples with genetic deficiency, especially in combination with each other. Particularly crucial, the assays retained sensitivity when comparing hyperinflammatory patients with and without a genetic defect, not only when familial patients were compared to healthy controls. Use of these safe, fast, robust

assays in diagnostics of monogenic hyperinflammatory syndromes can prioritise patients for genetic analyses and more invasive treatments such as HSCT.

Paper V identifies the need for novel diagnostic tools able to interpret the functionality of non-coding variants in monogenic rare disease. It first describes several parameters by which pathogenic non-coding variants are detectable in genome-wide screens, and implements this knowledge into an algorithm which may be applied to WGS, ATAC-seq, and RNA-seq data from a patient with suspected monogenic disease. Further chromatin accessibility analyses were performed to identify regions of normal biological imbalance in lymphocytes, and the extent of regional imbalance surrounding common, rare, or rare and known pathogenic non-coding variants. We also utilised the algorithm on four new patients and were able to propose a novel candidate variant in a promoter region of *MYB*, the gene coding for c-Myb, a previously identified disease gene for IEI.

6 Points of perspective

Two decades since the publication of the human genome sequence, a revolution has taken place in our understanding of human rare disease genes, and in turn of human biology. The potential for discoveries based on so-called 'experiments of nature' remains an exciting premise. Nonetheless, much work remains to be done to optimise diagnostics of these patients, both in variant detection and prioritisation, and in speed at which results can be attained.

Although the global pandemic driving work behind Papers I-III has subsided, patients from these papers remain of interest, especially if further characterisation or phenotypes emerge. Patients with variants in IRF7 and TLR7 or autoantibodies to type I IFN have been provided with their diagnoses, and may be the subject of follow-up studies in the future to better understand these afflictions and particularly the range of viruses they may be susceptible to.

Paper IV showcased a diagnostic platform for HLH based on flow cytometry assays, arguably making the need for further investigation redundant if defective cytotoxicity is indeed detected (given an HSCT would be the current gold standard treatment for all FHL cases). In reality, in cases of HLH where sensitive cellular assays have already been performed, a genetic diagnosis is still extremely valuable in identifying affected relatives and suitable donors for HSCT. Moreover, if therapies involving genome-editing become widely available, knowledge of the exact nucleotides affected becomes essential. However, the partnership of swift assays evaluating function, with the precision of genetic analyses is extremely powerful: the techniques are in symbiosis with each other.

Paper V increases the tools in the arsenal of a molecular diagnostician after a negative WGS, but work remains to be done to validate candidate variants from our test patients. As we accumulate more patients, the parameters can be optimised further, with the possibility to calculate thresholds which can increase the speed at which the pipeline can run. If we gather sufficient patients to form a larger pilot dataset, then the addition of supervised learning elements to the workflow also becomes possible.

We were particularly interested in whether use of ATAC-seq data marked an improvement on current use of transcriptomics data in rare disease diagnostics, which is becoming more widespread clinically since its initial implementation in 2017 (333,334). Reports from trials of this method have demonstrated its utility in

identifying or validating variants affecting splicing. However, there is so far a lack of evidence for the value of RNA-seq alone in identifying the root cause of disease caused by aberrant expression or imbalance of a gene. With the additional RNA-seq experiments we have planned, we will be able to investigate whether the optimal combination for robust pathogenic variant detection data is WGS with RNA-seq, ATAC-seq, or both.

The burgeoning number of referrals for molecular diagnostics and the amount of data it is now possible to generate from each can feel overwhelming, but it also offers incredible opportunity to benefit patients. The marriage of data availability from many different sequencing technologies, the dawn of predictive AI contribution to precision medicine, and use of CRISPR-based technologies for genome editing is a heady one. The challenge to improve clinical diagnostics is now suffused with fresh purpose, that not only can a diagnosis direct treatment, but possibly even provide a cure.

7 Acknowledgements

A word for the souls that have toil'd, and wrought, and thought with me.

First and greatest thanks, my gratitude and admiration eternal to my supervisor Yenan. The opportunities you have given me have been extraordinary. I will never

forget the 10pm phone call asking how I would like to go and live in New York for some of my PhD! Your belief in me and my projects has often exceeded my own, and I can't wait to see where we go from here.

My co-supervisors have all enriched my PhD immeasurably. Mikael, who has been my instructor in clinical matters: your dedication to your patients is an inspiration, and so important for we who work in the warehouse (so to speak) to experience. Thank you for letting me inhabit your world for a day, and for all the patient samples! Anna, you have always been available with useful comments, support, and invitations to meet Nobel prize winners whenever I needed. Bea, I treasure the memories both of working with you and of mentorship outside the lab as well. Many thanks to all three of you, and to my mentor Janne Lehtiö for the excellent career advice.

Bryceson group members past and present! Thanks to the greatest team: Lars, Anna-Rita, Caroline, Dona, Tak, Lamberto, and Paolo. Sigrid, thanks for being so positive! Petar, please never stop sending me memes about the PhD experience even long after we have both graduated (and we will both graduate). Jelve, you have made my day so often with your kind comments. Ram and Tim, your data has been instrumental in my projects, and Tim it was nice to have another Brit around! If it weren't for your understanding of EastEnders jokes I would have just been screaming cultural references into the abyss. Bianca and Marie, thanks for your formidable powers of diagnostics without which at least two of these publications would not have been possible. Sam, it was a joy to work on the exocytosis paper with you.

I did in fact jump at that chance to spend time in New York! Thanks to Jean-Laurent Casanova for hosting me in his lab, and to Qian Zhang, Yoann Seeleuthner, Bertrand Boisson, and Aurélie Cobat for providing science-based support. Equally, thanks to Peng Zhang, Ahmad Yatim, Yi Feng, Jessica Peel, Daniyel Lee, Baptiste Milisavljevic, and Matthieu Chaldebas for providing social life-based support.

Not only Rockefeller, but other departments at KI have welcomed me and provided help where they could. At Clinical Genomics, thanks to Valtteri for the affiliation; and to Jesper, Esmee, Anders, Ram, and Kristine for their knowledge of rare diseases and clinical diagnostic pipelines. At Clinical Immunology: Sofie, Peter, Anton, Malin, your energy seems boundless and it has been a sincere pleasure to work with you.

Back at HERM, we are blessed with incredible teams who run the core facilities, HR, and generally make everything tick. Thanks to the flow cytometry team, BEA, and over at Uppsala to UPPMAX; and to the wonderfully forbearing admin people of HERM during my PhD, Sri, Sara, Annette, and Elin. You are all treasures.

All past and present members of HERM, who have been so generous in many cases not only with their time and energy but also with their blood. Special thanks to Julia and Nicolai for being delightful human beings, and for help with RNA-seq (and that time I needed your thermal cycler booking. Sorry I ruined that experiment); Filip and Jonas, my co-PhD student representatives, who have somehow made both early morning systembolaget runs and administrative labour fun; Franca, thank heavens someone else was going through the defense craziness along with me, you will be amazing!; Pedro, for help with Seurat, grant proposals, and being totally indefatigable; Axel, your speed at Bayesian statistics is phenomenal, thank you for sharing your skills and your enthusiasm for art, literature, and wandering round Stockholm with me. Lucia, you have been my sanity check for such a long time, both in science and life! Also Caroline L, Elory, Nici, Stefy, Madde, Matilda, Sophia, Ece, and many more.

The rare disease diagnostics field has just the warmest community, and I feel lucky every day that I landed in the most welcoming discipline. My conference buddy Brian Schilder, thank you for letting me hang out with the infinitely cool Imperial gang! I also want to give special thanks to the incredible selfless women of rare disease who have gone out of their way to help me: from just making the time to talk to me, to putting me in touch with useful people, or suggesting me for a talk, interview, or prize, for no reason other than kindness. Lucia Pena Perez (yes you got name-checked twice), Gabrielle Lemire, Michelle Li, Sanna Gudmundsson, Sarah Stenton, Shilpa Kobren, you are my heroes. First among equals, Dianne Newbury, who hired me in her lab at Oxford when I was 17 years old. The chance you took on me changed my life.

Finally, thanks to my support crew these past 5 years. In the lab: Heinrich, your immunology expertise is truly extraordinary. Also, I want you to know that when you told me at 2am during our road trip in rural Italy that something was wrong with the car, it was one of the worst moments of my life. Tessa, my partner in crime and patient work, I have no idea how I will function without my other half! Giovi, my PhD sister, our lovely unicorn, the lab has been less bright without you. Elisa, I will miss getting the giggles in yoga with you, thank you for matching the depravity of my humour. I love you all to bits. Outside the lab, my book club Emma, Ida, and Sanna, have been my greatest supporters. Alice, who has been stalwart through school, weddings, and now a combined 6 degrees. We seriously need to leave education. Alex Roberts, my emotional and occasionally physical crutch for almost 15 years. Sean (only because the list is already so long that further recognition is meaningless; this doesn't mean I like you). My parents and sister, who have been so very tolerant of my weird inclinations to move to other countries and spend most waking moments either doing science, or talking about it. And Daniel, thank you always, for so much more than a mere PhD.

8 Declaration about the use of generative AI

No AI assisted tools were used in writing the thesis kappa. I take full responsibility for the content of the thesis kappa.

9 References

1. Vogel F. A Preliminary Estimate of the Number of Human Genes. Nature. 1964;201:847.

2. Salzberg SL. Open questions: How many genes do we have? Vol. 16, BMC Biology. BioMed Central Ltd.; 2018.

3. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze A V., Mikheenko A, et al. The complete sequence of a human genome. Science (1979). 2022 Apr 1;376(6588):44–53.

4. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. European Journal of Human Genetics. 2020 Feb 1;28(2):165–73.

5. Croft P, Altman DG, Deeks JJ, Dunn KM, Hay AD, Hemingway H, et al. The science of clinical practice: Disease diagnosis or patient prognosis? Evidence about 'what is likely to happen' should shape clinical practice. BMC Med. 2015 Jan 30;13(1).

6. Bauskis A, Strange C, Molster C, Fisher C. The diagnostic odyssey: insights from parents of children living with an undiagnosed condition. Orphanet J Rare Dis. 2022 Dec 1;17(1).

7. Spillmann RC, McConkie-Rosell A, Pena L, Jiang YH, Schoch K, Walley N, et al. A window into living with an undiagnosed disease: Illness narratives from the Undiagnosed Diseases Network. Orphanet J Rare Dis. 2017 Apr 17;12(1).

8. Benito-Lozano J, Arias-Merino G, Gómez-Martínez M, Arconada-López B, Ruiz-García B, de la Paz MP, et al. Psychosocial impact at the time of a rare disease diagnosis. PLoS One. 2023 Jul 1;18(7 JULY).

9. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009 Sep 10;461(7261):272–6.

10. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. New England Journal of Medicine. 2012 Nov 15;367(20):1921–9.

11. De Vries BBA, Pfundt R, Leisink M, Koolen DA, Vissers LELM, Janssen IM, et al. Diagnostic Genome Profiling in Mental Retardation. Am J Hum Genet. 2005;77:606–16.

12. Pickering DL, Eudy JD, Olney AH, Dave BJ, Golden D, Stevens J, et al. Array-based comparative genomic hybridization analysis of 1176 consecutive clinical genetics investigations. Genetics in Medicine. 2008 Apr;10(4):262–6.

13. Baris HN, Tan WH, Kimonis VE, Irons MB. Diagnostic utility of array-based comparative genomic hybridization in a clinical setting. Am J Med Genet A. 2007 Nov 1;143(21):2523–33.

14. Pfundt R, Del Rosario M, Vissers LELM, Kwint MP, Janssen IM, De Leeuw N, et al. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. Genetics in Medicine. 2017 Jun 1;19(6):667–75.

15. Retterer K, Scuffins J, Schmidt D, Lewis R, Pineda-Alvarez D, Stafford A, et al. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. Genetics in Medicine. 2015 Aug 6;17(8):623–9.

16. Smedley D, Smith KR, Martin A, Thomas EA, McDonagh EM, Cipriani V, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care — Preliminary Report. New England Journal of Medicine [Internet]. 2021 Nov 11;385(20):1868–80. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2035790

17. Mardis ER. The $1,000 genome, the $100,000 analysis? Genome Med. 2010;2(11).

18. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Vol. 526, Nature. Nature Publishing Group; 2015. p. 68–74.

19. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May 28;581(7809):434–43.

20. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes [Internet]. 2022. Available from: https://doi.org/10.1101/2022.03.20.485034

21. Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, et al. Insights into human genetic variation and population history from 929 diverse genomes. Science (1979). 2020 Mar 20;367(6484).

22. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nature. 2021 Nov 25;599(7886):628–34.

23. Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, et al. SweGen: A whole-genome data resource of genetic variability in a cross-section of the Swedish population. European Journal of Human Genetics. 2017 Nov 1;25(11):1253–60.

24. 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. Int J Epidemiol. 2015 Aug 1;44(4):1137–47.

25. Miller DT, Lee K, Gordon AS, Amendola LM, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2021 update: a policy statement of the American College of Medical Genetics and Genomics (ACMG). Genetics in Medicine. 2021 Aug 1;23(8):1391–8.

26. de Wert G, Dondorp W, Clarke A, Dequeker EMC, Cordier C, Deans Z, et al. Opportunistic genomic screening. Recommendations of the European Society of Human Genetics. European Journal of Human Genetics. 2021 Mar 1;29(3):365–77.

27. Birgmeier J, Haeussler M, Deisseroth CA, Steinberg EH, Jagadeesh KA, Ratner AJ, et al. AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature [Internet]. Vol. 12, Sci. Transl. Med. 2020. Available from: https://www.science.org

28. Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, et al. Deep Phenotyping on Electronic Health Records Facilitates Genetic Diagnosis by Clinical Exomes. Am J Hum Genet. 2018 Jul 5;103(1):58–73.

29. Deisseroth CA, Birgmeier J, Bodle EE, Kohler JN, Matalon DR, Nazarenko Y, et al. ClinPhen extracts and prioritizes patient phenotypes directly from medical records to expedite genetic disease diagnosis. Genetics in Medicine. 2019 Jul 1;21(7):1585–93.

30. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. Hum Mutat. 2015 Oct 1;36(10):915–21.

31. Haimel M, Pazmandi J, Heredia RJ, Dmytrus J, Bal SK, Zoghi S, et al. Curation and expansion of Human Phenotype Ontology for defined groups of inborn errors of immunity. Journal of Allergy and Clinical Immunology. 2022 Jan 1;149(1):369–78.

32. Bousfiha A, Moundir A, Tangye SG, Picard C, Jeddane L, Al-Herz W, et al. The 2022 Update of IUIS Phenotypical Classification for Human Inborn Errors of Immunity. J Clin Immunol. 2022 Oct 1;42(7):1508–20.

33. Dorsey M, Puck J. Newborn screening for severe combined immunodeficiency in the US: Current status and approach to management. Vol. 3, International Journal of Neonatal Screening. MDPI Multidisciplinary Digital Publishing Institute; 2017.

34. Barzaghi F, Aiuti A. Newborn screening for severe combined immunodeficiency: changing the landscape of post-transplantation survival. Vol. 402, The Lancet. Elsevier B.V.; 2023. p. 84–5.

35. Stray-Pedersen A, Sorte HS, Samarakoon P, Gambin T, Chinn IK, Coban Akdemir ZH, et al. Primary immunodeficiency diseases: Genomic approaches delineate heterogeneous Mendelian disorders. Journal of Allergy and Clinical Immunology. 2017 Jan 1;139(1):232–45.

36. Thaventhiran JED, Lango Allen H, Burren OS, Rae W, Greene D, Staples E, et al. Whole-genome sequencing of a sporadic primary immunodeficiency cohort. Nature [Internet]. 2020 Jul 2;583(7814):90–5. Available from: https://www.nature.com/articles/s41586-020-2265-1

37. Barry M, Bleackley RC. Cytotoxic T lymphocytes: All roads lead to death. Vol. 2, Nature Reviews Immunology. European Association for Cardio-Thoracic Surgery; 2002. p. 401–9.

38. Henter JI, Horne A, Aricó M, Egeler RM, Filipovich AH, Imashuku S, et al. HLH-2004: Diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. Pediatr Blood Cancer. 2007 Feb;48(2).

39. De Saint Basile G, Ménasché G, Fischer A. Molecular mechanisms of biogenesis and exocytosis of cytotoxic granules. Vol. 10, Nature Reviews Immunology. 2010. p. 568–79.

40. Al-Samkari H, Berliner N. Hemophagocytic Lymphohistiocytosis. Annual Review of Pathology: Mechanisms of Disease [Internet]. 2018;13:27–49. Available from: https://doi.org/10.1146/annurev-pathol-020117-

41. Chinn IK, Eckstein OS, Peckham-Gregory EC, Goldberg BR, Forbes LR, Nicholas SK, et al. Genetic and mechanistic diversity in pediatric hemophagocytic lymphohistiocytosis. Blood [Internet]. 2018 Jul 5;132(1):89–100. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29632024

42. Osińska I, Popko K, Demkow U. Perforin: An important player in immune response. Vol. 39, Central European Journal of Immunology. Termedia Publishing House Ltd.; 2014. p. 109–15.

43. Smyth MJ, Thia KYT, Street SEA, Macgregor D, Godfrey DI, Trapani JA. Perforin-mediated Cytotoxicity Is Critical for Surveillance of Spontaneous Lymphoma. J Exp Med [Internet]. 2000;192(5):755–60. Available from: http://www.jem.org/cgi/content/full/192/5/755

44. Stepp SE, Dufourcq-Lagelouse R, Deist F Le, Bhawan S, Certain S, Mathew PA, et al. Perforin Gene Defects in Familial Hemophagocytic Lymphohistiocytosis. Science (1979). 1999 Dec 3;286(5446):1957–9.

45. zur Stadt U, Rohr J, Seifert W, Koch F, Grieve S, Pagel J, et al. Familial Hemophagocytic Lymphohistiocytosis Type 5 (FHL-5) Is Caused by Mutations in Munc18-2 and Impaired Binding to Syntaxin 11. Am J Hum Genet. 2009 Oct 9;85(4):482–92.

46. Côte M, Ménager MM, Burgess A, Mahlaoui N, Picard C, Schaffner C, et al. Munc18-2 deficiency causes familial hemophagocytic lymphohistiocytosis type 5 and impairs cytotoxic granule exocytosis in patient NK cells. Journal of Clinical Investigation. 2009 Dec 1;119(12):3765–73.

47. zur Stadt U, Schmidt S, Kasper B, Beutel K, Diler AS, Henter JI, et al. Linkage of familial hemophagocytic lymphohistiocytosis (FHL) type-4 to chromosome 6q24 and identification of mutations in syntaxin 11. Hum Mol Genet. 2005 Mar 15;14(6):827–34.

48. rô me Feldmann J, Callebaut I, Raposo G, phanie Certain S, Bacq D, cile Dumont C, et al. Munc13-4 Is Essential for Cytolytic Granules Fusion and Is Mutated in a Form of Familial Hemophagocytic Lymphohistiocytosis (FHL3) cytotoxic granules at the immunological synapse. HMunc13-4 is therefore essential for the priming step of cytolytic granules secretion preceding vesicle membrane fusion [Internet]. Vol. 115, Cell. 2003. Available from: www.ncbi.nlm.nih.gov/mapview

49. Bryceson YT, Rudd E, Zheng C, Edner J, Ma D, Wood SM, et al. Defective cytotoxic lymphocyte degranulation in syntaxin-11-deficient familial hemophagocytic lymphohistiocytosis 4 (FHL4) patients. Blood. 2007 Sep 15;110(6):1906–15.

50. Löfstedt A, Ahlm C, Tesi B, Bergdahl IA, Nordenskjöld M, Bryceson YT, et al. Haploinsufficiency of UNC13D increases the risk of lymphoma. Cancer. 2019 Jun 1;125(11):1848–54.

51. Kalinichenko A, Perinetti Casoni G, Dupré L, Trotta L, Huemer J, Galgano D, et al. RhoG deficiency abrogates cytotoxicity of human lymphocytes and causes hemophagocytic lymphohistiocytosis. Blood. 2021 Apr 15;137(15).

52. Enders A, Zieger B, Schwarz K, Yoshimi A, Speckmann C, Knoepfle EM, et al. Lethal hemophagocytic lymphohistiocytosis in Hermansky-Pudlak syndrome type II. Blood. 2006 Jul 1;108(1):81–7.

53. Nagle DL, Karim MA, Woolf EA, Holmgren L, Bork3 P, Misumi1 DJ, et al. Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome. Nat Genet [Internet]. 1996;14:307–11. Available from: http://www.nature.com/naturegenetics

54. Ménasché G, Pastural E, Feldmann J, Certain S, Ersoy F, Dupuis S, et al. Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome. Nat Genet [Internet]. 2000;25:173. Available from: http://genetics.nature.com

55. Meeths M, Chiang SCC, Wood SM, Entesarian M, Schlums H, Bang B, et al. Familial hemophagocytic lymphohistiocytosis type 3 (FHL3) caused by deep intronic mutation and inversion in UNC13D. Blood. 2011 Nov 24;118(22):5783–93.

56. Seo JY, Song JS, Lee KO, Won HH, Kim JW, Kim SH, et al. Founder effects in two predominant intronic mutations of UNC13D, c.118-308C>T and c.754-1G>C underlie the unusual predominance of type 3 familial hemophagocytic lymphohistiocytosis (FHL3) in Korea. Ann Hematol. 2013 Mar;162:413–27.

57. Entesarian M, Chiang SC, Schlums H, Meeths M, Chan MY, Mya SN, et al. Novel deep intronic and missense UNC13D mutations in familial haemophagocytic lymphohistiocytosis type 3. Br J Haematol. 2013 Aug;162(3):413–5.

58. Tesi B, Rascon J, Chiang SCC, Burnyte B, Löfstedt A, Fasth A, et al. A RAB27A 5′ untranslated region structural variant associated with late-onset hemophagocytic lymphohistiocytosis and normal pigmentation. Journal of Allergy and Clinical Immunology. 2018 Jul;142(1):317-321.e8.

59. Meeths M, Bryceson YT. Genetics and pathophysiology of haemophagocytic lymphohistiocytosis. Vol. 110, Acta Paediatrica, International Journal of Paediatrics. John Wiley and Sons Inc; 2021. p. 2903–11.

60. Mariathasan S, Newton K, Monack DM, Vucic D, French DM, Lee WP, et al. Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. Nature [Internet]. 2004 Jul 8;(430):213–8. Available from: www.nature.com/nature

61. Broz P, Dixit VM. Inflammasomes: Mechanism of assembly, regulation and signalling. Vol. 16, Nature Reviews Immunology. Nature Publishing Group; 2016. p. 407–20.

62. Canna SW, De Jesus AA, Gouni S, Brooks SR, Marrero B, Liu Y, et al. An activating NLRC4 inflammasome mutation causes autoinflammation with recurrent macrophage activation syndrome. Nat Genet. 2014 Sep 26;46(10):1140–6.

63. Castro CN, Rosenzwajg M, Carapito R, Shahrooei M, Konantz M, Khan A, et al. NCKAP1L defects lead to a novel syndrome combining immunodeficiency, lymphoproliferation, and hyperinflammation. Journal of Experimental Medicine. 2020 Aug 6;217(12).

64. Gernez Y, de Jesus AA, Alsaleem H, Macaubas C, Roy A, Lovell D, et al. Severe autoinflammation in 4 patients with C-terminal variants in cell division control protein 42 homolog (CDC42) successfully treated with IL-1β inhibition. Journal of Allergy and Clinical Immunology. 2019 Oct 1;144(4):1122-1125.e6.

65. Lam MT, Coppola S, Krumbach OHF, Prencipe G, Insalaco A, Cifaldi C, et al. A novel disorder involving dyshematopoiesis, inflammation, and HLH due to aberrant CDC42 function. Journal of Experimental Medicine. 2019 Dec 1;216(12):2778–99.

66. Cook S, Lenardo MJ, Freeman AF. HEM1 Actin Immunodysregulatory Disorder: Genotypes, Phenotypes, and Future Directions. Vol. 42, Journal of Clinical Immunology. Springer; 2022. p. 1583–92.

67. Booth C, Gilmour KC, Veys P, Gennery AR, Slatter MA, Chapel H, et al. X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: A multicenter study on the manifestations, management and outcome of the disease. Blood. 2011 Jan 6;117(1):53–62.

68. Rigaud S, Fondanèche MC, Lambert N, Pasquier B, Mateo V, Soulas P, et al. XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome. Nature. 2006 Nov 2;444(7115):110–4.

69. Schmid JP, Canioni D, Moshous D, Touzot F, Mahlaoui N, Hauck F, et al. Clinical similarities and differences of patients with X-linked lymphoproliferative syndrome type 1 (XLP-1/SAP deficiency) versus type 2 (XLP-2/XIAP deficiency). Blood. 2011 Feb 3;117(5):1522–9.

70. Matsuda-Lennikov M, Biancalana M, Zou J, Ravell JC, Zheng L, Kanellopoulou C, et al. Magnesium transporter 1 (MAGT1) deficiency causes selective defects in N-linked glycosylation and expression of immune-response genes. Journal of Biological Chemistry. 2019 Sep 13;294(37):13638–56.

71. Li FY, Chaigne-Delalande B, Su H, Uzel G, Matthews H, Lenardo MJ. XMEN disease: a new primary immunodeficiency affecting Mg 21 regulation of immunity against Epstein-Barr virus. Blood [Internet]. 2014 Apr 3;2148–2152(123). Available from: http://ashpublications.org/blood/article-pdf/123/14/2148/1374984/2148.pdf

72. Passarelli C, Civino A, Rossi MN, Cifaldi L, Lanari V, Moneta GM, et al. IFNAR2 Deficiency Causing Dysregulation of NK Cell Functions and Presenting With Hemophagocytic Lymphohistiocytosis. Front Genet. 2020 Sep 18;11.

73. Gothe F, Stremenova Spegarova J, Hatton CF, Griffin H, Sargent T, Cowley SA, et al. Aberrant inflammatory responses to type I interferon in STAT2 or IRF9 deficiency. Journal of Allergy and Clinical Immunology. 2022 Oct 1;150(4):955-964.e16.

74. Gothe F, Hatton CF, Truong L, Klimova Z, Kanderova V, Fejtkova M, et al. A Novel Case of Homozygous Interferon Alpha/Beta Receptor Alpha Chain (IFNAR1) Deficiency With Hemophagocytic Lymphohistiocytosis. Clinical Infectious Diseases. 2022 Jan 1;74(1):136–9.

75. Duncan CJA, Skouboe MK, Howarth S, Hollensen AK, Chen R, Børresen ML, et al. Life-threatening viral disease in a novel form of autosomal recessive IFNAR2 deficiency in the Arctic. Journal of Experimental Medicine. 2022 Jun 6;219(6).

76. Vavassori S, Chou J, Faletti LE, Haunerdinger V, Opitz L, Joset P, et al. Multisystem inflammation and susceptibility to viral infections in human ZNFX1 deficiency. Journal of Allergy and Clinical Immunology. 2021 Aug 1;148(2):381–93.

77. Bianca Tesi, Elena Sieni, Conceicao Neves, Francesca Romano, Valentina Cetica, Ana Isabel Cordeiro, et al. Hemophagocytic lymphohistiocytosis in 2 patients with underlying IFN-g receptor deficiency. Journal of Allergy and Clinical Immunology. 2015 Jun 1;135(6):1638–41.

78. Mandey SHL, Schneiders MS, Koster J, Waterham HR. Mutational spectrum and genotype–phenotype correlations in mevalonate kinase deficiency. Hum Mutat. 2006 Aug;27(8):796–802.

79. Janka GE, Lehmberg K. Hemophagocytic syndromes - An update. Blood Rev. 2014;28(4):135–42.

80. Brunner KT, Mauel J, Cerottini JC, Chapuis B. Quantitative Assay of the Lytic Action of Immune Lymphoid Cells on 5"Cr-Labelled Allogeneic Target Cells In vitro; Inhibition by Isoantibody and by Drugs. Vol. 14, Immunology. 1968.

81. Bryceson YT, Pende D, Maul-Pavicic A, Gilmour KC, Ufheil H, Vraetz T, et al. A prospective evaluation of degranulation assays in the rapid diagnosis of familial hemophagocytic syndromes. 2012; Available from: http://ashpublications.org/blood/article-pdf/119/12/2754/1351142/zh801212002754.pdf

82. Chiang SCC, Theorell J, Entesarian M, Meeths M, Mastafa M, Al-Herz W, et al. Comparison of primary human cytotoxic T-cell and natural killer cell responses reveal similar molecular requirements for lytic granule exocytosis but differences in cytokine production. Blood. 2013 Feb 21;121(8):1345–56.

83. Böhm S, Wustrau K, Schmid JP, Prader S, Ahlmann M, Yacobovich J, et al. Survival in primary hemophagocytic lymphohistiocytosis, 2016 to 2021: etoposide is better than its reputation. Blood. 2024 Mar 7;10(143):872–81.

84. https://clinicaltrials.gov/study/NCT03113760.

85. Canna SW, Girard C, Malle L, de Jesus A, Romberg N, Kelsen J, et al. Life-threatening NLRC4-associated hyperinflammation successfully treated with IL-18 inhibition. Journal of Allergy and Clinical Immunology. 2017 May 1;139(5):1698–701.

86. Rosée P La, Rosée R, Horne A, Hines M, Von Bahr Greenwood T, Machowicz R, et al. Recommendations for the management of hemophagocytic lymphohistiocytosis in adults. Blood [Internet]. 2019 Jun 6;133(23). Available from: http://ashpublications.org/blood/article-pdf/133/23/2465/1553600/blood894618.pdf

87. Bratton SL, Duker H Van, Statler KD, Pulsipher MA, McArthur J, Keenan HT. Lower hospital mortality and complications after pediatric hematopoietic stem cell transplantation. Crit Care Med. 2008;36(3):923–7.

88. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281–308.

89. Frangoul H, Altshuler D, Cappellini MD, Chen YS, Domm J, Eustace BK, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. New England Journal of Medicine. 2021 Jan 21;384(3):252–60.

90. Longhurst HJ, Lindsay K, Petersen RS, Fijen LM, Gurugama P, Maag D, et al. CRISPR-Cas9 In Vivo Gene Editing of KLKB1 for Hereditary Angioedema . New England Journal of Medicine. 2024 Feb;390(5):432–41.

91. Dettmer-Monaco V, Weißert K, Ammann S, Monaco G, Lei L, Gräßel L, et al. Gene editing of hematopoietic stem cells restores T-cell response in familial hemophagocytic

lymphohistiocytosis. Journal of Allergy and Clinical Immunology. 2024 Jan 1;153(1):243-255.e14.

92. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. New England Journal of Medicine. 2020 Feb 20;382(8):727–33.

93. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet. 2020 Feb;395(10223):470–3.

94. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection. A narrative review. Ann Intern Med. 2020 Sep 1;173(5):362–8.

95. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, et al. Morbidity and Mortality Weekly Report Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility-King County, Washington, March 2020 [Internet]. Public Health-Seattle & King County. 2019. Available from: https://www.cdc.gov/mmwr

96. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Eurosurveillance. 2020 Sep 21;25(10).

97. Guan W jie, Ni Z yi, Hu Y, Liang W hua, Ou C quan, He J xing, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine. 2020 Apr 30;382(18):1708–20.

98. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA - Journal of the American Medical Association. 2020 Mar 17;323(11):1061–9.

99. Klok FA, Kruip MJHA, van der Meer NJM, Arbous MS, Gommers DAMPJ, Kant KM, et al. Incidence of thrombotic complications in critically ill ICU patients with COVID-19. Thromb Res. 2020 Jul 1;191:145–7.

100. Chen T, Wu D, Chen H, Yan W, Yang D, Chen G, et al. Clinical characteristics of 113 deceased patients with coronavirus disease 2019: Retrospective study. The BMJ. 2020 Mar 26;368.

101. Musuuza JS, Watson L, Parmasad V, Putman-Buehler N, Christensen L, Safdar N. Prevalence and outcomes of co-infection and superinfection with SARS-CoV-2 and other pathogens: A systematic review and metaanalysis. PLoS One. 2021 May 1;16(5 May).

102. Liotta EM, Batra A, Clark JR, Shlobin NA, Hoffman SC, Orban ZS, et al. Frequent neurologic manifestations and encephalopathy-associated morbidity in Covid-19 patients. Ann Clin Transl Neurol. 2020 Nov 1;7(11):2221–30.

103. World Health Organisation. https://covid19.who.int/?adgroupsurvey=%7Badgroupsurvey%7D&gclid=Cj0KCQiAw8OeBhCeARIsAGxWtUxR9wuVfKmK8awaDLSRrSm65bKiEJOcvWX34XpP2LX4eD_sR9ZfUqAaArNPEALw_wcB.

104. Wu Z, McGoogan JM. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. Vol. 323, JAMA - Journal of the American Medical Association. American Medical Association; 2020. p. 1239–42.

105. Stokes EK, Zambrano LD, Anderson KN, Marder EP, Raz KM, El Burai Felix S, et al. Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020. MMWR Morb Mortal Wkly Rep. 2020 Jun 19;69(24):759–65.

106. Ioannidis JPA. Reconciling estimates of global spread and infection fatality rates of COVID-19: An overview of systematic evaluations. Vol. 51, European Journal of Clinical Investigation. Blackwell Publishing Ltd; 2021.

107. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection fatality rates. International Journal of Infectious Diseases. 2020 Dec 1;101:138–48.

108. COVID-19 Forecasting Team. Variation in the COVID-19 infection-fatality ratio by age, time, and geography during the pre-vaccine era: a systematic analysis. Lancet. 2022 Apr 16;399(10334):1469–88.

109. Kompaniyets L, Goodman AB, Belay B, Freedman DS, Sucosky MS, Lange SJ, et al. Body Mass Index and Risk for COVID-19–Related Hospitalization, Intensive Care Unit Admission, Invasive Mechanical Ventilation, and Death — United States, March–December 2020. MMWR Morb Mortal Wkly Rep. 2021 Mar 12;70(10):355–61.

110. Lighter J, Phillips M, Hochman S, Stirling S, Johnson D, Francois F, et al. Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission. Clinical Infectious Diseases. 2020;896–7.

111. Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. PLoS Med. 2020 Sep 1;17(9).

112. Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: Prospective cohort study. The BMJ. 2020 May 22;369.

113. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020 Aug 20;584(7821):430–6.

114. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020 May 1;8(5):475–81.

115. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. JAMA. 2020 Apr 7;323(13):1239.

116. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. Lancet Infect Dis. 2020 Jun 1;20(6):669–77.

117. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020 May 26;323(20):2052.

118. Casanova JL, Su HC, Abel L, Aiuti A, Almuhsen S, Arias AA, et al. A Global Effort to Define the Human Genetics of Protective Immunity to SARS-CoV-2 Infection. Cell. 2020 Jun 11;181(6):1194–9.

119. Zhang SY, Zhang Q, Casanova JL, Su HC, Abel L, Bastard P, et al. Severe COVID-19 in the young and healthy: monogenic inborn errors of immunity? Vol. 20, Nature Reviews Immunology. Nature Research; 2020. p. 455–6.

120. Lemarquis A, Campbell T, Aranda-Guillén M, Hennings V, Brodin P, Kämpe O, et al. Severe COVID-19 in an APS1 patient with interferon autoantibodies treated with plasmapheresis. Vol. 148, Journal of Allergy and Clinical Immunology. Mosby Inc.; 2021. p. 96–8.

121. Schidlowski L, Iwamura APD, Abel L, Bastard P, Bustamante J, Casanova JL, et al. Diagnosis of APS-1 in Two Siblings Following Life-Threatening COVID-19 Pneumonia. J Clin Immunol [Internet]. 2022 May 19;42(4):749–52. Available from: https://link.springer.com/10.1007/s10875-022-01245-1

122. Bastard P, Orlova E, Sozaeva L, Lévy R, James A, Schmitt MM, et al. Preexisting autoantibodies to type I IFNs underlie critical COVID-19 pneumonia in patients with APS-1. J Exp Med. 2021 Jul 5;218(7).

123. Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann HH, Zhang Y, et al. Autoantibodies against type I IFNs in patients with life-threatening COVID-19. Science (1979). 2020 Oct 23;370(6515).

124. The risk of COVID-19 death is much greater and age dependent with type I IFN autoantibodies. 2022; Available from: https://doi.org/10.1073/pnas.2200413119

125. Koning R, Bastard P, Casanova JL, Brouwer MC, van de Beek D, van Agtmael M, et al. Autoantibodies against type I interferons are associated with multi-organ failure in COVID-19 patients. Vol. 47, Intensive Care Medicine. Springer Science and Business Media Deutschland GmbH; 2021. p. 704–6.

126. Solanich X, Rigo-Bonnin R, Gumucio VD, Bastard P, Rosain J, Philippot Q, et al. Pre-existing Autoantibodies Neutralizing High Concentrations of Type I Interferons in Almost 10% of COVID-19 Patients Admitted to Intensive Care in Barcelona. J Clin Immunol. 2021 Nov 1;41(8):1733–44.

127. Chauvineau-Grenier A, Bastard P, Servajean A, Gervais A, Rosain J, Jouanguy E, et al. Autoantibodies Neutralizing Type I Interferons in 20% of COVID-19 Deaths in a French Hospital. J Clin Immunol. 2022 Apr 1;42(3):459–70.

128. Abers MS, Rosen LB, Delmonte OM, Shaw E, Bastard P, Imberti L, et al. Neutralizing type-I interferon autoantibodies are associated with delayed viral clearance and intensive care unit admission in patients with COVID-19. Immunol Cell Biol. 2021 Oct 1;99(9):917–21.

129. Reizis B. Plasmacytoid Dendritic Cells: Development, Regulation, and Function. Vol. 50, Immunity. Cell Press; 2019. p. 37–50.

130. Schoggins JW. Interferon-Stimulated Genes: What Do They All Do? Annu Rev Virol [Internet]. 2019;14(1). Available from: https://doi.org/10.1146/annurev-virology-092818-

131. Evy RL, Zhang P, Bastard P, Dorgham K, Melki I, Hadchouel A, et al. Monoclonal antibody-mediated neutralization of SARS-CoV-2 in an IRF9-deficient child. Available from: https://doi.org/10.1073/pnas.2114390118

132. Campbell TM, Liu Z, Zhang Q, Moncada-Velez M, Covill LE, Zhang P, et al. Respiratory viral infections in otherwise healthy humans with inherited IRF7 deficiency. Journal of Experimental Medicine [Internet]. 2022 Jul 4;219(7). Available from: https://rupress.org/jem/article/219/7/e20220202/213267/Respiratory-viral-infections-in-otherwise-healthy

133. Abolhassani H, Landegren N, Bastard P, Materna M, Modaresi M, Du L, et al. Inherited IFNAR1 Deficiency in a Child with Both Critical COVID-19 Pneumonia and Multisystem Inflammatory Syndrome. J Clin Immunol. 2022 Apr 1;42(3):471–83.

134. Duncan CJA, Skouboe MK, Howarth S, Hollensen AK, Chen R, Børresen ML, et al. Life-threatening viral disease in a novel form of autosomal recessive IFNAR2 deficiency in the Arctic . Journal of Experimental Medicine. 2022 Jun 6;219(6).

135. Lévy R, Bastard P, Lanternier F, Lecuit M, Zhang SY, Casanova JL. IFN-α2a Therapy in Two Patients with Inborn Errors of TLR3 and IRF3 Infected with SARS-CoV-2. Available from: https://doi.org/10.1007/s10875-020-00933-0

136. Van Der Made CI, Simons A, Schuurs-Hoeijmakers J, Van Den Heuvel G, Mantere T, Kersten S, et al. Presence of Genetic Variants among Young Men with Severe COVID-19. JAMA - Journal of the American Medical Association. 2020 Aug 18;324(7):663–73.

137. Abolhassani H, Vosughimotlagh A, Asano T, Landegren N, Boisson B, Delavari S, et al. X-Linked TLR7 Deficiency Underlies Critical COVID-19 Pneumonia in a Male Patient with Ataxia-Telangiectasia. J Clin Immunol. 2022 Jan 1;42(1).

138. Asano T, Boisson B, Onodi F, Matuozzo D, Moncada-Velez M, Maglorius Renkilaraj MRL, et al. X-linked recessive TLR7 deficiency in ~1% of men under 60 years old with life-threatening COVID-19. Sci Immunol. 2021 Aug 10;6(62).

139. Dexamethasone in Hospitalized Patients with Covid-19. New England Journal of Medicine. 2021 Feb 25;384(8).

140. Zhang K, Jordan MB, Marsh RA, Johnson JA, Kissell D, Meller J, et al. Hypomorphic mutations in PRF1, MUNC13-4, and STXBP2 are associated with adult-onset familial HLH. Blood. 2011 Nov 24;118(22):5794–8.

141. Luo H, Liu D, Liu W, Wang G, Chen L, Cao Y, et al. Germline variants in UNC13D and AP3B1 are enriched in COVID-19 patients experiencing severe cytokine storms. European Journal of Human Genetics. 2021 Aug 19;29(8).

142. Reiff DD, Zhang M, Smitherman EA, Mannion ML, Stoll ML, Weiser P, et al. A Rare STXBP2 Mutation in Severe COVID-19 and Secondary Cytokine Storm Syndrome. Life. 2022 Feb 1;12(2).

143. Vagrecha A, Zhang M, Acharya S, Lozinsky S, Singer A, Levine C, et al. Hemophagocytic Lymphohistiocytosis Gene Variants in Multisystem Inflammatory Syndrome in Children. Biology (Basel). 2022 Mar 1;11(3).

144. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. New England Journal of Medicine [Internet]. 2020 Oct 15;383(16):1522–34. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2020283

145. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. Nature [Internet]. 2021 Mar 4;591(7848):92–8. Available from: http://www.nature.com/articles/s41586-020-03065-y

146. Downes DJ, Cross AR, Hua P, Roberts N, Schwessinger R, Cutler AJ, et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. Nat Genet. 2021 Nov 1;53(11):1606–15.

147. Kosmicki JA, Marcketta A, Sharma D, Di Gioia SA, Batista S, Yang XM, et al. Genetic risk factors for COVID-19 and influenza are largely distinct. Nat Genet [Internet]. 2024 Aug 5;56(8):1592–6. Available from: https://www.nature.com/articles/s41588-024-01844-1

148. Clementi R, Emmi L, Maccario R, Liotta F, Moretta L, Danesino C, et al. Adult onset and atypical presentation of hemophagocytic lymphohistiocytosis in siblings carrying PRF1 mutations. Blood. 2002 Sep 15;100(6).

149. Schulert GS, Zhang M, Fall N, Husami A, Kissell D, Hanosh A, et al. Whole-exome sequencing reveals mutations in genes linked to hemophagocytic lymphohistiocytosis and macrophage activation syndrome in fatal cases of H1N1 influenza. Journal of Infectious Diseases. 2016 Apr 1;213(7):1180–8.

150. Stranneheim H, Lagerstedt-Robinson K, Magnusson M, Kvarnung M, Nilsson D, Lesko N, et al. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. Genome Med. 2021 Dec 1;13(1).

151. Splinter K, Adams DR, Bacino CA, Bellen HJ, Bernstein JA, Cheatle-Jarvela AM, et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. New England Journal of Medicine. 2018 Nov 29;379(22):2131–9.

152. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep 6;489(7414):57–74.

153. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005 Aug;15(8):1034–50.

154. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005 Jul;15(7):901–13.

155. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010 Jan;20(1):110–21.

156. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. In: Bioinformatics. 2009.

157. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507(7493):455–61.

158. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Vol. 45, Nature Genetics. 2013. p. 580–5.

159. Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (1979). 2020 Sep 11;369(6509):1318–30.

160. Exner M, Minar E, Wagner O, Schillinger M. The role of heme oxygenase-1 promoter polymorphisms in human disease. Free Radic Biol Med. 2004 Oct 15;37(8):1097–104.

161. Gimeno-Ferrer F, Albuquerque D, García Banacloy A, Guzmán Luján C, Vidal Garcia C, Marcaida Benito G, et al. Genetic screening for MC4R gene identifies three novel mutations associated with severe familiar obesity in a cohort of Spanish individuals. Gene. 2019 Jul 1;704:74–9.

162. Vaché C, Torriano S, Faugère V, Erkilic N, Baux D, Garcia-Garcia G, et al. Pathogenicity of novel atypical variants leading to choroideremia as determined by functional analyses. Hum Mutat. 2019 Jan 1;40(1):31–5.

163. Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003 Jul 15;12(14):1725–35.

164. Panne D. The enhanceosome. Curr Opin Struct Biol. 2008 Apr;18(2):236–42.

165. Banerji J, Rusconi S, Schaffner W. Expression of a Beta-Globin Gene Is Enhanced by Remote SV40 DNA Sequences. Cell. 1981;27:299–308.

166. Lin X, Liu Y, Liu S, Zhu X, Wu L, Zhu Y, et al. Nested epistasis enhancer networks for robust genome regulation. Science (1979). 2022 Sep 2;377(6610):1077–85.

167. Lin D, Hong P, Zhang S, Xu W, Jamal M, Yan K, et al. Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. Nat Genet. 2018 May 1;50(5):754–63.

168. Kim TH, Dekker J. Preparation of Cross-Linked Chromatin for ChIP. Cold Spring Harb Protoc. 2018 Apr 2;2018(4):pdb.prot082602.

169. Zinzen RP, Girardot C, Gagneur J, Braun M, Furlong EEM. Combinatorial binding predicts spatio-temporal cis-regulatory activity. Nature. 2009 Nov 5;462(7269):65–70.

170. Lee D, Karchin R, Beer MA. Discriminative prediction of mammalian enhancers from DNA sequence. Genome Res. 2011 Dec;21(12):2167–80.

171. Gorkin DU, Lee D, Reed X, Fletez-Brant C, Bessling SL, Loftus SK, et al. Integration of ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. Genome Res. 2012 Nov;22(11):2290–301.

172. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009 May 7;459(7243):108–12.

173. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007 Mar;39(3):311–8.

174. Pomerantz MM, Ahmadiyeh N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet. 2009 Aug;41(8):882–4.

175. Viart V, Georges M Des, Claustres M, Taulan M. Functional analysis of a promoter variant identified in the CFTR gene in cis of a frameshift mutation. European Journal of Human Genetics. 2012 Feb;20(2):180–4.

176. Sproat Emison E, Mccallion AS, Kashuk CS, Bush RT, Grice E, Lin S, et al. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. 2005; Available from: http://www.genome.ucsc.edu/cgi-bin/

177. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017 Jan 1;45(D1):D896–901.

178. Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat Genet. 2018 Aug 1;50(8):1140–50.

179. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol. 2015;2015:21.29.1-21.29.9.

180. Guilford P, Hopkins J, Harraway J, Mcleod M, Mcleod N, Harawira P, et al. E-cadherin germline mutations in familial gastric cancer. Nature. 1998;392(6674):402–5.

181. Barbaux S, Niaudet P, Gubler\ MC, Grtinfeld4 JP, Jaubert F, Kuttenn F, et al. Donor splice-site mutations in WT1 are responsible for Frasier syndrome. Nat Genet [Internet]. 1997;17(4):467–70. Available from: http://www.nature.com/naturegenetics

182. Igarashi M, Masunaga Y, Hasegawa Y, Kinjo K, Miyado M, Saitsu H, et al. Nonsense-associated altered splicing of MAP3K1 in two siblings with 46,XY disorders of sex development. Sci Rep. 2020 Dec 1;10(1).

183. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Tolun A, Chessa L, et al. Splicing Defects in the Ataxia-Telangiectasia Gene, ATM: Underlying Mutations and Consequences. Am J Hum Genet. 1999;64:1617–31.

184. López-Bigas N, Rabionet R, De Cid R, Govea N, Gasparini P, Zelante L, et al. Splice-site mutation in the PDS gene may result in intrafamilial variability for deafness in pendred syndrome. Hum Mutat. 1999;14(6):520–6.

185. Busslinger M, Moschonas N, Flavell RA. Beta Thalassemia: Aberrant Splicing Results from a Single Point Mutation in an Intron. Cell. 1981;27:289–98.

186. Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, et al. Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. Nature. 2003;423:293–8.

187. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019 Jan 24;176(3):535-548.e24.

188. Danis D, Jacobsen JOB, Carmody LC, Gargano MA, McMurry JA, Hegde A, et al. Interpretable prioritization of splice variants in diagnostic next-generation sequencing. Am J Hum Genet. 2021 Sep 2;108(9):1564–77.

189. Lee M, Roos P, Sharma N, Atalar M, Evans TA, Pellicore MJ, et al. Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. Am J Hum Genet. 2017 May 4;100(5):751–65.

190. Jagadeesh KA, Paggi JM, Ye JS, Stenson PD, Cooper DN, Bernstein JA, et al. S-CAP extends pathogenicity prediction to genetic variants that affect RNA splicing. Nat Genet. 2019 Apr 1;51(4):755–63.

191. Cheng J, Nguyen TYD, Cygan KJ, Çelik MH, Fairbrother WG, Avsec Ž, et al. MMSplice: Modular modeling improves the predictions of genetic variant effects on splicing. Genome Biol. 2019 Mar 1;20(1).

192. Rosenberg AB, Patwardhan RP, Shendure J, Seelig G. Learning the Sequence Determinants of Alternative Splicing from Millions of Random Sequences. Cell. 2015 Oct 22;163(3):698–711.

193. Xiong HY, Alipanahi B, Lee LJ, Bretschneider H, Merico D, Yuen RKC, et al. The human splicing code reveals new insights into the genetic determinants of disease. Science (1979). 2015 Jan 9;347(6218).

194. Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. Nucleic Acids Res. 2008 Apr;36(7):2257–67.

195. Chapman KB, Boeke JD. Isolation and Characterization of the Gene Encoding Yeast Debranching Enzyme. Vol. 65, Cell. 1991.

196. Ruskin B, Green MR. An RNA Processing Activity That-Debranches RNA Lariats. Science (1979) [Internet]. 1985;229(4709):135–40. Available from: https://www.science.org

197. Burset M, Seledtsov IA, Solovyev V V. SpliceDB: database of canonical and non-canonical mammalian splice sites [Internet]. Vol. 29, Nucleic Acids Research. 2001. Available from: http://www.softberry.com/spldb/SpliceDB.html.

198. Hall SL, Padgett RA. Conserved Sequences in a Class of Rare Eukaryotic Nuclear Introns with Non-consensus Splice Sites. J Mol Biol. 1994;239:357–65.

199. Hall SL, Padgettt RA. Requirement of U12 snRNA for in Vivo Splicing of a Minor Class of Eukaryotic Nuclear Pre-mRNA Introns. Science (1979) [Internet]. 1996;271(5256):1716–8. Available from: https://www.science.org

200. Dietrich RC, Incorvaia R, Padgett RA. Terminal Intron Dinucleotide Sequences Do Not Distinguish between U2- and U12-Dependent Introns. Mol Cell. 1997;1:151–60.

201. Talhouarne GJS, Gall JG. Lariat intronic RNAs in the cytoplasm of vertebrate cells. Proc Natl Acad Sci U S A. 2018 Aug 21;115(34):E7970–7.

202. Pineda JMB, Bradley RK. Most human introns are recognized via multiple and tissue-specific branchpoints. Genes Dev. 2018 Apr 1;32(7–8):577–91.

203. Taggart AJ, Desimone AM, Shih JS, Filloux ME, Fairbrother WG. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat Struct Mol Biol. 2012 Jul;19(7):719–21.

204. Taggart AJ, Lin CL, Shrestha B, Heintzelman C, Kim S, Fairbrother WG. Large-scale analysis of branchpoint usage across species and cell lines. Genome Res. 2017 Apr 1;27(4):639–49.

205. Mercer TR, Clark MB, Andersen SB, Brunck ME, Haerty W, Crawford J, et al. Genome-wide discovery of human splicing branchpoints. Genome Res. 2015 Feb 1;25(2):290–303.

206. Sibley CR, Emmett W, Blazquez L, Faro A, Haberman N, Briese M, et al. Recursive splicing in long vertebrate genes. Nature. 2015 May 21;521(7552):371–5.

207. Zhang Q, Fan X, Wang Y, Sun MA, Shao J, Guo D. BPP: A sequence-based algorithm for branch point prediction. Bioinformatics. 2017 Oct 15;33(20):3166–72.

208. Signal B, Gloss BS, Dinger ME, Mercer TR. Machine learning annotation of human branchpoints. Bioinformatics. 2018 Mar 15;34(6):920–7.

209. Paggi JM, Bejerano G. A sequence-based, deep learning model accurately predicts RNA splicing branchpoints. 2018; Available from: http://www.rnajournal.org/cgi/doi/10.1261/rna.

210. Nazari I, Tayara H, Chong KT. Branch Point Selection in RNA Splicing Using Deep Learning. IEEE Access. 2019;7:1800–7.

211. Taggart AJ, Fairbrother WG. ShapeShifter: a novel approach for identifying and quantifying stable lariat intronic species in RNAseq data. Quantitative Biology. 2018 Sep 1;6(3):267–74.

212. Bitton DA, Rallis C, Jeffares DC, Smith GC, Chen YYC, Codlin S, et al. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. Genome Res. 2014;24(7):1169–79.

213. Leman R, Tubeuf H, Raad S, Tournier I, Derambure C, Lanos R, et al. Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants. BMC Genomics. 2020 Jan 28;21(1).

214. Zhang SY, Clark NE, Freije CA, Pauwels E, Taggart AJ, Okada S, et al. Inborn Errors of RNA Lariat Metabolism in Humans with Brainstem Viral Infection. Cell. 2018 Feb 22;172(5):952-965.e18.

215. Zhang P, Philippot Q, Ren W, Lei WT, Li J, Stenson P, et al. Genome-wide detection of human variants that disrupt intronic branchpoints. Proceedings of the National Academy [Internet]. 2022;119(44). Available from: http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2211194119/-/DCSupplemental.https://doi.org/10.1073/pnas.22111941191of12

216. Kozyrev S V., Bernal-Quirós M, Alarcón-Riquelme ME, Castillejo-López C. The dual effect of the lupus-associated polymorphism rs10516487 on BANK1 gene expression and protein localization. Genes Immun. 2012 Feb;13(2):129–38.

217. Onochie CI, Korngut LM, Vanhorne JB, Myers SM, Michaud D, Mulligan LM. Characterisation of the human GFR-3 locus and investigation of the gene in Hirschsprung disease. J Med Genet [Internet]. 2000;37:674–9. Available from: www.jmedgenet.com

218. Blakes AJM, Wai HA, Davies I, Moledina HE, Ruiz A, Thomas T, et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. Genome Med. 2022 Dec 1;14(1).

219. Zahler AM, Neugebauer KM, Lane WS, Roth MB. Distinct Functions of SR Proteins in Alternative Pre-mRNA Splicing. Science (1979) [Internet]. 1993;260(5105):219–22. Available from: https://www.science.org

220. Krainer A, Conway GC, Kozak D. The Essential Pre-mRNA Splicing Factor SF2 Influences 5′ Splice Site Selection by Activating Proximal Sites. Cell. 1990;62:35–42.

221. Ge H, Manley JL. A Protein Factor, ASF, Controls Cell-Specific Alternative Splicing of SV40 Early Pre-mRNA In Vitro. Cell. 1990;62:25–34.

222. Fu XD, Maniatis T. The 35-kDa mammalian splicing factor SC35 mediates specific interactions between UI and U2 small nuclear ribonucleoprotein particles at the 3′ splice site. Proc Natl Acad Sci USA. 1992;89:1725–9.

223. Caceres JF, Stamm S, Helfman DM, Krainer AR. Regulation of Alternative Splicing in Vivo by Overexpression of Antagonistic Splicing Factors. Science (1979). 1994;265(5179):1706–9.

224. Blanchette M, Chabot B. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. Vol. 18, The EMBO Journal. 1999.

225. Martinez-Contreras R, Fisette JF, Nasim FUH, Madden R, Cordeau M, Chabot B. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol. 2006 Feb;4(2):172–85.

226. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive Identification of Exonic Splicing Enhancers in Human Genes. Science (1979). 2002;297(5583):1007–13.

227. Lefebvre S, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, et al. Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene. Cell. 1995;80:155–65.

228. Lorson CL, Strasswimmer J, Yao JM, Baleja3 JD, Hahnen4 E, Wirth B, et al. SMN oligomerization defect correlates with spinal muscular atrophy severity. Nat Genet [Internet]. 1998;19:63–6. Available from: http://www.nature.com/naturegenetics

229. Lorson CL, Hahnen E, Androphy EJ, Wirth B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci USA [Internet]. 1999;96:6307–11. Available from: www.pnas.org.

230. Wang L, Ji Y, Chen Y, Bai J, Gao P, Feng P. A splicing silencer in SMN2 intron 6 is critical in spinal muscular atrophy. Hum Mol Genet. 2023 Mar 6;32(6):971–83.

231. Aznarez I, Zielenski J, Rommens JM, Blencowe BJ, Tsui LC. Exon skipping through the creation of a putative exonic splicing silencer as a consequence of the cystic fibrosis mutation R553X. J Med Genet. 2007 May;44(5):341–6.

232. Wang J, Chang YF, Hamilton JI, Wilkinson MF. Nonsense-Associated Altered Splicing: A Frame-Dependent Response Distinct from Nonsense-Mediated Decay. Mol Cell. 2002;10:951–7.

233. Vuoristo MM, Pappas JG, Jansen V, Ala-Kokko L. A stop codon mutation in COL11A2 induces exon skipping and leads to non-ocular stickler syndrome. Am J Med Genet. 2004 Oct 1;130 A(2):160–4.

234. Laimer M, Önder K, Schlager P, Lanschuetzer CM, Emberger M, Selhofer S, et al. Nonsense-associated altered splicing of the Patched gene fails to suppress carcinogenesis in Gorlin syndrome. British Journal of Dermatology. 2008 Jul;159(1):222–7.

235. Littink KW, Pott JWR, Collin RWJ, Kroes HY, Verheij JBGM, Blokland EAW, et al. A Novel Nonsense Mutation in CEP290 Induces Exon Skipping and Leads to a Relatively Mild Retinal Phenotype. Retina [Internet]. 2010;51(7):3646–52. Available from: www.iovs.org/cgi/content/full/51/7/3646/DC1.

236. Carvill GL, Mefford HC. Poison exons in neurodevelopment and disease. Vol. 65, Current Opinion in Genetics and Development. Elsevier Ltd; 2020. p. 98–102.

237. Titus MB, Chang AW, Olesnicky EC. Exploring the Diverse Functional and Regulatory Consequences of Alternative Splicing in Development and Disease. Vol. 12, Frontiers in Genetics. Frontiers Media S.A.; 2021.

238. Pervouchine D, Popov Y, Berry A, Borsari B, Frankish A, Guigó R. Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. Nucleic Acids Res. 2019 Jun 4;47(10):5293–306.

239. Eom T, Zhang C, Wang H, Lay K, Fak J, Noebels JL, et al. NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. Elife. 2013 Jan 22;2013(2).

240. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell. 2016 Aug 25;166(5):1147-1162.e15.

241. Carvill GL, Engel KL, Ramamurthy A, Cochran JN, Roovers J, Stamberger H, et al. Aberrant Inclusion of a Poison Exon Causes Dravet Syndrome and Related SCN1A-Associated Genetic Epilepsies. Am J Hum Genet. 2018 Dec 6;103(6):1022–9.

242. Steward CA, Roovers J, Suner MM, Gonzalez JM, Uszczynska-Ratajczak B, Pervouchine D, et al. Re-annotation of 191 developmental and epileptic encephalopathy-associated genes unmasks de novo variants in SCN1A. NPJ Genom Med. 2019 Dec 1;4(1).

243. Felker SA, Lawlor JM, Hiatt SM, Thompson ML, Latner DR, Finnila CR, et al. Poison exon annotations improve the yield of clinically relevant variants in genomic diagnostic testing [Internet]. 2023. Available from: https://doi.org/10.1101/2023.01.12.523654

244. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012 May 17;485(7398):376–80.

245. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012 May 17;485(7398):381–5.

246. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (1979). 2009 Oct 9;326(5950):289–93.

247. Ong CT, Corces VG. CTCF: An architectural protein bridging genome topology and function. Vol. 15, Nature Reviews Genetics. Nature Publishing Group; 2014. p. 234–46.

248. Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. In: Nature Genetics. 2011. p. 630–8.

249. Sauerwald N, Singhal A, Kingsford C. Analysis of the structural variability of topologically associated domains as revealed by Hi-C. NAR Genom Bioinform. 2020 Mar 1;2(1).

250. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014 Dec 18;159(7):1665–80.

251. Gong Y, Lazaris C, Sakellaropoulos T, Lozano A, Kambadur P, Ntziachristos P, et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. Nat Commun. 2018 Dec 1;9(1).

252. Sauerwald N, Kingsford C. Quantifying the similarity of topological domains across normal and cancer human cell types. In: Bioinformatics. Oxford University Press; 2018. p. i475–83.

253. Yang T, Zhang F, Yardımci GG, Song F, Hardison RC, Noble WS, et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. 2017 Nov 1;27(11):1939–49.

254. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. 2018 Oct 4;19(1).

255. Singh R, Berger B. Deciphering the species-level structure of topologically associating domains [Internet]. 2021. Available from: https://doi.org/10.1101/2021.10.28.466333

256. Sefer E. A comparison of topologically associating domain callers over mammals at high resolution. BMC Bioinformatics. 2022 Dec 1;23(1).

257. D'haene E, Vergult S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. Genetics in Medicine [Internet]. 2021;23(1):34–46. Available from: https://doi.org/10.1038/s41436-

258. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015 May 30;161(5):1012–25.

259. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. Vol. 352, Science. American Association for the Advancement of Science; 2016. p. 1413–6.

260. Kozak M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell. 1986 Jan;44(2):283–92.

261. Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes (mRNA structure/alternative initiator codons/scanning model/in vitro translation). Proc Nail Acad Sci USA. 1990;87:8301–5.

262. Cazzola M, Skoda RC. Translational pathophysiology: a novel molecular mechanism of human disease. Blood. 2000;95(11):3280–8.

263. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Nail Acad Sci USA [Internet]. 2009;106(18):7507–12. Available from: www.pnas.org/cgi/content/full/

264. Rendleman J, Pasha Mohammad M, Pressler M, Maity S, Hronová V, Gao Z, et al. Regulatory start-stop elements in 5′ untranslated regions pervasively modulate translation [Internet]. 2021. Available from: https://doi.org/10.1101/2021.07.26.453809

265. Evans DGR, Bowers N, Burkitt-Wright E, Miles E, Garg S, Scott-Kitching V, et al. Comprehensive RNA Analysis of the NF1 Gene in Classically Affected NF1 Affected Individuals Meeting NIH Criteria has High Sensitivity and Mutation Negative Testing is Reassuring in Isolated Cases With Pigmentary Features Only. EBioMedicine. 2016 May 1;7:212–20.

266. Whiffin N, Karczewski KJ, Zhang X, Chothani S, Smith MJ, Evans DG, et al. Characterising the loss-of-function impact of 5′ untranslated region variants in 15,708 individuals. Nat Commun. 2020 Dec 1;11(1).

267. Ferreira de Lima RLL, Hoper SA, Ghassibe M, Cooper ME, Rorick NK, Kondo S, et al. Prevalence and nonrandom distribution of exonic mutations in interferon regulatory factor 6 in 307 families with Van der Woude syndrome and 37 families with popliteal pterygium syndrome. Genetics in Medicine. 2009;11(4):241–7.

268. Kondo S, Schutte BC, Richardson RJ, Bjork BC, Knight AS, Watanabe Y, et al. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. Nat Genet. 2002 Oct 1;32(2):285–9.

269. Wright CF, Quaife NM, Ramos-Hernández L, Danecek P, Ferla MP, Samocha KE, et al. Non-coding region variants upstream of MEF2C cause severe developmental disorder through three distinct loss-of-function mechanisms. Am J Hum Genet. 2021 Jun 3;108(6):1083–94.

270. Zhang X, Wakeling M, Ware J, Whiffin N. Annotating high-impact 5′untranslated region variants with the UTRannotator. Bioinformatics. 2021 Apr 15;37(8):1171–3.

271. Philips A V, Timchenko LT, Cooper TA. Disruption of Splicing Regulated by aCUG-Binding Protein in Myotonic Dystrophy [Internet]. Vol. 8, S. H. Leppla, Methods Enzymol. 1995. Available from: https://www.science.org

272. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20;72(2):245–56.

273. Van Blitterswijk M, Dejesus-Hernandez M, Rademakers R. How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: Can we learn from other noncoding repeat expansion disorders? Vol. 25, Current Opinion in Neurology. 2012. p. 689–700.

274. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron. 2011 Oct 20;72(2):257–68.

275. Sato N, Amino T, Kobayashi K, Asakawa S, Ishiguro T, Tsunemi T, et al. Spinocerebellar Ataxia Type 31 Is Associated with "Inserted" Penta-Nucleotide Repeats Containing (TGGAA)n. The American Journal of Human Genetics. 2009 Nov;85(5):544–57.

276. Andrew SE, Paul Goldberg Y, Kremer B, Telenius H, Theilmann J, Adam S, et al. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. Nat Genet. 1993 Aug;4(4):398–403.

277. Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. Science. 1992 Mar 6;255(5049):1253–5.

278. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, et al. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell. 1992 Feb 21;68(4):799–808.

279. Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology. 2001 Jul 10;57(1):127–30.

280. Dolzhenko E, Bennett MF, Richmond PA, Trost B, Chen S, van Vugt JJFA, et al. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. Genome Biol. 2020 Dec 28;21(1):102.

281. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. Vol. 21, Nature Reviews Genetics. Nature Research; 2020. p. 102–17.

282. Allou L, Balzano S, Magg A, Quinodoz M, Royer-Bertrand B, Schöpflin R, et al. Non-coding deletions identify Maenli lncRNA as a limb-specific En1 regulator. Nature. 2021 Apr 1;592(7852):93–8.

283. Carvill GL, Heavin SB, Yendle SC, McMahon JM, O'Roak BJ, Cook J, et al. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. Nat Genet. 2013 Jul;45(7):825–30.

284. Ganesh VS, Riquin K, Chatron N, Lamar KM, Aziz MC, Monin P, et al. Novel syndromic neurodevelopmental disorder caused by de novo deletion of CHASERR, a long

noncoding RNA Coalition to Cure CHD2, USA. 2024; Available from: https://doi.org/10.1101/2024.01.31.24301497

285. LaFlamme CW, Rastin C, Sengupta S, Pennington HE, Russ-Hall SJ, Schneider AL, et al. Diagnostic utility of DNA methylation analysis in genetically unsolved pediatric epilepsies and CHD2 episignature refinement. Nat Commun. 2024 Aug 6;15(1):6524.

286. Ferguson-Smith AC. Genomic imprinting: The emergence of an epigenetic paradigm. Vol. 12, Nature Reviews Genetics. 2011. p. 565–75.

287. Lucifero D, Mann MR, Bartolomei MS, Trasler JM. Gene-specific timing and epigenetic memory in oocyte imprinting. Hum Mol Genet. 2004 Apr 15;13(8):839–49.

288. Dechiara TM, Robertson EJ, Efstratiadis A. Parental Imprinting of the Mouse Insulin-like Growth Factor II Gene. Cell. 1991;64:849–59.

289. Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. Nature. 1991 May 9;351:153–5.

290. Plass C, Shibata H, Kalcheva Iveta, Mullins L, Kotelevtseva N, Mullins J, et al. Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. Nat Genet [Internet]. 1996 Sep 1;14:106–9. Available from: http://www.nature.com/naturegenetics

291. Barlow D, Stöger R, Herrmann B, Saito K, Schweifer N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature. 1991 Jan 3;349:84–7.

292. Gigante S, Gouil Q, Lucattini A, Keniry A, Beck T, Tinning M, et al. Using long-read sequencing to detect imprinted DNA methylation. Nucleic Acids Res. 2019 May 1;47(8).

293. Netchine I, Rossignol S, Dufourg MN, Azzi S, Rousseau A, Perin L, et al. 11p15 Imprinting Center Region 1 Loss of Methylation Is a Common and Specific Cause of Typical Russell-Silver Syndrome: Clinical Scoring System and Epigenetic-Phenotypic Correlations. J Clin Endocrinol Metab. 2007 Aug 1;92(8):3148–54.

294. Dutly F, Baumer A, Kayserili H, Yuksel-Apak M, Zerova T, Hebisch G, et al. Seven cases of Wiedemann-Beckwith syndrome, including the first reported case of mosaic paternal isodisomy along the whole chromosome 11. Am J Med Genet. 1998 Oct 12;79(5):347–53.

295. Magenis RE, Toth-Fejel S, Allen LJ, Black M, Brown MG, Budden S, et al. Comparison of the 15q deletions in Prader-Willi and Angelman syndromes: Specific regions, extent of deletions, parental origin, and clinical consequences. Am J Med Genet. 1990 Mar 5;35(3):333–49.

296. Wilkie AO, Malcolm S, Pembrey ME. Isodisomy in BWS chromosomes. Nature. 1991 Oct 31;353:802.

297. King DA, Fitzgerald TW, Miller R, Canham N, Clayton-Smith J, Johnson D, et al. A novel method for detecting uniparental disomy from trio genotypes identifies a significant excess in children with developmental disorders. Genome Res. 2014;24(4):673–87.

298. Klein CJ, Botuyan MV, Wu Y, Ward CJ, Nicholson GA, Hammans S, et al. Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. Nat Genet. 2011 Jun;43(6):595–600.

299. Sun Z, Wu Y, Ordog T, Baheti S, Nie J, Duan X, et al. Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. Epigenetics. 2014 Jul 7;9(8):1184–93.

300. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread Monoallelic Expression on Human Autosomes. Science (1979). 2007 Nov 16;318(5853):1136–40.

301. Zwemer LM, Zak A, Thompson BR, Kirby A, Daly MJ, Chess A, et al. Autosomal monoallelic expression in the mouse. Genome Biol. 2012 Feb 20;13(2).

302. Matuozzo D, Talouarn E, Marchal A, Zhang P, Manry J, Seeleuthner Y, et al. Rare predicted loss-of-function variants of type I IFN immunity genes are associated with life-threatening COVID-19. Genome Med. 2023 Dec 1;15(1).

303. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods. 2017 Oct 1;14(10):959–62.

304. Gustafsson C, Hauenstein J, Frengen N, Krstic A, Luc S, Månsson R. T-RHEX-RNAseq – a tagmentation-based, rRNA blocked, random hexamer primed RNAseq method for generating stranded RNAseq libraries directly from very low numbers of lysed cells. BMC Genomics. 2023 Dec 1;24(1).

305. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012 Aug;40(15).

306. Gudmundsson S, Singer-Berk M, Watts NA, Phu W, Goodrich JK, Solomonson M, et al. Variant interpretation using population databases: Lessons from gnomAD. Vol. 43, Human Mutation. John Wiley and Sons Inc; 2022. p. 1012–30.

307. Kousathanas A, Pairo-Castineira E, Rawlik K, Stuckey A, Odhams CA, Walker S, et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. Nature. 2022 Jul 7;607(7917):97–103.

308. Holmes TD, Pandey RV, Helm EY, Schlums H, Han H, Campbell TM, et al. The transcription factor Bcl11b promotes both canonical and adaptive NK cell differentiation. Sci Immunol. 2021 Mar 12;6(57).

309. GATK. https://gatk.broadinstitute.org/hc/en-us/articles/360035531112--How-to-Filter-variants-either-with-VQSR-or-by-hard-filtering. 2024.

310. Martin M, Patterson M, Garg S, Fischer SO, Pisanti N, Klau GW, et al. WhatsHap: fast and accurate read-based phasing. Available from: https://doi.org/10.1101/085050

311. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell. 2018 Nov 29;175(6):1701-1715.e16.

312. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010 May 28;38(4):576–89.

313. Ciancanelli MJ, Huang SXL, Luthra P, Garner H, Itan Y, Volpi S, et al. Life-threatening influenza and impaired interferon amplification in human IRF7 deficiency. Science (1979). 2015 Apr 24;348(6233):448–53.

314. Zhang Q, Liu Z, Moncada-Velez M, Chen J, Ogishi M, Bigio B, et al. Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. Science (1979). 2020 Oct 23;370(6515).

315. Gervais A, Marchal A, Fortova A, Berankova M, Krbkova L, Pychova M, et al. Autoantibodies neutralizing type I IFNs underlie severe tick-borne encephalitis in ~10% of patients. J Exp Med. 2024 Oct 7;221(10).

316. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. New England Journal of Medicine [Internet]. 2020 Oct 15;383(16):1522–34. Available from: http://www.nejm.org/doi/10.1056/NEJMoa2020283

317. Seo S, Zhang Q, Bugge K, Breslow DK, Searby CC, Nachury M V., et al. A novel protein LZTFL1 regulates ciliary trafficking of the BBSome and smoothened. PLoS Genet. 2011 Nov;7(11).

318. Wei Q, Chen ZH, Wang L, Zhang T, Duan L, Behrens C, et al. LZTFL1 suppresses lung tumorigenesis by maintaining differentiation of lung epithelial cells. Oncogene. 2016 May 19;35(20):2655–63.

319. Ragab D, Salah Eldin H, Taeimah M, Khattab R, Salem R. The COVID-19 Cytokine Storm; What We Know So Far. Front Immunol. 2020 Jun 16;11.

320. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020 Feb;395(10223).

321. Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. The Lancet. 2020 Mar;395(10229).

322. Tesi B, Bryceson YT. HLH: genomics illuminates pathophysiological diversity. Blood. 2018 Jul 5;132(1).

323. Cook SA, Comrie WA, Poli MC, Similuk M, Oler AJ, Faruqi AJ, et al. HEM1 deficiency disrupts mTORC2 and F-actin control in inherited immunodysregulatory disease. Science (1979). 2020 Jul 10;369(6500).

324. Henter JI, Ehrnst A, Andersson J, Elinder G. Familial hemophagocytic lymphohistiocytosis and viral infections. Acta Paediatr. 1993 Apr;82(4).

325. Miao Y, Zhu HY, Qiao C, Xia Y, Kong Y, Zou YX, et al. Pathogenic Gene Mutations or Variants Identified by Targeted Gene Sequencing in Adults With Hemophagocytic Lymphohistiocytosis. Front Immunol. 2019 Mar 7;10.

326. Meyer L, Nichols K. Redirect your attention: new CTL assay for HLH. Vol. 144, Blood. Elsevier B.V.; 2024. p. 802–4.

327. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008 Nov 27;456(7221):470–6.

328. Cummings BB, Karczewski KJ, Kosmicki JA, Seaby EG, Watts NA, Singer-Berk M, et al. Transcript expression-aware annotation improves rare variant interpretation. Nature. 2020 May 28;581(7809):452–8.

329. Melé M, Pedro G Ferreira, Ferran Reverter, David S DeLuca, Jean Monlong, Michael Sammeth, et al. The human transcriptome across tissues and individuals. Science (1979). 2015 May 8;348(6235):660–5.

330. Ardlie KG, DeLuca DS, Segrè A V., Sullivan TJ, Young TR, Gelfand ET, et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science (1979). 2015 May 8;348(6235):648–60.

331. Sadikovic B, Levy MA, Kerkhof J, Aref-Eshghi E, Schenkel L, Stuart A, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genetics in Medicine. 2021 Jun 1;23(6):1065–74.

332. Smail C, Montgomery SB. RNA Sequencing in Disease Diagnosis. Annual Review of Genomics and Human Genetics Downloaded from www.annualreviews.org Guest [Internet]. 2024;54:54. Available from: https://doi.org/10.1146/annurev-genom-021623-

333. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat Commun. 2017 Jun 12;8.

334. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med [Internet]. 2017 Apr 17;9(386):eaal5209. Available from: https://www.science.org

335. Wright CF, Campbell P, Eberhardt RY, Aitken S, Perrett D, Brent S, et al. Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. New England Journal of Medicine. 2023 Apr 27;388(17):1559–71.

336. Pucel J, Briere L, Reuter C, Gochyyev P, Undiagnosed Diseases Network. Exome and genome sequencing in a heterogeneous population of patients with rare disease: Identifying predictors of a diagnosis. Genetics in Medicine. 2024 Jun 24;26(6).

337. Ungar RA, Goddard PC, Jensen TD, Degalez F, Smith KS, Jin CA, et al. Impact of genome build on RNA-seq interpretation and diagnostics. Am J Hum Genet. 2024 Jul 11;111(7):1282–300.