MEDICAL UNIVERSITY OF VIENNA

Systems biology approaches for the phenotypic and network-based characterization of inborn errors of immunity

Doctoral thesis at the Medical University of Vienna
for obtaining the academic degree

Doctor of Philosophy

Submitted by

Julia Pazmandi, MSc

Supervisor:
Assoc. Prof. Priv.-Doz. Dr. Kaan Boztug
Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases
CeMM Research Center for Molecular Medicine, Austrian Academy of Sciences
St. Anna Children's Cancer Research Institute (CCRI)
St. Anna Children's Hospital, Department of Pediatrics and Adolescent Medicine, Medical University of Vienna
Department of Pediatrics and Adolescent Medicine, Medical University of Vienna

Co - Supervisor:
Prof. Dr. rer. Nat. Jörg Menche
Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna
Faculty of Mathematics, University of Vienna
CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences

Vienna, 06/2021

Declaration

The work presented in this thesis has been carried out in the Research Center for Molecular Medicine of the Austrian Academy of Sciences (CeMM), and at the Ludwig Boltzmann Institute of Rare and Undiagnosed Diseases (LBI-RUD), in Vienna, Austria. The thesis has been written in a cumulative format and includes a first author review, a shared first author publication, and a first author publication. The work that I present in this thesis was carried out in the laboratory of Kaan Boztug, at the Ludwig Boltzman institute of Rare and Undiagnosed Diseases, Vienna, Austria and was co-supervised by Jörg Menche at the Max Perutz Labs, Vienna, Austria.

#1 First author review published in Immunological Reviews:

Julia Pazmandi, Artem Kalinichenko, Rico Chandra Ardy, Kaan Boztug. 2019. "Early-Onset Inflammatory Bowel Disease as a Model Disease to Identify Key Regulators of Immune Homeostasis Mechanisms." Immunological Reviews 287 (1): 162-85.
DOI: 10.1111/imr.12726

#2 Shared first author publication published in the Journal of Allergy and Clinical Immunology:

Matthias Haimel PhD*, Julia Pazmandi MSc*, Raúl Jiménez Heredia MSc, Jasmin Dmytrus PhD, Sevgi Köstel Bal M.D.,PhD, Samaneh Zoghi PhD, Paul van Daele M.D., Tracy A. Briggs PhD, Carine Wouters M.D., Brigitte Bader-Meunier M.D., Florence A. Aeschlimann M.D., Roberta Caorsi M.D., Despina Eleftheriou M.D., Esther Hoppenreijs M.D., Elisabeth Salzer M.D.,PhD, Shahrzad Bakhtiar M.D., Beata Derfalvi M.D., Francesco Saettini M.D., Maaike A. A. Kusters M.D.,PhD , Reem Elfeky M.D., Johannes Trück M.D., Jacques G. Rivière M.D., Mirjam van der Burg PhD , Marco Gattorno M.D., Markus G. Seidel M.D., Siobhan Burns M.D., Klaus Warnatz M.D., Fabian Hauck M.D.,PhD , Paul Brogan M.D., Kimberly C. Gilmour PhD, Catharina Schuetz M.D., Anna Simon M.D.,PhD, Christoph Bock PhD , Sophie Hambleton PhD , Esther de Vries PhD , Peter Robinson M.D., Marielle van Gijn PhD †#, Kaan Boztug M.D.†#. 2021. "Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity". The Journal of Allergy and Clinical Immunology. Published:May 11,2021
DOI: https://doi.org/10.1016/j.jaci.2021.04.033
* and †, these authors contributed equally # to whom correspondence should be addressed.

#3 First author paper under review in Nature Communications since 20th May 2021:

Julia Pazmandi, Sevgi Köstel Bal, Felix Müller, Celine Sin, Christiane V. R. Hütter, Jörg Menche*#, Kaan Boztug*#. "AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation". * these authors contributed equally, # to whom correspondence should be addressed.

The project "Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity" was conceived by Kaan Boztug, including coordination of an international

consortium comprised of clinical immunologists, biologists, and bioinformaticians/computational biologists. The computational part of the project was designed by Matthias Haimel and Julia Pazmandi.

The project “AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation” was conceived by Kaan Boztug and Jörg Menche. The immunology part of the project was designed by Kaan Boztug, the computational part of the project was designed by Julia Pazmandi and Jörg Menche. Julia Pazmandi carried out the computation part of the work with the guidance of Jörg Menche.

The manuscripts resulting from this project are my first-author publications. They constitute this thesis and have been reprinted according to the Reprints & Permissions policies of Elsevier, Nature Publishing Group and John Wiley & Sons.

I, the author, wrote all the chapters of this thesis.

Table of Contents

List of figures

Abbreviations

AGS	Aicardi Gutiérrez Syndrome
APECED	Autoimmune polyendocrinopathy, candidiasis, and ectodermal dystrophy
BCR	B-cell receptor
CAPS	Cryopyrin-associated periodic syndromes
CEMM	Research Centre for Molecular Medicine of the Austrian Academy of Sciences
CGD	Chronic granulomatous disease
CID	Combined immunodeficiency
CTL	Cytotoxic T cell
CVID	Common variable immunodeficiency
DC	Dendritic cell
DO	Disease Ontology
EHR	Electronic health records
ESID	European Societies of Immunodeficiency
FA	Fanconi anemia
FMF	Familial Mediterranean Fever
GARD	Genetic and Rare Diseases Information center
GM-CSF	Granulocyte-macrophage colony stimulating factor
GOF	Gain of function
GWAS	Genome Wide Association Studies
HIPPIE	Human Integrated Protein Protein Interaction rEference
HOOM	HPO and ORDO Ontological Module
HPO	Human Phenotype ontology
IBD	Inflammatory bowel disease
IC	Information content
ICD	International Statistical Classification of Diseases and Related Health Problems
IEI	Inborn errors of immunity
IFN	Interferon
IL	Interleukin

IPEX	Immune dysregulation, polyendocrinopathy, enteropathy, X-linked
IUIS	International Union of Immunological Societies
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIR	Killer cell immunoglobulin receptors
LBI-RUD	Ludwig Boltzmann Institute of Rare and Undiagnosed Diseases
LOF	Loss of function
MAF	Minor allele frequency
MHC	Major histocompatibility complex
MPO	Mammalian Phenotype Ontology
NCR	Neural concept recognizer
NGS	Next generation sequencing
NK	Natural killer cell
NLP	Natural language processing
OMIM	Online Mendelian Inheritance of Man
OWL	Web Ontology Language
PAMP	Pathogen-associated molecular pattern
PID	Primary immunodeficiency
PRR	Pattern recognition receptor
RaDaR	Rare Diseases Registry Program
ROS	Reactive oxygen species
SCID	Severe combined immunodeficiency
SHM	Somatic hypermutation
SLE	Systemic lupus erythematosus
SNOMED CT	Systematized Nomenclature Of Medicine clinical terms
STAT	Signal transducer activator of transcription
STRING	Search Tool for Recurring Instances of Neighboring Genes
TCR	T cell receptor
Tfh	Follicular helper T cell
TNF-a	Tumor necrosis factor alpha
TRAPS	Tumor necrosis factor receptor-associated periodic fever syndrome
Treg	Regulatory T cell

UDP and UDN	NIH Undiagnosed Disease Program and Network
UMLS	Unified Medical Language System
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
Y2H	Yeast two-hybrid

Acknowledgements

I would first like to express my gratitude to CeMM, and the CeMM PhD program. I have spent the majority of my PhD between the walls of an exceptional research institute. The people, the atmosphere, the scientific rigor and curiosity, and the work ethic that I have experienced at CeMM have made me the scientist that I am today. I will forever cherish the memories that I have made during these years. Special thanks for all the administrative and personal help from the administration team and other teams. And of course I am grateful for the people in my year for the companionship, the sharing of our struggles and successes: Alex, Fede, Martin, Lydia, Rob, and Rouchen - the journey would not have been the same without you.

Of course, the work presented in this thesis would not have been possible without my supervisor, Kaan Boztug. I would like to thank Kaan for giving me a chance and offering me a position in his lab in 2015. The opportunity provided through this PhD experience was truly life-changing. The work with Kaan has taught me what determination and scientific excellence look like, and has inspired me to continue to pursue it in my future. Thank you for everything!

As stated in the cover page of the thesis, my work and PhD has been co-supervised by Jörg Menche. I would like to thank Jörg for adopting me into his team, and making me feel a true part of the “homies”. You have shown me how beautiful and exciting science can be, and have encouraged me to get over my fears and feelings of inadequacy (which is difficult when one is working among so many inspiring and brilliant people!). I am truly privileged to have worked with you and your team.

I am also incredibly thankful for my colleagues in the Boztug group. This tight-knit community made especially the first, most timid years of my PhD feel like I was working among family. Thank you for the many discussions, debates, encouragement, laughs and shared difficult moments. You have taught me what hard work, resilience and creativity look like. Furthermore, I must say thank you to my adoptive group, Menche group. I am one of the lucky few people who got to call not just one, but two groups as home during their PhD. I will always cherish and remember all the discussions (fueled by coffee, beer or both) among this incredibly diverse group of people. I feel very fortunate to have met and got to know you all, to have been on adventures that were both scientific and not-so scientific (but interesting nonetheless!).

I would like to express my gratitude to my parents. Without their support, I would not have been able to do it all. Thank you for providing a great example of the value of hard work and dedication, and the importance of education. Thank you for supporting me all the way from the start of my academic journey at 6 years old, and further down the line, making it possible for me to move abroad to pursue science first in Sweden, then in England, and now in Austria. I would also like to thank my siblings, Mendi and Jenci. Thank you for being such great siblings, and such (annoyingly) high-achieving people. Jenci, although I don't tell you, you always inspire me with your creativity and endless business and now scientific ideas. We have even survived living together as adults, which is a great accomplishment in my book. Thank you Mendi for always being my friend, my cheerleader and an example of a good scientist. I will try to forgive you for finishing your PhD before I finished mine. I would also like to thank all of my friends - the Hungarian ladies who have been by my side for over ten years, and all the international friends that I have made along the way. You have made these last 5,5 years so much fun!

Towards the end, I would like to thank my fiancé Marcus. Thank you for making these past years the best years! You have listened to my stresses and complaints, and have made me feel supported and cared for. Thank you very much for that. Finally, I would like to thank Otis. You have not been born yet, but you have already provided me with the greatest motivation to complete my PhD and write this thesis.

Abstract

Inborn errors of immunity (IEI) are a heterogeneous group of rare diseases that affect the immune system. All IEI in total affect a considerable fraction of the population and pose a significant demand on the healthcare system. Although individual components of these diseases have been studied in detail, the available clinical, phenotypic and molecular data is sparse, and large-scale comparative studies that reveal general properties of these diseases are still lacking. This has resulted in a considerable diagnostic delay for IEI patients, as well as an incomplete understanding of core genes and pathways that lead to the observed disease pathobiology. In order to bridge the diagnostic gap of IEI, and to get a better understanding of the core genes and pathways of the immune system, accurate disease-associated knowledge bases and tailored approaches are required.

To this end, this thesis introduces three manuscripts, including a review of a specific rare disease to illustrate the genetic and molecular diversity of the condition, and two research manuscripts for i) creating accurate phenotyping data for rare diseases in an expert-driven, machine-learning aided approach, and ii) applying network medicine to elucidate the genetic and molecular heterogeneity behind Mendelian autoimmune and autoinflammatory diseases.

The manuscript in the introduction focuses on early-onset monogenic IBD, and details the genetic lesions and pathobiological changes that lead to the bowel inflammation phenotype. The subsequent research manuscripts presented in the results section focus on specific applications of systems-methods for IEI. Research manuscript one details our effort to revise and expand the available phenotyping data to accurately describe IEI. We have initiated an international and interdisciplinary collaboration of IEI experts. Within this collaboration, we have developed a framework for the revision and expansion of the phenotypic representation of IEI using the Human Phenotype Ontology (HPO). Four major branches of the HPO tree were revised, focusing on four separate subgroups of IEI as a proof of concept. As a result of this revision, over 206 changes in the ontology structure were requested including 137 new, IEI-relevant phenotypic terms. We have developed an expert-reviewed ontology-guided machine learning method to reannotate IEI with HPO terms. With this method, we achieved a significant, 4.7-fold increase in the available phenotypic terms per disease which has translated into a significant improvement in HPO-based diagnostic accuracy and disease-similarity. Our directed expansion of the HPO corpus has

enhanced the precision of phenotypic annotation of IEI, which will enable the characterization of these diseases in the community, and benefit both IEI diagnostics and research.

The overarching ambition of the second research manuscript was to develop a systems-level view for IEI with autoimmunity and autoinflammation, and showcase its utility for addressing a wide range of important biomedical questions. We started by building a state-of-the-art interactome combining several relevant interactome resources from the literature. Using rare IEI that present with autoimmunity/autoinflammation, we identified the AutoCore, a tightly connected subnetwork as the set of core genes and their interaction essential for the homeostasis of immune function. We showed that within the AutoCore, autoimmunity and autoinflammation do not separate but are molecularly linked, and that the monogenic AutoCore is at the topological center of complex, polygenic autoimmune and autoinflammatory diseases. Furthermore, we used the topology of the AutoCore to identify 19 distinct molecular subclusters of monogenic autoimmune and autoinflammatory diseases. Finally, we used the AutoCore and network distance to pinpoint potential novel therapeutically targetable pathways.

Overall, this thesis illustrates how to use tailored systems-based methods to both expand the available phenotypic knowledge-base of IEI, and to develop a network-based framework for unraveling the molecular landscape of autoimmunity and autoinflammation to quantify previously only descriptive clinical phenomena. Both of these applications of different systems-methods significantly increase the current available knowledge on rare diseases of the immune system, contribute to bridging the diagnostic gap, and offer lasting novel platforms to systematically describe and explore the molecular and phenotypic origins of immune homeostasis and dysregulation.

Zusammenfassung

Angeborene Fehler des Immunsystems (English: *Inborn errors of immunity* (IEI)) sind eine heterogene Gruppe seltener Erkrankungen des Immunsystems. In ihrer Gesamtheit betreffen IEI einen erheblichen Teil der Bevölkerung und stellen eine beträchtliche Belastung des Gesundheitssystems dar. Obwohl einzelne Komponenten dieser Erkrankungen detailliert untersucht wurden, sind vergleichsweise wenig klinische, phänotypische und molekulare Daten verfügbar. Große Vergleichsstudien, die allgemeingültige Eigenschaften dieser Erkrankungen aufzeigen, liegen nicht vor. Dies führt einerseits zu einer erheblichen Verzögerung der Diagnose von IEI-Patienten und äußert sich zudem in einem unvollständigen Verständnis von Schlüsselgenen und Signalwegen der jeweiligen Erkrankung. Um diese diagnostische Lücke bei IEI zu überbrücken und die zentralen Gene und Signalwege des Immunsystems besser zu verstehen, sind präzisere Datenbanken und maßgeschneiderte wissenschaftliche Ansätze erforderlich.

Dazu werden in dieser Arbeit drei Manuskripte vorgestellt. In einem Review zu einer spezifischen seltenen Erkrankung, wird die genetische und molekulare Vielfalt der Erkrankung veranschaulicht. Zudem behandeln die zwei Forschungsmanuskripte zwei neuartige Methoden um i) in einem von Experten gesteuerten und durch maschinelles Lernen unterstützten Ansatzes genaue phänotypische Daten für seltene Erkrankungen zu kategorisieren und ii) mittels systembasierter Netzwerkansätze in der Medizin zur Aufklärung der genetischen und molekularen Heterogenität hinter Mendelschen Autoimmunerkrankungen und autoinflammatorischen Erkrankungen beizutragen.

Die Forschungsmanuskripte im Ergebnisteil konzentrieren sich auf spezifische Anwendungen von Systemansätzen für IEI. Das erste Forschungsmanuskript beschreibt detailliert unsere Bemühungen, die verfügbaren Daten zur Phänotypisierung zu überarbeiten und zu erweitern, um IEI genauer beschreiben zu können. Zur Umsetzung haben wir eine internationale und interdisziplinäre Zusammenarbeit von IEI-Experten initiiert. Im Rahmen dieser Zusammenarbeit haben wir ein Konzept für die Überarbeitung und Erweiterung der phänotypischen Darstellung von IEI unter Verwendung der *Human Phenotype Ontology* (HPO) entwickelt. Vier Hauptzweige des HPO-Baumes wurden überarbeitet, wobei vier separate Untergruppen der IEI als Machbarkeitsnachweis im Mittelpunkt standen. Infolge dieser Überarbeitung wurden über 206 Änderungen in der Ontologiestruktur angefordert, darunter 137 neue, IEI-relevante

phänotypische Begriffe. Zusammengefasst, haben wir eine von Experten überprüfte Ontologie-gesteuerte Methode für maschinelles Lernen entwickelt, um IEI mit HPO-Begriffen neu zu annotieren. Mit dieser Methode haben wir eine signifikante, 4,7-fache Erhöhung der verfügbaren phänotypischen Begriffe pro Krankheit erreicht, was zu einer signifikanten Verbesserung der HPO-basierten diagnostischen Genauigkeit und Krankheitsähnlichkeit geführt hat. Unsere gezielte Erweiterung der HPO Begriffe hat die Präzision der phänotypischen Annotation von IEI verbessert. Dies ermöglichte eine präzise Charakterisierung dieser Krankheiten in der Bevölkerung und kommt sowohl der IEI-Diagnostik als auch der Forschung zugute.

Das übergeordnete Ziel des zweiten Forschungsmanuskripts bestand darin, eine systembasierte Analyse von IEI, die sich durch Autoimmunität oder Autoinflammation manifestieren, zu entwickeln und ihren Nutzen für die Beantwortung einer Vielzahl wichtiger biomedizinischer Fragen zu demonstrieren.

Wir begannen mit der Konstruktion eines neuartigen Interaktom-Netzwerks, das relevante und anerkannte Interaktome kombiniert. Innerhalb des kombinierten Interaktoms mit seltenen IEI, bei denen Autoimmunität/Autoinflammation auftritt, identifizierten wir eine Unterguppe von Schlüsselgenen und deren Interaktionspartner, die für die Aufrechterhaltung der Homoöstase unerlässlich sind, und daher als „AutoCore“ bezeichnet werden. Wir zeigten hier, dass innerhalb des AutoCores, Autoimmunität und Autoinflammation eng miteinander verknüpft sind, und dass sich das monogene AutoCore im Zentrum des Netzwerkes komplexer polygener Autoimmun- und Autoinflammationserkrankungen befindet. Darüber hinaus verwenden wir die Topologie des Autocores, um 19 unterschiedliche molekulare Untergruppen monogener Autoimmun- und Autoinflammationserkrankungen zu identifizieren. Schließlich verwenden wir die Autocore- und Netzwerkentfernung, um neue potenzielle Signalwege für therapeutische Interventionen zu ermitteln.

Insgesamt zeigt diese Arbeit, wie man maßgeschneiderte systembasierte Methoden verwendet, um die Charakterisierung klinischer Phänotypen von IEI zu erweitern und einen netzwerkbasierten Rahmen für die Entdeckung molekularer Zusammenhänge von Autoimmun- und Autoinflammationserkrankungen schafft. Dieser Systemansatz ermöglicht daher die Quantifizierung von zuvor nur beschriebenen klinischen Phänomenen. Beide Methoden erhöhen das aktuell verfügbare Wissen über seltene Erkrankungen des Immunsystems erheblich und tragen dazu bei die diagnostische Lücke zu schließen. Sie etablieren zudem neuartige

Plattformen, um systematisch die molekularen und phänotypischen Ursprünge von Homoöstase und Dysregulierung des Immunsystems zu beschreiben und zu erkunden.

Overview

This thesis has been written in a cumulative form. It includes a review article on inflammatory bowel diseases “Early-Onset Inflammatory Bowel Disease as a Model Disease to Identify Key Regulators of Immune Homeostasis Mechanisms.” in the introduction, and two research articles, “Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity” and “AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation”, that are the main results of this thesis and presented in chapter 3.

The first part of the thesis contains **the introduction (1)** starting from a chapter on rare diseases of the immune system (1.1), including a general introduction (1.1.1) and a more detailed discussion on rare diseases of the immune system (1.1.2). This is followed by an introduction to the databases, registries, nomenclatures and data structures available for rare diseases of the immune system (1.1.3 and 1.1.4), the human phenotype ontology (1.1.5), and the assessment of ontological similarity (1.1.6). The next section introduces the different elements of the immune system, focusing on innate (1.2.1) and adaptive immunity (1.2.2) and autoimmune and autoinflammatory diseases (1.2.3) in particular. Then, gene and molecular defects leading to early-onset inflammatory bowel disease, a prototypic autoimmune/autoinflammatory condition are introduced in detail (1.2.4), with a review. In the third part of the introduction, network medicine (1.3) and network-based methods are discussed in detail, covering topics such as types of biological networks (1.3.1.1), interactomes as maps of cellular function (1.3.2) and basic network properties (1.3.2.1-1.3.2.4). Special emphasis is given to how network medicine is used for different types of diseases (1.3.3), including disease gene identification (1.3.3.1), methods for the investigation of disease-disease relationships (1.3.3.2) and applications to assess drug efficacy (1.3.3.3).

The second chapter introduces the detailed **aims of the thesis (2)**, by formulating the research question and the objectives of this work.

The main **results (3)** of this doctoral thesis are detailed in chapter three, with two research articles in chapter (3.1) and (3.2), respectively. Each article is preceded by a short summary of the work. The research articles are followed by their respective supplementary material which provide details on the applied methodology.

Finally, the fourth chapter of the doctoral thesis contains a detailed **discussion (4)** and future outlook of the results presented in the two research articles. The discussion begins with a general introduction and discussion (4), followed by a separate discussion and outlook on each research article in sections (4.1) and (4.2).

1 Introduction

1.1 Rare diseases of the immune system

1.1.1 Introduction to rare diseases

It is estimated that there are between 6000-8000 rare diseases (McKusick 2007; Pavan et al. 2017), that taken together affect about 3.5 – 5.9% of the global population, about 263-446 million persons (Wakap et al. 2019). There is no official, unifying definition of rare diseases to date. In Europe, a disease is considered rare if it affects less than 1 in 2000 people (Nugent and Rhinard 2015; Eurodis), while in the United States a disease is defined as rare if it affects fewer than 1 in 1500 people (Reaves 2003). Rare diseases are often chronic, progressive, degenerative and life threatening diseases that represent a challenge on healthcare systems. Many of these diseases lack effective treatments today (Braun et al. 2010). About 80% of rare diseases are considered to be of genetic origin, many of them following the inheritance pattern proposed by Georg Mendel in 1865, therefore termed Mendelian diseases (Wakap et al. 2019; MENDEL G 1865). In contrast to common diseases that are usually brought on by a combination of factors including environmental factors such as lifestyle, exposures to environmental agents and genetic susceptibility, rare diseases tend to be the results of single detrimental and deleterious lesions in the genome. These lesions often affect the functions of proteins, and the function of cellular processes severely enough to elicit a strong phenotype (Figure 1). These function-affecting, deleterious genetic lesions are rare, and can be found in less than 1% of the population (minor allele frequency, MAF < 1 %), as opposed to the common polymorphisms in the genome (MAF >= 1%) (McCarthy et al. 2008). The genetic lesions that cause rare disease can present as biallelic mutations, whereby both alleles of an affected gene carry a variation from the reference genome, or they can be heterozygous with only one affected allele. Because of their genetic origin, rare diseases tend to be severe and present at an early age, illustrated by the fact that the majority of affected are children (Wakap et al. 2019).

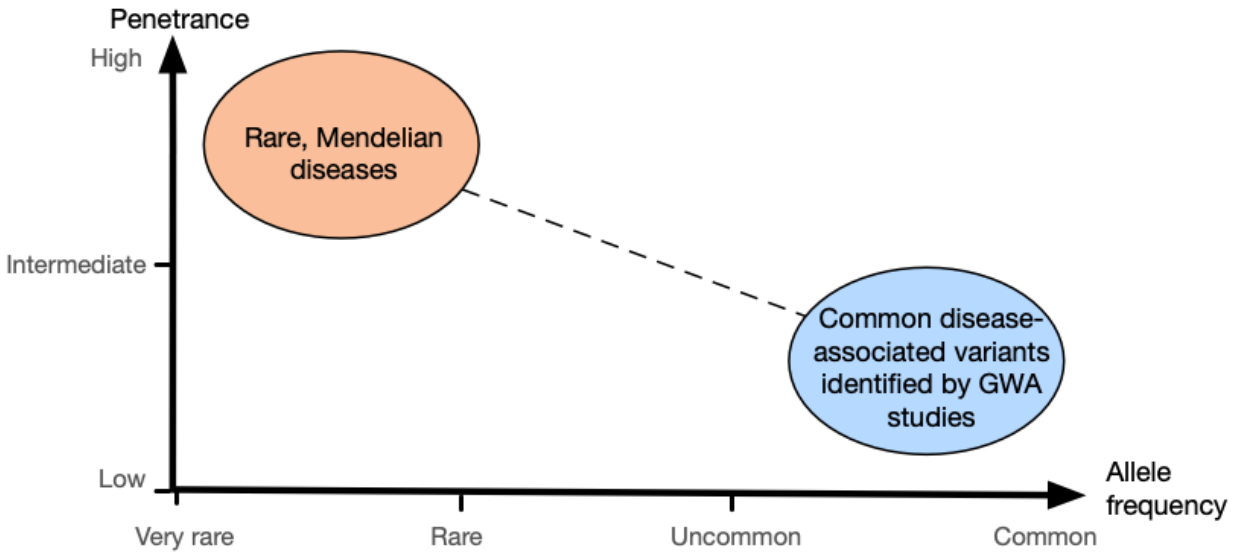


Figure 1. Genetic architecture of common, polygenic and rare, monogenic diseases. Reprinted by permission and adapted from Springer Nature: Nature, “Finding the missing heritability of complex diseases” (McCarthy et al. 2008).

The strong genetic component displayed by rare diseases has prompted research elucidating the genetic lesions underlying the phenotypes. Since the rapid expansion of next-generation sequencing techniques (NGS), NGS has become one of the gold standard tools of investigation and diagnosis of rare diseases (Vinkšel et al. 2021; Z. Liu et al. 2019). Although approaches to identify the pathogenic variants may vary, they often consist of sequencing a panel of genes previously identified to be causal for a specific phenotype, then moving onto whole exome (WES) or whole genome sequencing (WGS) if the causal genetic link remains elusive (Z. Liu et al. 2019). The NGS result of a single patient often contains hundreds of variants of unknown significance, which have to be validated and investigated. The subsequent prioritization and contextualization of these variants and genes is time and cost intensive. Therefore, although NGS-based diagnostic tools have accelerated the rate of disease gene discovery and diagnosis, they still only yield a clear genetic diagnosis for about 30% of rare disease patients, and the majority of rare disease patients do not receive a genetic diagnosis, despite being referred to expert centers (Gahl et al. 2012; Ramoni et al. 2017).

1.1.2 Rare diseases of the immune system

There are more than 400 rare diseases that affect the immune system, linked to over 430 different gene defects (A. Bousfiha et al. 2020; A. A. Bousfiha et al. 2013). These diseases, collectively termed IEI, have great social and economic impact, as they affect the very young and often present with a serious clinical course (Cannizzo et al. 2018). Due to their monogenic nature, the genotype to phenotype relationship displayed by IEI provides vital insights into immune homeostasis and the lack thereof. IEI usually manifest as increased susceptibility to infections, autoinflammation and/or autoimmunity, allergy and malignancy. As most rare diseases, IEI are caused by deleterious genetic lesions that often result in a loss-of-function (LOF) or gain-of-function (GOF) of the encoded protein (A. Bousfiha et al. 2020). Although most mutations underlying IEI are LOF biallelic mutations that impair the function of the encoded proteins, there is also evidence of heterozygous lesions that are inherited in an autosomal dominant manner. These heterogeneous mutations can be GOF, or deleterious lesions causing haploinsufficiency or a dominant negative effect.

The classification of IEI is a difficult task, as no two patients are the same and even in the same disease category there is ample heterogeneity (Vinkšel et al. 2021). To date, the International Union of Immunological Societies (IUIS) disease classification (A. Bousfiha et al. 2020), updated every two years, is one of the de facto ways to classify IEI disease in a phenotypic manner. This classification divides diseases based on an expert consensus on a phenotypic basis, and diseases are often re-classified. The IUIS clinical and phenotypical classification of IEI currently groups them into 10 categories (Figure 2).

Immunodeficiencies affecting cellular and humoral immunity (Table I) present with severe combined immunodeficiency (SCID) or combined immunodeficiency (CID). Combined immunodeficiencies with associated or syndromic features (Table II) include CIDs that present with thrombocytopenia and hyper IgE syndrome. Predominantly antibody deficiencies (Table III) contain B cell and immunoglobulin defects that usually present as common variable immunodeficiencies (CVIDs). Diseases of immune dysregulation (Table IV) represent most autoimmune diseases, as well as autoinflammatory diseases with susceptibilities to infection. Congenital defects of phagocyte number or function (Table V) group together diseases of phagocytic cells. Defects in intrinsic and innate immunity (Table VI) include diseases that present with susceptibility to bacterial, fungal and viral infection. Autoinflammatory disorders (Table VII)

group together classical autoinflammatory disorders. Complement deficiencies (Table VIII) contain defects of the complement system. The bone marrow failure (Table IX) group lists diseases such as those that cause Fanconi anemia, dyskeratosis congenita and others. Finally, phenocopies of IEI (Table X), a group of disorders presenting with similar phenotypes to IEI, but instead of germline mutations are results of somatic mutations in certain subsets of cells.

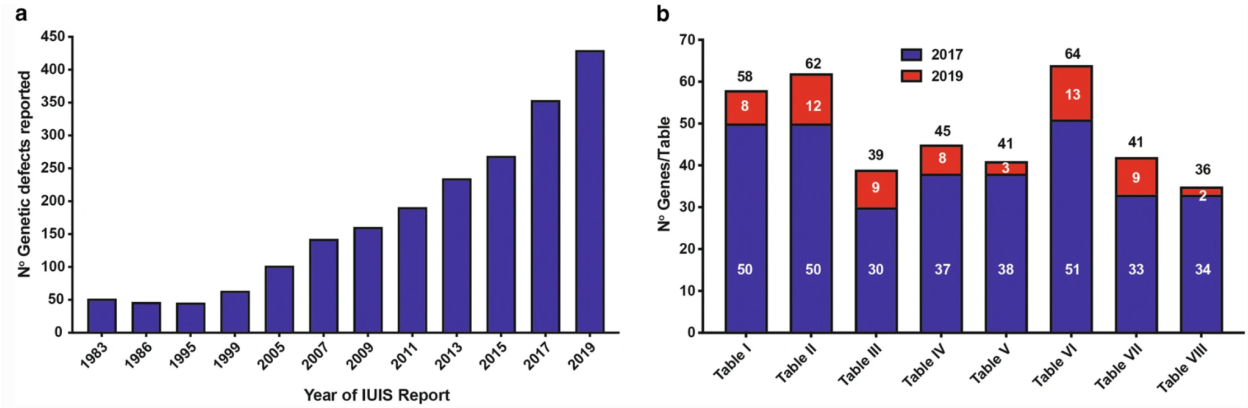


Figure 2. Rate of discovery, and types of IEI according to the International Union of Immunological Societies (IUIS). a) Number of IEI from 1983-2019. b) Number of diseases in each IUIS clinical group in 2017 and 2019. Reprinted by permission and adapted from Springer Nature: Journal of Clinical Immunology, "Human Inborn Errors of Immunity: 2019 Update of the IUIS Phenotypical Classification" (A. Bousfiha et al. 2020).

1.1.3 Databases and registries for inborn errors of immunity

In order to address the need to improve the level of diagnosis and care of rare diseases across the globe, multiple programmes and databases have been established in the past years. These include The NIH Undiagnosed Disease Program and Network (UDP and UDN) (Gahl et al. 2016), the Undiagnosed Diseases Network International (Taruscio et al. 2020), the RD-connect project to create a global infrastructure for rare diseases in the EU and beyond (Gainotti et al. 2018) and SOLVE-RD (Ferlini 2017), a research project dedicated to provide diagnosis for unsolved rare disease patients.

Finding a second patient with similar phenotype, or phenotypic matching to a cohort of rare disease patients is key to accelerate the diagnostic process for IEI. To facilitate this, registries and data sharing platforms have emerged in recent years. Registries such as the European

Societies of Immunodeficiency (ESID) registry (Grimbacher and ESID Registry Working Party 2014), or US-based registries such as the Rare Diseases Registry Program (RaDaR) developed by the NIH ("Rare Diseases Registry Program (RaDaR)" 2017), or the IAMRARE registry organised by the National Organization for Rare Disorders (NORD) ("IAMRARE® Registry Program" 2017) aim to provide a more centralized way of storing and accessing patient-related information. Data sharing platforms further facilitate the identification of patients with similar genetic makeup and phenotypes. These include GeneMatcher (N. Sobreira et al. 2015), a platform that matches researchers together based on interests in the same genotypes, and Matchmaker Exchange (N. L. M. Sobreira et al. 2017), a platform developed to pair up centers and research groups with patients with similar phenotypes.

1.1.4 Nomenclatures and data structures for inborn errors of immunity

As the available data on IEI patients increased, so did the need for objective disease classification systems and data nomenclatures. Alongside the clinical IUIS classification of diseases that is most often referred to by clinical experts, Online Mendelian Inheritance in Man (OMIM) (Amberger et al. 2018) emerged as a database on Mendelian (and non-Mendelian) diseases, and is also widely used in the IEI community. In parallel, nomenclatures have emerged that order disease classification into ontologies that provide a hierarchical structure to organize diseases. Disease ontologies such as ORDO (Vasant et al. 2014), developed by OrphaNet, as well as Disease Ontology (DO) (Schriml et al. 2012) aim to provide an ontological structure to classify and group rare diseases. Both of these ontologies have developed their own identifier systems to refer to diseases. Unfortunately, the representation of IEI in both ORDO and DO is sparse.

The ideal operation of registries and databases that aim to centralize patients relies on common standardized terminologies to refer to the highly multidimensional data that represents phenotypes and clinical identifiers. To meet this demand, nomenclatures that refer to both clinical and phenotypical data have emerged in recent years (Gkoutos, Schofield, and Hoehndorf 2018). Systematized Nomenclature Of Medicine Clinical Terms (SNOMED CT) (Bhattacharyya 2016) is a systematically organized collection of medical terms developed for clinical documentation and reporting. To date, SNOMED CT provides the core general terminology for electronic health records (EHR). SNOMED CT includes clinical findings, symptoms, diagnoses, procedures, body structures, organisms, etiologies, substances, pharmaceuticals, devices and specimens. Along

with SNOMED CT, Unified Medical Language System (UMLS) aims to provide a standardized biomedical terminology (Bodenreider 2004). The International Statistical Classification of Diseases and Related Health Problems (ICD) defines the universe of diseases, disorders, injuries and other related health conditions developed by the World Health Organization (WHO) (Kreutzer, DeLuca, and Caplan 2011). Although to some extent all these terminologies are in use for IEI especially in clinical care and EHR, many of the specific needs to represent phenotypes and their relationships in IEI are not depicted.

Along with the aforementioned terminologies, in recent years ontologies were developed specifically to represent disease phenotypes and the logical relationships between them. Ontologies are data structures to describe taxonomies and classification networks, simultaneously providing a vocabulary and defining the structure of knowledge among the data. The Mammalian Phenotype Ontology (MPO) (Smith, Goldsmith, and Eppig 2005), was the first entity of its kind that aimed to capture phenotypic aberrations to the entire phenome of an organism. Similar to the MPO, The Human Phenotype ontology (HPO) was more recently developed to provide description of abnormal phenotypes in humans. The approach of HPO assumes the presence of a reference organism (humans) and abnormal phenotypes represent a deviation from the reference norm.

1.1.5 Human phenotype ontology

The HPO phenotype vocabulary and ontology initially published in 2008 (Robinson et al. 2008; Köhler, Kindle, and Robinson 2021) was created to enable accurate phenotyping for diseases. HPO currently contains over 13,000 terms, describes phenotypic information regarding over 7000 diseases, and is currently the de facto standard for deep phenotyping of rare diseases. Each term in HPO describes a distinct phenotypic feature (e. g. abnormality of body height), and these terms are ordered in an ontology. This means that phenotype terms are ordered and linked in a hierarchical manner, with more general terms close to the root, followed by terms with increasing specificity below (Figure 3). The disease-based information content (IC) of each term in HPO can be estimated through its frequency among the entire OMIM annotation corpus.

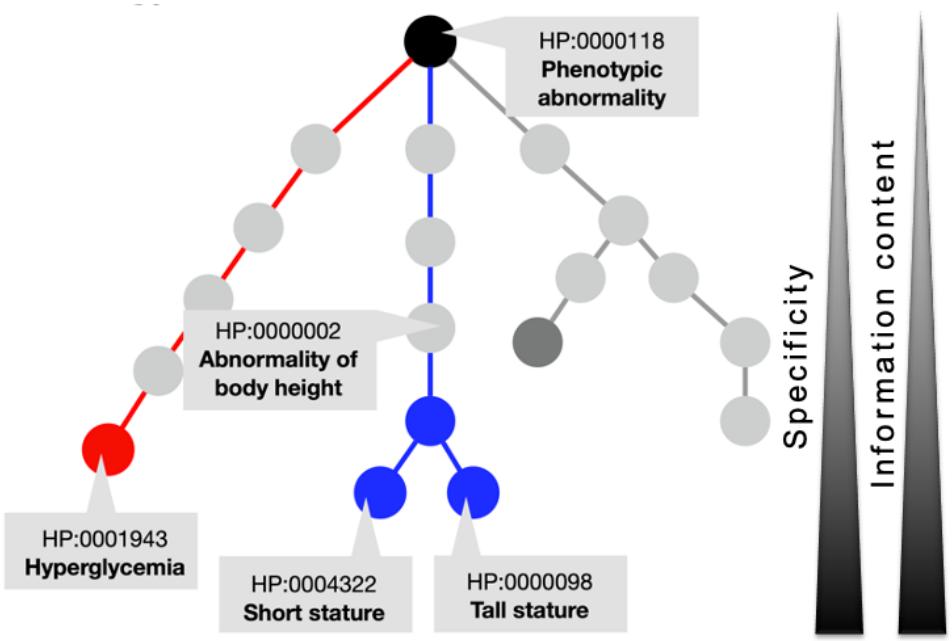


Figure 3. An example of an HPO branch. More general terms are located closer to the root, while more specific terms are further below. Information content and specificity increases further from the root.

Similar to most modern ontologies, HPO utilizes the Web Ontology Language (OWL) (Grau et al. 2008) format to store content. HPO is a community based tool and increasingly adapted into everyday use as the standard to describe phenotypic abnormalities (Köhler, Kindle, and Robinson 2021). Crowd-sourcing initiatives exist that aim to translate the original English HPO terms to other languages. In addition, a translation of HPO terms into common, everyday language exists to facilitate the use of HPO outside of strict clinical and research bounds (Vasilevsky et al. 2018). As a result, to date HPO provides the most comprehensive resource for deep phenotyping of diseases for researchers, clinicians, bioinformaticians and EHR systems.

HPO has also emerged as a go-to vocabulary for bioinformatics-based phenotype and disease-analysis. Specific tools and resources have been developed for gene prioritization such as Exomiser (Smedley et al. 2015) and Lirical (Robinson et al. 2020), or PhenoRank (Cornish, David, and Sternberg 2018) that use HPO to annotate and filter potentially causal variants from NGS sequencing files. Tools have been developed that use HPO to analyze genome-wide association (GWAS) data (Sveinbjornsson et al. 2016; Beck, Shorter, and Brookes 2020), and to understand genomic variation (Posey et al. 2017).

The wide use of HPO has enabled ample improvements in diagnostic accuracy and variant prioritization. Usage of HPO is customary for the analysis of WES and WGS data. This is often achieved through tools and platforms such as Exomiser and Lirical, or by using in-house, cohort-specific approaches (Taylor et al. 2017; Fang et al. 2017; Posey et al. 2016; Retterer et al. 2016; Zhu et al. 2015; T. Fujiwara et al. 2018; Thiffault et al. 2019; Stokman et al. 2018; Trujillano et al. 2017). In addition to clinical care and diagnosis, HPO is also increasingly used in commercial applications and data sharing platforms (Segal et al. 2017). Phenotips is a free and open source software for collecting and analyzing phenotypic information of patients with genetic disorders that is widely used in the rare disease community (Girdea et al. 2013). Data sharing platforms such as GeneMatcher and Matchmaker exchange benefit from the use of HPO and its standardized nomenclature which facilitates accurate patient matching.

1.1.6 Assessment of the ontological similarity of phenotypic terms

For each term in the HPO ontology, its information content (IC) can be calculated as a function of its frequency among OMIM annotations. Therefore, for each *t* term of HPO, the IC can be determined as the negative logarithm of the probability of finding term t in the whole OMIM annotation corpus: $-\log p(t)$.

Multiple IC-based methods exist for assessing the semantic similarity of two ontological terms, or sets of ontological terms. Node based measures such as the Resnik similarity measure (Resnik 1995) only considered the IC of the most specific concept which is an ancestor of both terms t1 and t2 terms, which is also termed the lowest common subsumer (lcs). The Resnik similarity of two terms is given by:

$$Sim_{Res}(t1, t2) = IC(lcs(t1, t2))$$

where *lcs(t*1,*t*2) is the lcs of concepts *t*1 and *t*2, and *IC* returns the IC of a term.

The Lin semantic similarity measure (Lin and Others 1998) is a variation of the Resnik measure and defined by:

$$Sim_{Lin}(t1, t2) = \frac{2xIC\big(lcs(t1, t2)\big)}{IC(t1) + IC(t2)}$$

After a general introduction to rare diseases and rare diseases of the immune system, the next part of the introduction expands on immune processes diseases in detail that are particularly relevant for autoimmune and autoinflammatory diseases, as they constitute the topic of the research article included in Results 3.2.

1.2 The immune system, diseases of autoimmunity and autoinflammation

Our immune system has developed as our main defense system against foreign invaders. It is a multi-leveled and complex system, with two main lines of response to infectious and other threats, called innate and adaptive immunity (C. A. Janeway et al. 2005). Both of these components of immunity work together to identify and eliminate threats. The innate immune system constitutes the first line, non-specific response to threats and foreign invaders. Adaptive immune system constitutes the second line of defense against non-self invaders. It is only found in vertebrates, and is specific to the pathogen presented. In both adaptive and innate immunity, the immune response is mounted against foreign proteins and polysaccharide molecules that are non-self, called foreign antigens (C. A. Janeway et al. 2005). The ability of our immune system to recognize and differentiate non-self from self, and mount an effective immune response is the basis of immune homeostasis. In case this homeostasis is compromised, and an immune response is mounted against self-antigens, autoimmunity and autoinflammation can arise (Delves 1998; Theofilopoulos, Kono, and Baccala 2017; Kastner, Aksentijevich, and Goldbach-Mansky 2010). The following subchapters of the introduction introduce the basic elements of innate and adaptive immunity, in particular focusing on those mechanisms and cell types that are crucial in the context of autoimmunity and autoinflammation.

1.2.1. The innate immune system

The first line of this defense system against non-self-invaders is the innate immune system (Monie 2017), and is constituted of physical, chemical and cellular defenses against pathogens. The main functions of the innate immune system include i) serving as a physical and chemical barrier, ii) the recruitment of more immune cells to the infection site through the production of chemicals

such as cytokines, iii) the activation of the complement cascade, iv) to identify and remove foreign substances and v) the activation of the adaptive immune response (Paulsen, Garreis, and Bräuer 2013). The various epithelial surfaces that constitute the barriers of our body and organ systems are impermeable to most infectious agents, acting as the first defense system against invaders. These surfaces provide an unsuitable environment for the survival of most microbes (Koyama et al. 2008). One of the first acute responses to infection or irritation is inflammation initiated by chemical agents released by injured cells (Stvrtinová, Jakubovsk\`y, and Hulín 1995). Inflammation is initiated by various cells residing in the tissues, and mainly include white blood cells (or leukocytes) such as phagocytes, histiocytes (tissue macrophages or dendritic cells), Kupffer cells and mast cells. Leukocytes are able to move freely within the body and are able to interact and capture debris, foreign particles and microorganisms. Although different types of cells are specialized for various functions, all of these cell types present receptors on their cell surface that are able to recognize foreign molecules that are distinguishable from self and shared by pathogens. These receptors, collectively termed pattern recognition receptors (PRRs), are therefore able to recognize pathogen-associated molecular patterns (PAMPs) (Takeuchi and Akira 2010).

1.2.1.1. Innate immune cell types

Innate immune cells are the products of multipotent hematopoietic stem cells present in the bone marrow, and unlike many other cells in the body are unable to divide and reproduce on their own (Blackstone 2003). The innate leukocytes include mast cells, natural killer (NK) cells, eosinophils and basophils, and phagocytes. Phagocytes are leukocytes that are capable of engulfing or “phagocytosing” particles and pathogens. They include macrophages, neutrophils and dendritic cells. Macrophages are efficient phagocytes that are able to move through capillaries to pursue invading pathogens. The binding of a PAMP to the PRR on the surface of a macrophage triggers a mechanism whereby the macrophage engulfs and destroys the agent through the generation of respiratory burst, causing the release of reactive oxygen species (ROS) (N. Fujiwara and Kobayashi 2005). Specialized macrophages include Kupffer cells (stellate macrophages) (Decker 1990) that are localized in the liver sinusoids, and histiocytes that are tissue-specific (Cline 1994). Neutrophils are granulocytes for the granules present in their cytoplasms. These granules contain various chemicals designed to neutralize or inhibit growth of bacteria and fungi. Similarly to macrophages, neutrophils rely on respiratory burst and oxidizing agents to eliminate foreign

agents. Neutrophils are the most abundant phagocytes and represent about 50-60% of total circulating leukocytes (Nauseef and Borregaard 2014). Dendritic cells (DCs) are specialized phagocytes that are in contact with the external environment, mainly the skin (where they are called Langerhans cells), inner mucosal lining of the nose, stomach and intestines. Dendritic cells, as well as being able to phagocytose and eliminate foreign molecules, are essential links between the innate and adaptive immune system, as they are prevalent players in antigen presentation - a process whereby foreign molecules are presented on the surface of antigen presenting cells for the engagement and activation of other cell types, mainly T cells (Banchereau and Schmitt 2012; Guermonprez et al. 2002).

Mast cells reside in connective tissues and mucous membranes. When activated by PAMPs, mast cells release histamine and heparin rich granules, along with chemokines and cytokines for the recruitment of other immune cells (Metcalfe, Baram, and Mekori 1997). Histamine, an organic nitrogenous compound is responsible for dilation of local blood vessels, and recruits neutrophils and macrophages to the place of infection or inflammation (*Advances in Experimental Medicine and Biology* 1967).

Basophils and eosinophils are granulocytes related to neutrophils. When activated through their PRRs, basophils release histamine and are important players in the defense against parasites (Galli, Chatterjea, and Tsai 2014). Basophils also play a role in allergic reactions and asthma (Marone et al. 2005). Eosinophils secrete highly toxic chemicals as a result of activation that are effective in killing parasites (Hogan et al. 2008).

In contrast to other innate cell types specialized in attacking and eliminating microbes and foreign agents, NK cells destroy host cells that have been compromised, such as tumor or virus infected cells. On healthy host cells, the surface expression of major histocompatibility complex (MHC) I molecule is abundant, whereas compromised cells have low levels of MHC-I. NK cells harbor their killer cell immunoglobulin receptors (KIR) (Raulet 2004) on their surface, and are able to recognize and get activated if a cell has low-levels of MHC-I.

1.2.1.2. The complement system

The complement system or complement cascade is designed to enhance the ability of phagocytic cells and antibodies to eliminate microbes and damaged host cells, promote inflammation and attack the cell membrane of pathogens (Charles A. Janeway). Although part of the innate immune system, the complement cascade can be activated by antibodies generated during the adaptive

immune response (Carroll 2004). The complement system triggers the attack of the membranes of pathogens by phagocytic cells, phagocytosis by opsonization of the antigens, and inflammation by attracting macrophages and neutrophils. The complement system is made up of over 30 proteins and protein fragments. The majority of these proteins are synthesized in the liver by hepatocytes, but some of them are also produced by macrophages, genitourinary epithelial cells and epithelial cells of the gastrointestinal tract. These proteins circulate in the bloodstream as inactive precursors. When activated through one of the triggers (discussed below) proteases in the complement system cleave specific complement proteins to initiate the release of cytokines and the cleavage of further protein complexes. The complement system can be activated through three different biochemical ways (Sarma and Ward 2011).

In the classical pathway, the complement is activated by the C1-complex consisting of C1q binding to IgM or IgG antibodies. The majority of complement activation happens through the alternative pathway, without the presence of specific antibodies. The activation through the alternative pathway is a result of spontaneous hydrolysis of C3, or can be brought on by foreign material, pathogens, or damaged cells. Finally, the complement can be also activated through the lectin pathway, which is similar to the classical pathway, but instead of C1q and antibodies, mannose-binding lectin and ficolins acting as the opsonins. All these three pathways converge in the activation of homologous variants of the protease C3-convertase, which initiates the opsonization of particles, the release of inflammatory molecules and C5 convertase formation and cell lysis (Ricklin, Reis, and Lambris 2016; Charles A. Janeway, n.d.). The end result of the complement activation is the stimulation of phagocytes to clear and eliminate foreign and damaged material, inflammation and the attraction of more phagocytes, and the activation of the membrane attack complex (MAC), a protein complex designed to kill cells. The MAC forms on the surface of pathogen cell membranes, and places pores on the membranes which leads to cell lysis and death of pathogens (Charles A. Janeway, n.d.).

1.2.1.3. Inflammation

Upon the onset of an infection or injury, leukocytes are activated when the PRRs recognize PAMPs and release various inflammatory mediators such as histamines, bradykinin or serotonin (C. A. Janeway et al. 2005). These chemicals kickstart inflammation, and sensitize pain receptors and cause dilation of local blood vessels, and attract more cells such as neutrophils and macrophages. The neutrophils and macrophages recruited release chemicals that attract more

cells including further innate immune cells and lymphocytes. These chemicals include cytokines, which are small proteins specialized in cell-to-cell signaling and immunomodulation. Cytokines act through cell surface receptors, and induce various changes such as the maturation, growth, and responsiveness of particular cells (Paul and Seder 1994). The most prevalent cytokines secreted in the inflammatory response are interleukin-1 (IL-1), IL- 12, IL-18, tumor necrosis factor alpha (TNF-a), interferon gamma (IFN-gamma) and granulocyte-macrophage colony stimulating factor (GM-CSF) (Glauser 1996). As a result of the physiological, cellular and chemical changes in the local environment, the inflammatory response is characterized by a cascade of symptoms, including skin redness, increased local temperature or systemic fever, swelling of the affected tissues, increased mucus production and the sensation of local or global pain.

1.2.2. The adaptive immune system

The adaptive immune system, also referred to as the acquired immune system, is composed of specialized white blood cells that are designed to eliminate pathogens. In contrast to the general immune response elicited by the activation of the innate immune system, the adaptive immune response is highly specific to the pathogen encountered by the organism (Davies 1997). The adaptive immune system can be triggered by a pathogen that evades the innate immune response, and danger signals are generated and activate DCs. The major tasks of the adaptive immune system include recognition of non-self-antigens in the presence of self-antigens in the process of antigen presentation, the generation of pathogen specific responses and responses to pathogen infected cells, and the development of immunological memory (Klenerman 2017). The adaptive immune response is carried out by specialized white blood cells, or lymphocytes, discussed in more detail below.

1.2.2.1. Antigen presentation

The acquired immune response relies on the capacity of the immune cells involved to distinguish between the body's own cells and foreign invaders. Antigen presentation is the immune process by which cellular molecules are displayed on cell surfaces bound to the MHC (Lindsay Whitton 2013). At all times, the host's cells express self-antigens (endogenous molecules) on their surface in their MHC-I molecules. These self-antigens are different from antigens from foreign invaders

such as bacteria or virus infected host cells. Antigens that are from the extracellular space (extracellular antigens) are presented on the surface of specific antigen presenting cells in MHC-II molecules (Neefjes et al. 2011). Specialized antigen presenting cells include DCs, macrophages and B cells. These extracellular antigens within the MHC-II are presented to specialized lymphocytes (T helper cells, discussed below) and activate them to kickstart the adaptive immune response (Guermonprez et al. 2002).

1.2.2.2. Adaptive immune cell types

The cells involved in the adaptive immune response are commonly known as lymphocytes. Lymphocytes, as innate immune cells are derived from multipotent hematopoietic stem cells. T and B cells are the two main types of lymphocytes that carry out the cell-mediated immune response and antibody response (M. D. Cooper and Alder 2006).

T cells contain a protein complex on their surface called the T-cell receptor (TCR), which is designed to recognize fragments of antigens bound to MHC molecules. Cytotoxic T cells (or CD8 T cells, or CTL) are a subtype of T cells that are specialized in eliminating host cells that are infected with viruses, or which are damaged or dysfunctional. Cytotoxic T cells are activated when their TCR interacts with an MHC-I molecule that is presenting an antigen (Andersen et al. 2006). Once activated, cytotoxic T cells undergo clonal selection, a process whereby activated by a specific antigen, a specific type of T cell rapidly multiplies and produces identical clones of itself (Burnet 2015). When in contact with those cells expressing the favorable antigen, CTLs release perforin and granulysin, chemicals that form pores on the plasma membrane, and granzyme that induces apoptosis (Andersen et al. 2006). In contrast to cytotoxic T cells, CD4 T cells do not have cytotoxic activity, but are specialized to manage and direct other types of immune cells. CD4 T cells also express TCR on their surface, and are activated through contact with an APC that is presenting with an antigen in their MHC-II. Th1 CD4 T cells are involved in immune response against bacteria and viruses. They are characterized by INF-gamma production, which subsequently activates macrophages and B cells (Mosmann and Coffman 1989). Th2 Cd4 T cells orchestrate the response against extracellular bacteria and parasites. Th2 CD4 T cells release IL-5, which induces eosinophils and IL-4 that facilitates B cell isotype switching (Mosmann and Coffman 1989). Regulatory T cells (Tregs) regulate adaptive immunity by limiting and suppressing

the immune system's response to self-antigens. Tregs are immunosuppressive, and downregulate the expansion of other T cell types (Weinberg 2006), and have a major role in preventing immunity against self – otherwise known as autoimmune disease (La Cava 2009). These cells express biomarkers such as FOXP3 on their surface. Th17 T cells are pro-inflammatory T cells that are characterized by their production of IL-17. The triggers that signal to cause Th17 activation and differentiation inhibit the differentiation of Tregs. Th17 cells are involved in pathogen clearance and maintaining the homeostasis in mucosal barriers (Weinberg 2006; Korn et al. 2009). Follicular helper T cells (Tfh cells) help B-cell dependent humoral immunity. Tfh cells are able to migrate to follicular B cells and provide them signals to enable the generation of antibodies (Fazilleau et al. 2009).

Gamma delta T cells (or γδ T cells) house an alternative TCR on their surface. They share the traits of helper T cells, NK cells and CTLs. γδ T cells do not require a presentation of an antigen by an APC for activation, and are considered to be at the border of innate and adaptive immunity (Girardi et al. 2001).

B lymphocytes are the other major group of lymphocytes involved in the adaptive immune response. B cells are able to produce antibodies, Y shaped proteins that can attach to antigens, and circulate in blood plasma and lymph. There are five types of antibodies in mammals: IgA, IgD, IgE, IgG and IgM that possess different biological properties and are designed to bind to and handle different types of antigens (Hoffman, Lakkis, and Chalasani 2016). The type of immunity that is mediated by these peptides is called antibody-mediated or humoral immunity (Elgueta, de Vries, and Noelle 2010). B cells express an unique receptor on their surface, called the B cell receptor (BCR), which is a membrane-bound antibody. All the BCRs on the surface of a B cell only recognize one particular antigen. In contrast to T cells, B cells can recognize antigens in their native, non MHC-bound form. B cells are activated when encountering a matching antigen. Activated B cells produce antibodies that recognize unique antigens. The binding of antibodies to antigens triggers different immune mechanism: i) agglutination, that reduces the number of infectious elements, ii) complement activation, iii) opsonization - the antibody coating of antigens to facilitate phagocytosis, iv) the recruitment of macrophages, eosinophils, and NK cells to the antibody coated elements - also called antibody dependent cell-mediated cytotoxicity, and v) neutralization, or blocking adhesin of bacteria and viruses to mucosa (Elgueta, de Vries, and Noelle 2010). Once encountering a specific antigen, and receiving activation signaling from Th2

cells, B cells differentiate into effector B cells, also called plasma cells. Plasma cells have a short life span and their main function is to secrete antibodies (Kölmel 1977; Bernasconi, Traggiai, and Lanzavecchia 2002). During affinity maturation, Tfh cells stimulate B cells to produce antibodies with increased affinity for the antigen during the immune response. Affinity maturation involves two processes that both occur in the germinal centers of lymphoid organs. Somatic hypermutation (SHM) diversifies BCRs to recognize foreign elements, by a programmed process of mutation of the variable regions of immunoglobulin encoding genes (Di Noia and Neuberger 2007). Those B cells that have undergone SHM compete for growth resources in the lymph node. The B cells which are highly competitive and are able to conjugate Tfh cells in a stable manner receive T cell-dependent survival signals, while those B cells that are unable to stably bind and engage Tfh cells are deleted. Through this process, only those B cells remain that are able to produce highly effective antibodies to eliminate the foreign threat (Burnet 2015).

1.2.2.3. Immunological diversity

Most antigens contain multiple different epitopes, which are antigenic determinants that interact with an antibody of a receptor of a lymphocyte. Only a small percentage of lymphocytes can recognize and bind to a particular antigen (C. A. Janeway et al. 2005). Immune cells have to be able to differentiate between a multitude of different antigens. For this, the cell surface receptors that recognize antigens have to be produced in a large variety of configurations. Countless receptors are produced through a process called clonal selection. The clonal selection theory states that an animal is born with a diversity of randomly generated lymphocytes that each bear a unique antigen receptor (Burnet 2015). In order to generate each of these unique antigen receptors, the genes encoding different parts of the receptors undergo a process termed V(D)J recombination, that happens in the bone marrow (B cells) and thymus (for T cells). During this process, one gene segment recombines with other gene segments to produce unique genes (Ferrier 2009). V(D)J recombination and the translation of these unique receptor encoding genes generates a large diversity of receptors and antibodies and enables the immune system to respond to an almost unlimited diversity of antigens (C. A. Janeway et al. 2005).

1.2.2.4. Immunological tolerance

Immunological tolerance is defined as a state of unresponsiveness of the immune system to elements that are able to elicit an immune response in the organism. This tolerance is classified into central or peripheral tolerance. Central tolerance is originally induced in the thymus and bone marrow, while peripheral tolerance is induced in lymph nodes (Medawar 1961). Central tolerance is an essential process by which the immune system learns to discriminate self from non-self. Peripheral tolerance on the other hand is vital to preventing over-reactivity of the immune system to environmental entities such as allergens or gut microbes. Defects in either of the tolerance mechanisms can cause autoimmune or autoinflammatory diseases (Anderson and Kuchroo 2007).

1.2.2.4.1. Central tolerance

Central tolerance is the tolerance achieved by the elimination of autoreactive T and B lymphocytes before they become fully developed and immunocompetent. This process of negative selection enables the elimination of those T and B cells which are able to initiate a strong immune response to the hosts own cells and tissues, while preserving the ability to recognize foreign antigens.

This process of elimination is carried out in the thymus and bone marrow, where the developing lymphocytes are presented with self-antigens by thymic epithelial cells and thymic DCs, or bone marrow cells. T cell tolerance mechanisms are carried out in the thymus, where the cells first undergo a positive then a negative selection (Romagnani 2006). First, during positive selection T cells are tested for their ability to bind to MHC-antigen complexes. Those T cells which are unable to bind MHC-I or MHC-II undergo apoptosis from not receiving survival signals. Next, during negative selection, T cells are probed for their affinity to elicit an immune response to self. The transcription factor AIRE is a major orchestrator of the self-antigen presentation process (Perniola 2018). The lymphocytes that are able to strongly bind to self-antigens are eliminated by the induction of apoptosis or anergy (Fleit 2012). Those T cells that recognize the MHC-peptide complexes but do not bind to self are either CD4 or CD8 positive T cells and migrate to secondary lymphoid organs. Treg selection is carried out in the thymus as well, accompanied by the expression of the transcription factor FOXP3. Some T cells that are able to recognize self-

antigens in a weak manner are differentiated to Treg cells that circulate and act as sentinel cells in the periphery to identify cells with potential of T cell autoreactivity (Sakaguchi et al. 2008).

While in development, immature B cells undergo negative selection in the bone marrow if they are able to bind to self-peptides. In case a B cell becomes reactive upon binding a self-antigen, it will either undergo apoptosis, anergy, or receptor editing. During receptor editing, the B cell rearranges the genes encoding their BCR that does not respond to self (Meffre and Wardemann 2008). B cells that present weak autoreactivity might be ignored, as they don't respond to the stimulation of their BCR (Fleit 2012). Compared to T cells, the deletion threshold is less strict for B cells as they are unable to cause tissue damage directly.

1.2.2.4.2. Peripheral tolerance

The main purpose of peripheral tolerance is to eliminate the autoreactive T and B cells that have escaped the filtering mechanisms of central tolerance. Peripherial tolerance mechanisms take place in peripheral lymphoid organs. As T cell activation has to occur to elicit an adequate immune response, antigens that are presented in low numbers are generally ignored by the immune system. Immune-privileged organs - organs that can tolerate the presence of antigens without immune response, such as the eyes, the testis, the central nervous system or the placenta and fetus - have special mechanisms to ensure ignorance (Benhar, London, and Schwartz 2012). These mechanisms include the presence of physical barriers where lymphocytes cannot get through, or low levels of antigen presentation. Expression of apoptotic markers such as FAS ligand and the expression and presence of anti-inflammatory cytokines such as TGF-beta and IL-10 ensure that lymphocyte activation does not occur (Mueller 2010).

Most of the self-reactive T cells are deleted during central tolerance mechanisms, but low affinity self-reactive T cells are able to escape (Malhotra et al. 2016). During peripheral tolerance mechanisms, T cells that present with low affinity can be either deleted or converted to Treg cells when recognizing an antigen presented by an immature DC in secondary lymphoid organs (Steinman, Hawiger, and Nussenzweig 2003). In addition to DCs, other cell populations have been identified that are able to induce antigen-specific tolerance in T cells. These include lymph node stromal cells, which are able to present endogenous antigens on MHC-I, and peptide MHC-II complexes, and thereby have the capacity to induce CD4 and CD8 T cell tolerance (Fletcher

et al. 2010). Tregs, both the ones generated in the thymus and ones in the periphery are able to suppress autoreactive T cells by mechanisms such as depletion of IL-2 from the environment, and the expression of immunosuppressive and tolerogenic cytokines such as TGF-beta and IL-10 (La Cava 2009). Finally, T cells can be rendered non-responsive if the antigen is presented without further co-stimulatory signals such as the engagement of co-stimulatory molecules that are upregulated by pro-inflammatory cytokines (Mueller 2010).

1.2.2.5. Immunological memory

Our immune system encounters a variety of different threats presented by different antigens throughout our lifetime. Immunological memory is the ability of the immune system to swiftly recognize an antigen that it has encountered before, and quickly initiate a specific immune response to eliminate the threat (Ahmed and Gray 1996). During the immune response to an antigen, specific adaptive immune cells are created whose purpose is to serve as “memory” cells in the next immune response to the same pathogen. These cells are called memory T cells and memory B cells.

Memory T cells can both be CD4 positive, or CD8 positive T cells. In comparison to non-memory T cells, these memory T cells do not need a signal via their MHC to proliferate. Based on the surface expression of marker CCR7, a chemokine receptor, memory T cells can be grouped into two subsets. Effector memory T cells do not express CCR7 are able to migrate to the site of infection and represent a population of cells that is able to immediately participate in the immune response, and produce IFN-gamma, IL-4 and IL-5. Central memory T cells express CCR7 on their surface and lack proinflammatory and cytotoxic function, These cells are able to migrate to the lymph node and stimulate DCs, then differentiate into effective memory T cells.
A small fraction, about 10% of plasma cells survive and become memory B cells that are already primed to produce those specific antibodies designed for the antigen that their progenitors have encountered (Kölmel 1977; Bernasconi, Traggiai, and Lanzavecchia 2002). These memory cells have already undergone affinity maturation and are able to produce specific antigens. Both T and B memory cells can persist for decades in the body (Sallusto et al. 1999; Bernasconi, Traggiai, and Lanzavecchia 2002), enabling a swift and effective immune response.

1.2.3 Autoimmune and autoinflammatory diseases

Autoimmunity and autoinflammation arise when either the adaptive or innate immune processes are over-reactive or chronically activated, leading to an attack on self instead of external invaders. Autoimmunity is defined as the system of immune responses that an organism elicits against its own healthy cells and tissues (Delves 1998; Theofilopoulos, Kono, and Baccala 2017). Autoimmunity is usually characterized by the break of self-tolerance of adaptive immune cells that leads to phenotypes such as fatigue, fever, malaise, muscle and joint aches and rashes (L. Wang, Wang, and Eric Gershwin 2015). Traditionally, autoimmune phenotypes that affect different organ systems have been classified as separate diseases. These diseases include inflammatory bowel disease (IBD) that presents with chronic inflammation of the intestines (Denmark and Mayer 2014; Abraham and Cho 2009), coeliac disease that presents with intestinal inflammation as a result of exposure of gluten (Meresse et al. 2009), multiple sclerosis, a neurodegenerative disorder where T cells attack the myelin sheath of brain neurons (Frohman, Racke, and Raine 2006), rheumatoid arthritis, a chronic inflammation of joints (Turesson and Matteson 2009), systemic lupus erythematosus (SLE) that is associated with a wide-loss of immune tolerance and chronic inflammation (Tsokos 2011) and autoimmune anemia, a failure of the body to produce adequate number of lymphocytes.

Autoinflammation arises when innate immune cells become over-activated, as a result of dysregulated secretion of pro-inflammatory cytokines (Kastner, Aksentijevich, and Goldbach-Mansky 2010). Autoinflammatory diseases are marked by fevers, rashes, joint and muscle pain and systemic inflammation (Moreira et al. 2017). Most autoimmune and autoinflammatory diseases are classically considered complex heterogeneous diseases, brought on by the combination of various genetic and environmental risk factors (Eyre, Orozco, and Worthington 2017). The majority of these diseases display a heterogeneous combination and penetrance of autoimmune and autoinflammatory phenotypes, and are frequent in Western countries (G. S. Cooper, Bynum, and Somers 2009). The genetic architecture of common autoimmune and autoinflammatory diseases is usually investigated with GWAS, which have revealed thousands of genomic loci associated with different common autoimmune and autoinflammatory diseases (Caliskan, Brown, and Maranville 2021).

Intriguingly, as immune dysregulation is a permanent feature of IEI, autoimmune and autoinflammatory phenotypes have been documented in rare diseases of the immune system

(Köstel Bal et al. 2020; A. Bousfiha et al. 2020; Fischer et al. 2017). Indeed, it has been recently shown that autoimmune and autoinflammatory manifestations are increasingly common complications seen in IEI patients, especially for CVID and CID patients (Fischer et al. 2017). The next chapter introduces the genetics of rare autoimmune and autoinflammatory diseases in more detail.

1.2.3.1. Mechanisms of rare autoimmunity and autoinflammation

A considerable fraction of IEI have been documented to present with either autoimmune or autoinflammatory phenotypes (Figure 4). The overall survival of IEI patients with autoimmune and autoinflammatory phenotypes is generally less favorable as compared to IEI without these manifestations (Fischer et al. 2017). The gene defects and affected molecular mechanisms leading to autoimmune and autoinflammatory phenotypes can generally be grouped into six categories:

I. Defects of lymphocyte development, differentiation, activation and selection. These usually constitute defects in VDJ recombination and T cell receptor (TCR) signaling.
II. Diseases of immune dysregulation and loss of tolerance. In these diseases, loss of central tolerance or defect in regulatory T cells are observed, but also signal transducer activator of transcription (STAT) defects are common.
III. Antibody disorders that are often genetically undiagnosed and affect various B cell functions.
IV. Phagocytic disorders that arise as a defect of oxidative burst due to mutations in the NADPH genes.
V. Complement defects that affect elements of the complement cascade.
VI. Autoinflammatory diseases that arise due to defects in IL-1, IFN-alpha or NFkB signaling, or the inflammasome.

The next chapter introduces the gene defects and molecular pathomechanisms underlying early-onset IBD, a prototypic autoimmune and autoinflammatory disease.

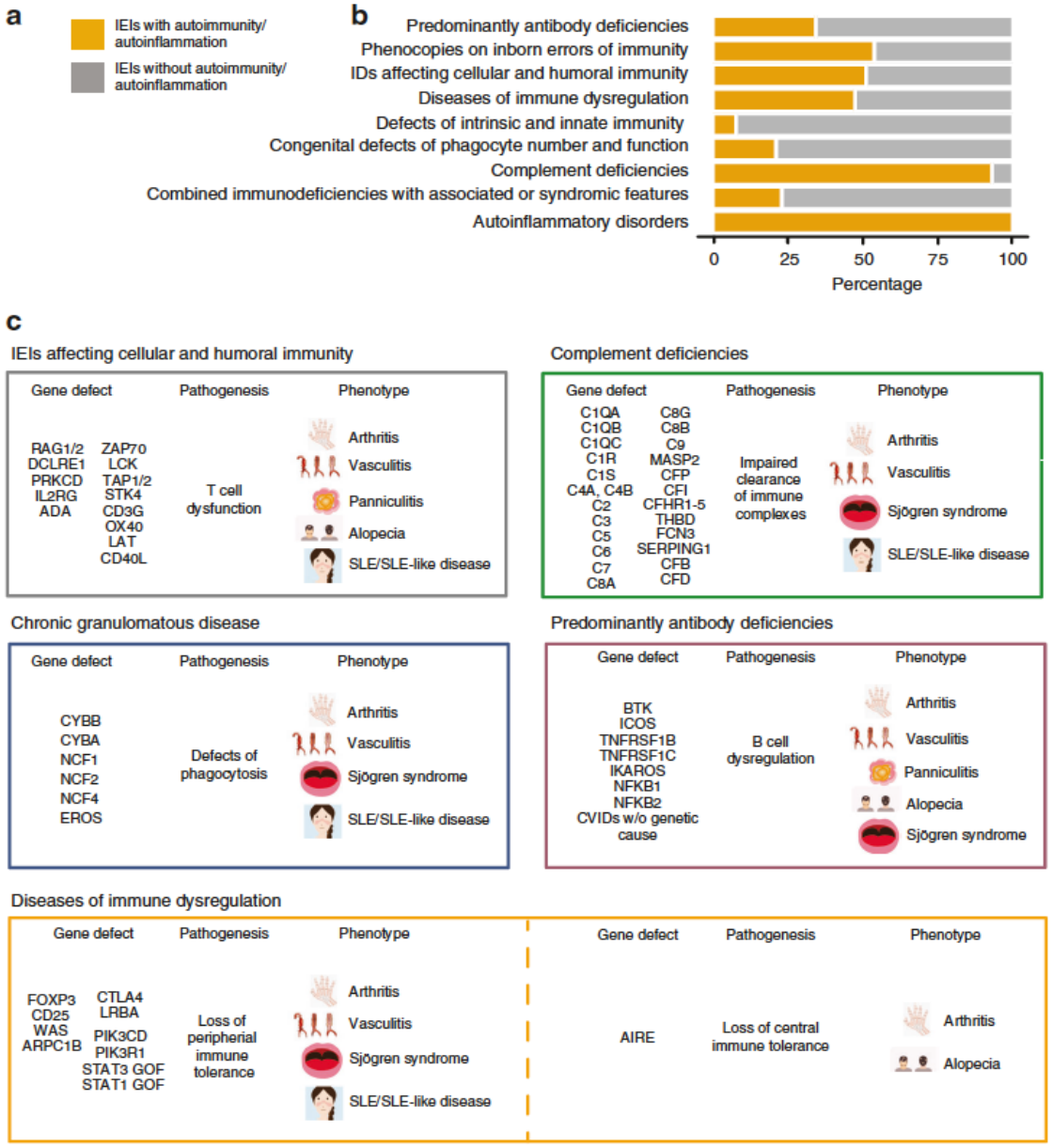


Figure 4. Genetics and phenotypes of monogenic autoimmune and autoinflammatory diseases. a,b) Percentage of autoimmune and autoinflammatory phenotypes in IEI. c) Phenotypic presentation of autoimmunity and autoinflammation in IEI. Reprinted by permission from Springer Nature: Pediatr. Res., “Rheumatological manifestations in inborn errors of immunity” (Köstel Bal et al. 2020).

1.2.4. Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms

Julia Pazmandi, Artem Kalinichenko, Rico Chandra Ardy, Kaan Boztug. 2019. **Immunological Reviews** 287 (1): 162-85. DOI: 10.1111/imr.12726

This review article introduces the gene defects and affected mechanisms of early-onset IBD, one of the model diseases of autoimmunity and autoinflammation. In order to fully appreciate the heterogeneity of rare diseases of the immune system on the genetic, molecular and phenotypic level, it is crucial to get a detailed account of the genetic perturbations that lead to specific rare diseases. The manuscript gives a general overview of monogenic, Mendelian IBD, and the postulated disease pathomechanisms that lead to the observed bowel inflammation phenotype. The article discusses the cell types and molecular mechanisms known to be involved underlying monogenic IBD, as well as introduces how recent advances in genomic technologies have influenced therapeutic guidelines and clinical care. In addition, bowel inflammation in the context of the microbiome and organoid technologies are discussed. Next, the review gives an overview of the future of genetics in IBD, on topics such as NGS technologies and variant prioritization and interpretation. The manuscript further details the available data sharing platforms including a chapter on nomenclatures and ontologies. Finally, systems-biology and other more integrative methods are discussed that show promise in elucidating research questions in connection with early-onset IBD.

Received: 17 September 2018 | Accepted: 23 September 2018

DOI: 10.1111/imr.12726

INVITED REVIEW

WILEY Immunological Reviews

Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms

Julia Pazmandi[1,2] | Artem Kalinichenko[1,2] | Rico Chandra Ardy[1,2] | Kaan Boztug[1,2,3,4]

[1]Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria

[2]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

[3]Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria

[4]Department of Pediatrics, St. Anna Kinderspital and Children's Cancer Research Institute, Medical University of Vienna, Vienna, Austria

Correspondence
Kaan Boztug, MD, Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria.
Email: kaan.boztug@rud.lbg.ac.at

Summary

Rare, monogenetic diseases present unique models to dissect gene functions and biological pathways, concomitantly enhancing our understanding of the etiology of complex (and often more common) traits. Although inflammatory bowel disease (IBD) is a generally prototypic complex disease, it can also manifest in an early-onset, monogenic fashion, often following Mendelian modes of inheritance. Recent advances in genomic technologies have spurred the identification of genetic defects underlying rare, very early-onset IBD (VEO-IBD) as a disease subgroup driven by strong genetic influence, pinpointing key players in the delicate homeostasis of the immune system in the gut and illustrating the intimate relationships between bowel inflammation, systemic immune dysregulation, and primary immunodeficiency with increased susceptibility to infections. As for other human diseases, it is likely that adult-onset diseases may represent complex diseases integrating the effects of host genetic susceptibility and environmental triggers. Comparison of adult-onset IBD and VEO-IBD thus provides beautiful models to investigate the relationship between monogenic and multifactorial/polygenic diseases. This review discusses the present and novel findings regarding monogenic IBD as well as key questions and future directions of IBD research.

KEYWORDS

genetics, inborn errors of immunity, inflammatory bowel disease, pathomechanisms

1 | BACKGROUND

1.1 | Inflammatory bowel disease (IBD)

The gastrointestinal (GI) tract is the largest lymphoid organ in the body and contains a multitude of diverse cell types including enterocytes, Goblet cells, enteroendocrine cells, Paneth cells, but also T and B cells, macrophages, dendritic cells, and innate lymphoid cells.[1-3] Despite the fact that these cells are constantly confronted with antigens, primarily in the form of food and bacteria, immune responses in the gut are tightly regulated to maintain homeostasis. IBD refers to a heterogeneous group of diseases that present with bowel inflammation and intractable diarrhea[4] as a result of an inappropriate inflammatory response and unbalanced crosstalk between the gut lumen and mucosal immune system. IBD is often classified according to histopathological features as Crohn's disease, ulcerative colitis, or indeterminate colitis.[5] Adult-onset IBD is common and generally considered a complex, multifactorial disease where a combination of factors, including host genetics and environmental factors (including the microbiome), influence disease onset.[6,7] Due to the complex nature of adult IBD, research unraveling the genetic aberrations behind this phenotype has focused

This article is part of a series of reviews covering Lessons primary immunodeficiencies teach about the healthy and diseased immune system appearing in Volume 287 of *Immunological Reviews*.

on identifying genetic risk factors using genome-wide association studies (GWAS). In the last decade, intensive research using GWAS has identified over 230 IBD-associated loci comprising approximately 300 potentially associated genes,[8-11] including *NOD2*, *ATG16L1*, *IRGM*, *IL23R*, *CARD9*, *RNF186*, and *PRDM1*. Although there are only a few GWAS SNPs with evidence of biological involvement in IBD, such as missense SNPs in *NOD2*[12] and *ATG16L1*,[13] identification of such associations pinpointed crucial mechanisms such as autophagy, pattern-recognition, Th17 involvement, and maintenance of the epithelial barrier in IBD pathogenesis.[11] Recent efforts have focused on meta-analysis and fine mapping of existing GWAS datasets using innovative approaches such as Bayesian analysis,[14] as well as adding novel, valuable cohorts to identify new loci. Among newly identified loci are SNPs pointing to integrin genes[10] *ITGA4* and *ITGB8*. Integrins are transmembrane receptors that facilitate extracellular matrix adhesion, thereby are important in the homeostasis of the epithelial barriers.

Interest in the potential common component of immune-mediated diseases has lead to inter-disease comparisons and identification of shared loci between IBD and other autoimmune or inflammatory conditions such as juvenile idiopathic arthritis, primary sclerosing cholangitis, psoriasis, multiple sclerosis, and ankylosing spondylitis.[15-17] These pleiotropic loci point to shared pathways and molecular mechanisms underlying the heterogeneous immune-mediated diseases.

Despite current advances in data collection and analysis, our understanding of SNPs outside the coding regions is still elusive. It has been shown that SNPs for autoimmune disease tend to be enriched in regulatory regions, and in differentially expressed genes, and that risk variants for autoimmune diseases show particular enrichment in active chromatin regions of immune cells.[18-20] In addition, several efforts have been made to unravel how SNPs at a locus affect mRNA expression of genes. These efforts combining GWAS with transcriptome analysis have revealed that pinpointing the causal SNP in a haplotype block is a non-trivial task and that many of the SNPs have a detectable effect only in a cell type-dependent or stimulus dependent context.[21-23]

As our knowledge for non-coding regions of the genome is growing relating SNPs to regulatory regions, as well as assaying the cell type specificity of loci, will be important goals for the future. Notably, it is unclear how the identified susceptibility loci and associated genes identified in these GWAS studies relate to the early-onset, Mendelian form of IBD. Identification of high impact SNPs in *NOD2* that are associated with adult Crohn's disease with clear involvement in IBD pathogenesis has illustrated a genetic continuum between adult and early-onset IBD, in contrast to the classical view of two genetically independent diseases.[24] In this context, we can hypothesize that adult and Mendelian IBD arise as a result of a spectrum of varyingly pathogenic genetic lesions that impact common key pathways in IBD. Despite these advances, the exact relationship between adult and Mendelian IBD is still poorly understood. The lack of understanding of (adult) IBD is also reflected in the fact that there are currently only a few stratified/personalized treatment strategies despite the recent expansion of therapies based on immune modulation, mostly using monoclonal antibodies.[25] Given these challenges, the precise mechanisms of IBD disease pathogenesis, the relationship between adult and early-onset IBD, and the complex interplay between host genetics and environmental factors have remained partially elusive with major gaps in our understanding in the genetic processes governing IBD pathology.[8]

1.2 | Monogenic and Mendelian IBD

Very early-onset IBD (VEO-IBD) denotes a subgroup of IBD patients with a disease onset before the age of 6 years.[27] In contrast to adult IBD, VEO-IBD is a rare disease where mutations in causal genes may be inherited in a Mendelian fashion, as illustrated by our discovery of IL10R deficiency.[26] VEO-IBD patients usually present with a severe clinical course including (often bloody) diarrhea and abdominal pain.[27] Most patients with VEO-IBD receive immunosuppressive treatment, and many patients require surgical intervention during the course of their disease.[28] To date, there are only a handful of monogenic defects that result in a predominant IBD phenotype, including *ADAM17*, *IL10*, *IL10RA*, *IL10RB*, *GUCY2C*, *IL21*, *LRBA*, *TTC7A*, and *XIAP*.[26,29-34] Identification of these gene defects have provided proof of concept for genetic diagnosis and stratified therapeutic choices, shaping our understanding of the immune system and illustrating molecular mechanisms underlying the delicate balance in keeping the homeostasis in the gut.

Intriguingly, a spectrum of inborn errors of immunity (IEI) can present with an IBD-like phenotype, sometimes as the initial disease manifestation. IEIs are a heterogeneous group of more than 330 different disorders with around 300 genes currently identified to be associated with monogenic, Mendelian forms.[35] The main characteristics of IEIs are increased susceptibility to infections due to improper function or dysregulation of key players of the immune system. These observations highlight the interesting fact that a spectrum of different immunopathological processes can underlie GI inflammation and point to the GI tract as an exceptionally sensitive site to immune disturbances. Current consensus estimates that about 20% of genetic defects underlying IEIs can develop bowel inflammation (Figure 1A). The International Union of Immunological Societies (IUIS) recognizes 9 phenotypic groups of IEIs.[35] Among the functional groups of IEIs, diseases of immune dysregulation present most often with an IBD-like phenotype in up to 40% of the different genetic defects. On the other hand, complement deficiencies tend to present without bowel inflammation (95% of known gene defects do not cause IBD, Figure 1A). To date, considerably accelerated by the advent of next-generation sequencing, >60 monogenic diseases that present with IBD have been described[27,36] (Figure 1A). Between the year 2015 and 2018 alone, several new gene defects have been identified that underlie some type of bowel inflammation (Figure 1B).

Interestingly, some gene defects in subgroups of IEI do present with bowel inflammation, while other gene defects in the same

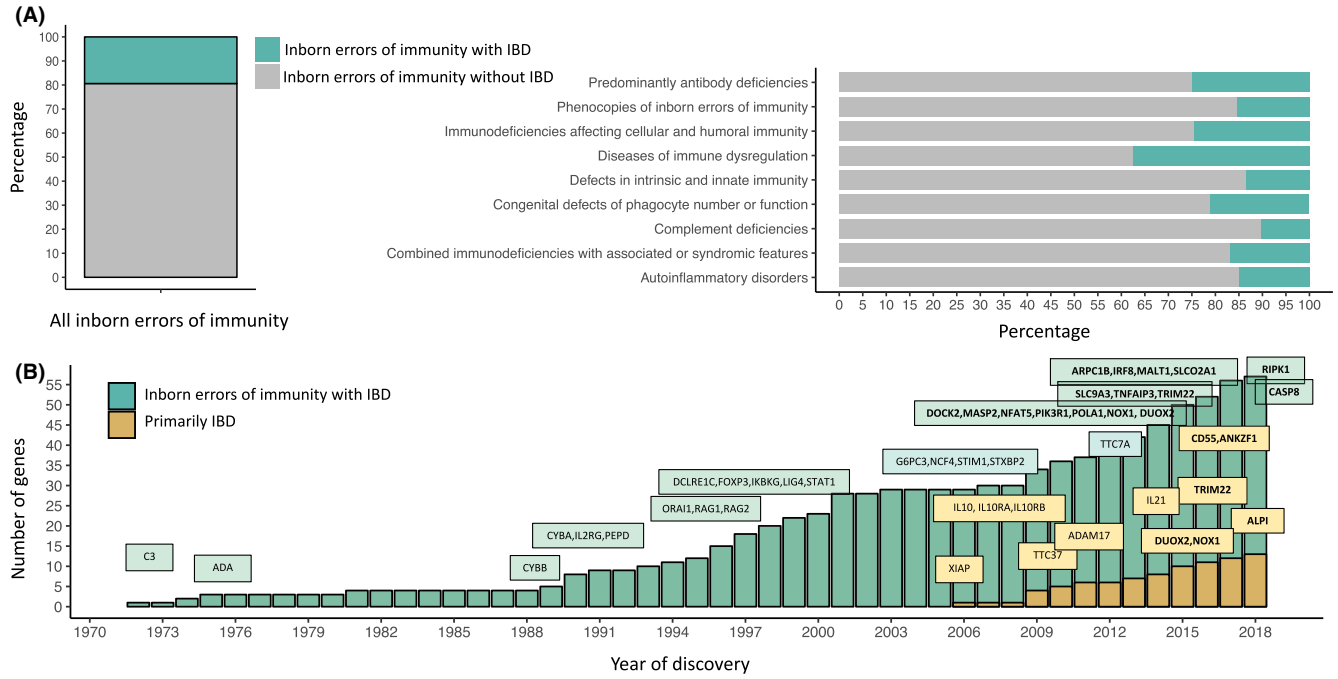


FIGURE 1 Advances in identification of genetic etiologies underlying inflammatory bowel disease and inborn errors of immunity. (A) The percentage of inborn errors of immunity with IBD. Classification according to the 2017 International Union of Immunological Societies (IUIS) phenotypic classification of inborn errors of immunity.[35] (B) Discoveries of inborn errors of immunity with IBD and VEO-IBD genes through the years. Gene defects that were described between 2015 to 2018 are highlighted in bold

group do not. While currently, there is no comprehensive and satisfactory explanation for the varying frequency of the IBD phenotype in individual gene defects, one can speculate that (a) due to the few patients and therefore small sample size in rare diseases, it is possible that certain phenotypes of inborn errors of immunity have not yet been captured, especially when it comes to disease with only one patient described at present, (b) our knowledge of the explicit effects of genetic aberrations is incomplete; therefore, it is plausible that in some gene defects counter-mechanism are in place and can maintain a pseudo-homeostatic state in the gut, therefore not inducing an IBD-like phenotype, and (c) since our understanding of the influence of factors extrinsic to genetic triggers is incompletely studied and understood in EO-IBD, it is likely that (similarly to adult IBD) in some cases the EO-IBD phenotype only emerges as a result of strong non-genetic triggers on a genetically susceptibly host.

Investigating the consequences of genetic aberrations in patients with monogenic defects causing IBD allows for a precise dissection of genotype-phenotype relationship. Moreover, through understanding of the mechanistic effect of pathogenic mutations on gene regulation, we have widened our knowledge on principal immune processes. Therefore identification of monogenic defects underlying IBD has not only provided genetic diagnosis to patients, but also proven to yield invaluable insights into how the immune system works. We here review monogenic defects underlying IBD and how dissection of their molecular pathophysiology has contributed to our understanding of immune homeostasis in the gut in health and disease.

2 | MONOGENIC FORMS OF INFLAMMATORY BOWEL DISEASE

2.1 | Epithelial barrier defects

The intestinal epithelium forms both a physical and biochemical barrier between gut microbiota and the immune cells within the mucosa. Therefore, dysregulation of the gut epithelium can result in immune overactivation that culminates in bowel inflammation. The onset of IBD can arise through the following mechanisms: (a) defects of epithelial organization, (b) defects leading to epithelial apoptosis and necroptosis, and (c) defects of epithelial-intrinsic cellular function.

TTC7A, a member of TPR domain-containing proteins is thought to have diverse functions in cell cycle control, protein transport, phosphate turnover, and protein trafficking and secretion. Patients with TTC7A deficiency typically present with features of severe combined immunodeficiency (SCID), along with severe exfoliative apoptotic enterocolitis.[30,37,38] In previous studies, mutations in TTC7A were reported to have multiple intestinal atresias (MIA) possibly due to the constant inflammation and apoptosis of the epithelium. It appears that patients with complete loss-of-function typically present with MIA-SCID phenotype, whereas milder (hypomorphic) mutations may present with EO-IBD as a predominant phenotype.[37,38] TTC7A-deficient patient-derived organoids show defective apicobasal polarity and have increased apoptosis that may cause a physical breach of the epithelium therefore aggravating the bowel inflammation.[37] However, the involvement thymic stromal-intrinsic

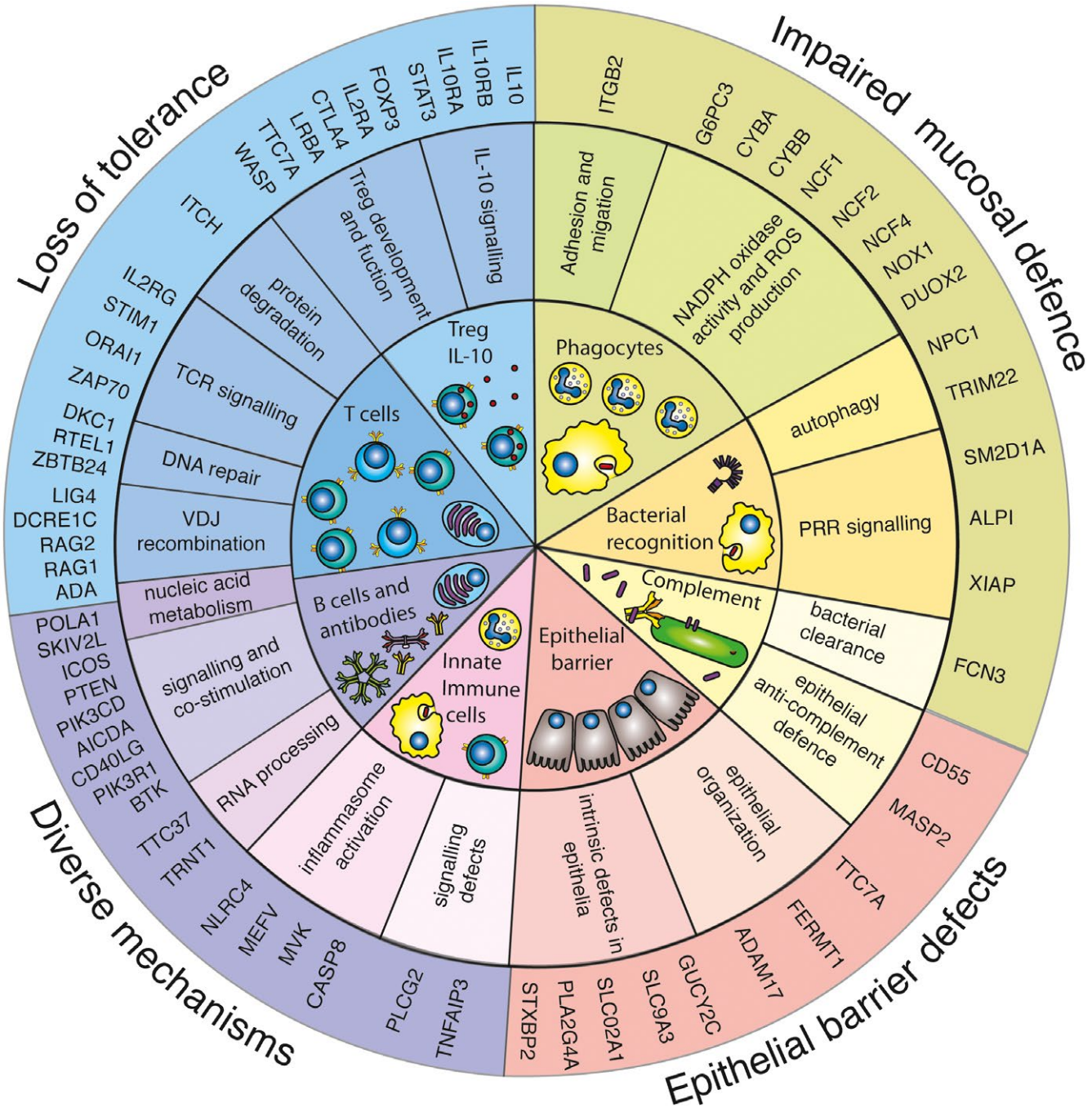


FIGURE 2 Cell types and molecular mechanisms involved in the pathogenesis of inflammatory bowel disease. The inner circle represents cell types and cell components involved in IBD pathogenesis, as detailed in the text. The middle circle depicts the molecular mechanisms affected by mutations in genes presenting with an IBD phenotype. The outer circle represents the molecular pathomechanisms leading to IBD. Treg IL10: T-cell immunodeficiencies with bowel inflammation and Defects in Tregs or IL10 signaling. Phagocytes: Congenital defects of phagocyte number or function. Complement: Complement deficiencies. Bacterial recognition: Defects in host-microbiota interactions, bacterial sensing. Epithelial barrier: Epithelial barrier defects. B cells and antibodies: Predominantly antibody deficiencies with IBD. Innate immune cells: Systemic autoinflammatory diseases and IBD. PRR: pattern-recognition receptor

TTC7A deficiency in the context of T-cell maturation and TTC7A in T-cell intrinsic defect of activation has to be considered as potential cause for bowel inflammation in TTC7A deficiency.

In Kindler syndrome, mutations in *FERMT1* lead to lack of Kindlin 1 and an induction of inflammatory response in keratinocytes via paracrine communication. Kindlin 1 is involved in integrin signaling and the linkage of the actin cytoskeleton to the extracellular matrix. Patients with Kindler syndrome have been reported to have ulcerative colitis,[39-41] and *Fermt1*$^{-/-}$ mouse model shows gut epithelial detachment due to a lack of epithelial integrin activation.[42] This was

hypothesized to cause epithelial barrier breach, which culminated in bowel inflammation in this model. Mutations in the *COL7A1* gene elicit an autoimmune response and autoantibodies to type VII collagen and cause epidermolysis bullosa dystrophica.[43] The mutated COL7A1 leads to a deficiency in anchoring fibrils, which in turn impairs the adherence between the epidermis and the underlying dermis similarly resulting in an impaired gut epithelial barrier.

Mutations in *guanylate cyclase 2C* (*GUCY2C*), an intestinal receptor for bacterial heat-stable enterotoxins cause relatively mild early-onset chronic diarrhea and is associated with increased susceptibility to IBD, small-bowel obstruction, and esophagitis.[44] Although the exact molecular mechanism behind the familial diarrhea is yet to be determined, it has been shown that the expression of mutant GUCY2C results in increased production of cGMP, possibly underlying the hyperactivation of CFTR, leading to increased chloride and water secretion from enterocytes. Missense, splicing, and truncation mutations in *SLC9A3*, identified in nine patients from eight families lead to congenital sodium diarrhea (CSD).[45] Two of these nine patients developed IBD at 4 and 16 years of age.[45] SLC9A3 is an epithelial brush-border Na/H exchanger that uses an inward sodium ion gradient to expel acids from the cell. Several members of the *SLC9A* family of Na^+/H^+ exchangers are expressed in the gut, with varying expression patterns and cellular localization. They participate in the regulation of basic epithelial cell functions, including control of transepithelial Na^+ absorption, intracellular pH, cell volume, and nutrient absorption, and also in cellular proliferation, migration, and apoptosis. In addition, these proteins modulate the extracellular milieu to facilitate other nutrient absorption and to regulate the intestinal microbial microenvironment.[46] The functional consequence of loss-of-function *SLC9A3* gene variants (ie, reduced sodium uptake and proton exchange at the luminal surface) appears similar to that of gain-of-function (GOF) variants in the *GUCY2C* gene, showcasing a potential overlapping molecular mechanism. However, the underlying mechanism of bowel inflammation in these patients is unclear. One potential hypothesis includes physical epithelial damage due to distended bowel resulting in microbiota-mediated immune activation and bowel inflammation.

Loss-of-function (LOF) mutations in the *SLCO2A1* gene, encoding a prostaglandin transporter have been described to cause pediatric-onset chronic nonspecific multiple ulcers of the small intestine, accompanied with persistent blood and protein-losing enteropathy[47,48] in the Japanese population. Mutations in *SLCO2A1* have been previously reported as the cause of primary hypertrophic osteoarthropathy (PHO).[49,50] Three out of five male patients with chronic enteropathy associated with *SLCO2A1* had all of the major clinical features of PHO as well, such as digital clubbing, periostosis, and pachydermia. SLCO2A1, naturally expressed on the cellular membrane of vascular endothelial cells in the small intestinal mucosa, was absent from the patients' epithelium, pointing to a potential epithelial-intrinsic cell defect. Similarly, a LOF mutation in the *PLA2G4A* gene, encoding for cytosolic phospholipase 2-α, has been identified in patients with cryptogenic multifocal ulcerating stenosing enteritis (CMUSE).[51] It was shown that these patients lack protein expression in their gut epithelium. Phospholipase 2-α is an enzyme important in the formation of prostaglandin. Together, these gene defects point toward the role of prostaglandin in gut epithelial homeostasis, specifically in the context of epithelium-intrinsic defects. However, the exact molecular mechanism of prostaglandin-associated enteropathy is still unclear.

Familial hemophagocytic lymphohistiocytosis (FHL) is caused by recessive mutations that impair cytotoxic function and is characterized by fever, splenomegaly, bicytopenia, high triglycerides/low fibrinogen, hemophagocytosis, high ferritin, low natural killer (NK) cell cytotoxicity, and high soluble CD25.[52] FHL type 5 is initiated by mutations in the *STXBP2* (Munc18-2) gene, encoding a protein involved in intracellular trafficking, the control of soluble NSF attachment protein receptor (SNARE) assembly, and the release of cytotoxic granules by NK cells.[53] Notably, MUNC18-2 deficiency (unlike other FHL) is often accompanied by colitis,[53] although GI symptoms are not a common feature of FHL type 2, FHL type 3, or Griscelli Syndrome type 2 patients, suggesting that the pathology of FHL does not necessarily lead to GI disease, even in the more severe FHL subtypes. Munc18-2 proteins have been described to have widespread expression in epithelial tissues, such as the kidney and intestines, with localization to the apical surface of the plasma membrane.[54,55] Thereby, Munc18-2 might be essential for maintaining epithelial integrity in GI epithelial cells, but more mechanistic studies are required to determine the how Munc18-2 deficiency lead to bowel inflammation.

2.2 | Congenital defects of phagocyte number or function

Emerging evidence suggests that neutrophil function plays an important role in intestinal integrity, as highlighted by IBD in patients with either quantitative or qualitative neutrophil deficiencies. Neutrophil function in the gut is not restricted to the killing of bacteria that have translocated across mucosal epithelium. During the inflammatory response, neutrophils also contribute to the recruitment of other immune cells and facilitate mucosal healing by releasing mediators necessary for the resolution of inflammation.[56] Even though our understanding of neutrophils' role in intestinal homeostasis and their complex interactions with intestinal epithelial cells is still incomplete, gut pathologies in patients with neutrophil defects has revealed several important mechanisms.

Neutrophil nicotinamide adenine dinucleotide phosphate oxidase (NOX) is the enzyme complex responsible for generation of superoxide and other reactive oxygen species (ROS) in phagocytic cells. Mutations in the *CYBB* and *CYBA*, *NCF1*, *NFC2*, and *NCF4* genes, encoding for the cytosolic subunits of NOX, abrogate its activity and compromise host immunity against certain bacteria and fungi. These defects cause chronic granulomatous disease which are characterized by immunodeficiency and can cause IBD-like intestinal inflammation.[57] Inflammatory reactions in CGD patients (namely colitis) might be a result of impaired anti-bacterial protection due to impaired NOX activity, resembling defects in ephithelial-specific NADPH Oxidase 1

(NOX1) and Dual Oxidase 2 (DUOX2) in patients with severe EO-IBD. Both NOX1 and DOUX2 are epithelial NADPH oxidases involved in the generation of ROS in the gut epithelium.[58] Mutations in *NOX1* and *DUOX2* result in reduced ROS production and cause a 10-fold increase in bacterial invasion.[59] Impaired mucosal defense may represent a key pathomechanism that results in intestinal inflammation and development of IBD. Another possible pathomechanism leading to colitis in CGD patients is inflammmasome hyperactivation. Intriguingly, NOX-deficient mice exhibited a skewed Th17 phenotype suggesting a possible role of pathogenic Th17 cells in development of inflammatory reactions.[60] These data indicate that while reactive oxygen species are used by the immune system to eliminate infections they may also serve as signaling intermediates to coordinate the efforts of the innate and adaptive immune systems resulting in a complex etiology underlying phagocyte defects.

Although the exact molecular link has not been established yet, it has been shown that mononuclear phagocytes from CGD patients have increased secretion of IL-1β that could be controlled by IL-1 receptor antagonist (IL-1RA) ex vivo and during treatment with anakinra.[61]

Impaired mucosal defense underlying colitis might be one of the pathomechanisms in the other types of neutropenias that result in impaired function or recruitment of neutrophils. Mutations in *G6PC3*, encoding the catalytic subunit of glucose-6-phosphatase (G6Pase) cause severe congenital neutropenia type IV (SCN IV) and predispose patients to IBD.[62-64] SCN IV has been linked to glycogen storage disease type 1b as both disorders involve disruption of the glucose-6-phosphatase/glucose-6-phosphate transporter complex, leading to developmental or functional defects in neutrophils. The function of NADPH oxidase in phagocytes from patients with G6PC3 was diminished, abrogating normal ROS production.[65] These defects suggest loss of protective function perhaps may be the main pathomechanism underlying predisposition to IBD in a subset of G6PC3-mutant patients.

Leukocyte adhesion deficiency type 1 (LAD1) is caused by mutations in the *ITGB2* gene, an integrin participating in cell adhesion and cell surface-mediated signaling. The disease is characterized clinically by delayed umbilical cord separation, recurrent life-threatening infections, impaired pus formation, poor wound healing, and persistent leukocytosis. These clinical features are consequences of defective leukocyte adhesion to endothelial cells, the absence of transmigration into inflamed tissues as well as deficient phagocytosis and chemotaxis of granulocytes, monocytes, and lymphoid cells.[66] Some patients develop an IBD-like phenotype, most likely due to the complex pathology caused by dysregulated recruitment of leukocytes into the intestine that abrogates mucosal defense and regulation of immune response.[67,68]

2.3 | Defects in host-microbiota interactions, bacterial sensing

Nucleotide-binding and oligomerization domain (NOD)-like receptors act as a first line of defense against invading bacteria. Within the NOD family, NOD2 functions as an intracellular sensor for peptidoglycans from the bacterial cell wall. NOD2 has long been studied and is recognized as a critical player in Crohn's disease pathogenesis, where it was shown to regulate innate immunity through NF-κB-induced proinflammatory responses.[12] Intriguingly, single gene defects involving *NOD2* cause Blau syndrome, an inflammatory disorder phenotypically characterized by the triad of granulomatous polyarthritis, dermatitis and uveitis, however without bowel inflammation.[69] In this context, it is postulated, that gene defects that do not directly disrupt NOD2 function, but rather de-regulate proper NOD2 signaling, do present with IBD, whereas at least the Blau syndrome-associated mutations in *NOD2* do not. The very first of discovery relating IBD to defective NOD2 signaling without *NOD2* mutations was XIAP deficiency.[29,70]

X-linked lymphoproliferative (XLP) disease is a rare immunodeficiency caused by mutations in the *SH2D1A/SAP* or *XIAP* genes, respectively. XLP is characterized by severe immune dysregulation that presents with susceptibility to EBV-triggered lymphoproliferative disease (EBV-LPD) or hemophagocytic lymphohistiocytosis (HLH), lymphoma, and dysgammaglobulinemia.[71,72] *SH2D1A* encodes the signaling lymphocyte activation molecule (SLAM)-associated protein (SAP). SAP is involved in the function of cytotoxic lymphocytes and is a key regulator of normal immune function in T and NK cells, as well as the of NK-cell apoptosis.[73-75] Mutations that disrupt the SAP protein impair proper signalling to induce immune response toward viral (EBV) infection and led to the development of lymphomas due to defective lymphocytes apoptosis. Large gene deletions in the *SH2D1A* gene (up to 11 Mb) including those involving the whole gene were identified in 5 families. Three of these larger deletions were associated with GI symptoms of colitis and gastritis.[71] XIAP plays an essential role in the regulation of apoptotic cell death induced by viral infection or an over-production of caspases. In addition to this role, XIAP is also responsible of the regulation of RIPK2, a protein vital in NOD2 signaling. Mutations in *XIAP* cause a unique IEI, similar to X-linked familial hemophagocytic lymphohistiocytosis and X-linked Lymphoproliferative syndromes. Patients with *XIAP* mutations can also develop very early-onset IBD.[71,76] The IBD phenotype in XLP2 is hypothesized to be brought on by abrogated NOD2-mediated signalling and result in innate and adaptive immune defects including granulomatous colitis and perianal disease. Therefore, it is postulated hat colitis may be clinically and pathologically different between XLP1 and XLP2.[77]

Two additional novel gene defects that influence NOD2 signaling and present with bowel inflammation have been described recently. Mutations in the *NPC1* gene, encoding a protein that mediates intracellular cholesterol trafficking of endosomes and lysosomes, cause a neurodegenerative lysosomal storage disease, coupled with fistuling colitis with granuloma formation.[70] The pathogenic mutations in NPC1 is thought to elicit impaired autophagy due to defective autophagosome function. Similar to XIAP deficiency, mutations in NPC1 abolishes NOD2-mediated bacterial handling. However, *NPC1* mutations do not impair RIPK2-XIAP dependent cytokine production.

Identification of patients with homozygous *TRIM22* mutations provided additional links of NOD2 to VEO-IBD. TRIM22 is a ubiquitin ligase that influences NOD2 activity by ubiquitination.[78] Mutations in *TRIM22* disrupt the ability of TRIM22 to regulate NOD2-dependent activation of IFN-β signaling and NFkβ. Intriguingly, LOF variants in NOD2 have been shown to result in the loss of NF-κB-induced proinflammatory cytokine response to muramyl dipeptide (MDP),[12] mirroring the defects observed in patients with *TRIM22* mutation.

Expanding the spectrum of disorders of bacterial sensing underlying bowel inflammation are novel biallelic-inherited LOF mutations in *ALPI*. ALPI is an intestinal alkaline phosphatase that is thought to function in the detoxification of lipopolysaccharide (LPS) and prevention of bacterial translocation in the gut. Mutations in *ALPI* abrogate the regulation of host-microbiota interactions and restrain host inflammatory responses causing early-onset severe diarrhea, weight loss, and severe ulcerations from transverse colon to the rectum.[79]

2.4 | Predominantly antibody deficiencies with IBD

The molecular mechanisms of IBD in patients with defects in humoral immunity are not yet completely understood. Impaired antibody production, especially low IgA, may contribute to the development of gut dysbiosis, but the defects in antibody deficiency alone do not result in intestinal disease. The pathomechanism in this case is most likely due to combined T- and B-cell defects.[80] Some of the predominantly antibody deficiencies that can present with IBD include: (a) selective IgA deficiency with unknown gene defect resulting in defective B-cell maturation into IgA-secreting plasma cells, (b) agammaglobulinemia due to BTK or PIK3R1 deficiency leading to the lack of mature B cells and absent IgM, IgG, and IgA,[80,81] (c) X-linked hyper IgM syndrome due to CD40LG deficiency resulting in defective co-stimulation signaling vital for B-cell proliferation and class-switch,[80] (d) activation-induced cytidine deaminase (*AICDA*) deficiency with abrogated somatic hypermutation, gene conversion, and class-switch recombination of immunoglobulin genes in B cells. Additionally, mutations in *PIK3CD* causing Hyper IgM syndrome (HIGM) result in intrinsic defects in both B and T cells. Clinical heterogeneity in patients with *PIK3CD* GOF mutations correlates with differences in immunological findings and suggests that development of bowel inflammation correlates with more pronounced T- and B-cell defects.[82]

Three novel gene defects associated with impaired humoral immunity and gut abnormalities have recently been described. These gene defects, although all affecting humoral immunity, most likely have distinct mechanisms underlying the observed phenotypes.

Patients with PTEN Hamartoma Tumor Syndrome (PHTS) develop autoimmunity, extensive adenoid lymphoid hyperplasia requiring steroid treatment and adenotomy, thymic hyperplasia, and indeterminate colitis.[83] PTEN is a multifunctional dual phosphatase targeting both lipid and protein targets. It mainly dephosphorylate phosphatidyl inositol-3,4,5-triphosphate (PIP3), an activator of PKB/Akt kinase. Therefore, PTEN is a negative regulator of the PI3K/Akt signaling. Reduced PTEN activity in PHTS affects the homeostasis of germinal centers in B cells by aberrant PI3K/Akt/mTOR pathway thereby disturbing antiapoptotic signals. Patients with heterozygous germline mutations in *PTEN* have been reported to present with B-cells defects, including impaired class-switching, decreased somatic hypermutation frequency and hypogammaglobulinemia.[84] These patients show similarities to patients with GOF mutations in *PIK3CD* where B-cell defects and increased Akt activity can be observed. Differences in clinical presentation such as hemartomas, GI polyps and lipomas not seen in PIK3CD-mutant patients, might be explained by the broader expression pattern of PTEN.

Tricho-hepato-enteric syndrome (THES), also known as syndromic or phenotypic diarrhea, is a congenital enteropathy due to mutations in the *TTC37* gene. Patients with THES present with diarrhea, growth retardation, hair and facial abnormalities, and immunodeficiency. The associated malabsorption leads to malnutrition and failure to thrive. While the exact function of the *TTC37* protein is not known, some studies reported TTC37 as a component of the Ski complex which is crucial for the accurate processing of nuclear RNA precursors and degradation of both cytoplasmic and nuclear RNA.[85] Preliminary studies of brush-border ion transporters in enterocytes from 5 patients demonstrated their reduced expression or mislocalization.[86] While this study suggests that the diarrhea in THES patients might be a result of intrinsic defects in enterocytes, most of the patients also develop humoral immune defects with low protective immunoglobulin (Ig) levels or poor vaccination response. A recent discovery identified a patient with *TTC37* mutation, presenting with immunodeficiency but without diarrhea.[87] While quantitative Ig concentrations were normal, response to pneumococcal vaccination was abnormal with rapid loss of protective titers, pointing to a B-cell defect characteristic for this deficiency. It is unclear why this patient did not develop defects in GI and diarrhea, but these findings may indicate an unexpected genotype-phenotype spectrum in this disease.

TRNT1 enzyme deficiency is a novel metabolic disease caused by defective post-transcriptional modification of mitochondrial and cytosolic transfer RNAs. TRNT1 functions as a CCA-adding enzyme by catalyzing the addition of the conserved nucleotide triplet CCA to the 3′ terminus of tRNA molecules. Mutations in *TRNT1* cause a complex multisystem disease, including B lymphocyte immunodeficiency and infantile-onset cyclical aseptic febrile episodes with vomiting and diarrhea, characterized by global electrolyte imbalance during these episodes.[88] Although the IBD phenotype is currently attributed to intrinsic defects of the gut tissues, whether defects in humoral immunity may contribute to GI inflammation is still to be investigated. Bone marrow transplantation in two patients led to encouraging results, although more long-term data are needed to clarify the disease etiology.

Inducible T-cell costimulatory (ICOS) is an activation-induced member of the CD28 family on T cells. Mutations in the *ICOS* gene cause ICOS deficiency, presenting with common variable immunodeficiency (CVID) including splenomegaly, autoimmune manifestations, recurrent bacterial infections, and IBD.[89] Absence of ICOS results in abrogation of germinal center formation leading to severe reduction of class-switched memory B cells, as well as reduction in naïve B cells. The presumed cause of the IBD phenotype in ICOS deficiency is insufficient IL-10 production by ICOS-deficient T cells.[89]

2.5 | T-cell immunodeficiencies with bowel inflammation

Gene defects that disturb adaptive immune cell selection, activation, and differentiation can all manifest in complex immune signaling disturbances, which can result in immunodeficiency, autoimmunity, and intestinal inflammation. Monogenic gene defects underlying IEI and IBD have been essential in improving our understanding of the complex machinery of immune regulatory cascades and identified novel players in immune processes. SCID denotes a group of disorders of genetic defects that abrogate T-cell development. Mutations in any of the genes that underlie SCID can cause an IBD-like pathology. In particular, hypomorphic mutations where the proteins and/or molecular functions are impaired but residual activity can be observed often lead to IBD. Hypomorphic mutations in SCID-causing genes that affect development of TCR repertoire may allow development of oligoclonal and poorly functioning T cells and are associated with a broad clinical phenotype that may include inflammatory and autoimmune manifestations, including intestinal inflammation.[90,91] Therefore, all partial T-cell defects can potentially be associated with (severe) immune dysregulation and IBD. Here, we discuss a few examples of genetic defects in this group.

Ommen syndrome, namely impaired V(D)J recombination due to mutations in *RAG1* and *RAG2*,[92-94] and defective DNA repair after V(D)J recombination by mutations in *DCLRE1C/ARTEMIS*[95] cause SCID characterized by erythroderma, desquamation, alopecia, eosinophilia, hepatosplenomegaly, elevated serum IgE levels, and often, colitis.[96] Moreover, defects in *DNA ligase 4* (*LIG4*) encoding an ATP-dependent DNA ligase that joins double stranded breaks during non-homologous end joining pathway, and is essential for V(D)J recombination, can cause SCID, and can develop IBD.[92]

Impaired V(D)J recombination results in an emergence of an oligoclonal T-cell repertoire, which indicates that the thymic selection in patients with Omenn syndrome is restricted to the T cell in which recombinase activity is sufficient to generate a functional TCR.[97]

Adenosine deaminase (ADA) deficiency leads to an accumulation of toxic purine degradation by-products, most potently affecting lymphocytes, but other manifestations include skeletal abnormalities, neurodevelopmental affects, and pulmonary manifestations associated with pulmonary-alveolar proteinosis.[98] The major consequences of *ADA* mutations are severe depletion of T and B lymphocytes and NK cells. The underlying mechanisms of this deleterious effect are the increased apoptosis due to the buildup of dATP in cells especially in developing thymocytes and T cells.[99] Although patients present with severe B-lymphocytopaenia and hypogammaglobulinaemia, B-cell development seems to be unaffected.[100]

Interleukin receptor common gamma chain (IL2RG), is a cytokine receptor subunit that is common to the receptor complexes of at least six different interleukin receptors: IL-2, IL-7, IL-9, IL-15, and IL-21.[101] Lack of IL2RG function results in the near-complete absence of T and NK lymphocytes and nonfunctional B lymphocytes, although abrogated γc cytokine-dependent lymphocyte survival. The phenotype presents as SCID with often chronic diarrhea, a phenotype very similar to Omenn syndrome.[102,103]

Combined immunodeficiency due to mutations in *DOCK2*, an activator of Rho GTPases such as RAC1 and RAC2, lead to early-onset invasive bacterial and viral infections, lymphopenia, and various defective T-cell, B-cell, and NK-cell responses. In a large international cohort, we and others showed that one of five unrelated children with defective DOCK2 developed diarrhea. *DOCK2* mutations impaired RAC1 activation in T cells and chemokine-induced migration and actin polymerization in the T cells, B cells, and NK cells. Adding to the cellular phenotype, IFN-α and IFN-λ production by peripheral-blood mononuclear cells was diminished after viral infection.[104] Impaired T-cell activation may account for the immune dysregulation in DOCK2 deficiency, leading to bowel inflammation.

ZAP70 is a membrane protein found on the surface of T and NK cells. It is part of the T-cell receptor signaling cascade, crucial in the context of TCR signaling. ZAP70 deficiency, characterized by CD4 and CD8 T-cell deficiency due to defective T-cell receptor signaling, can present with IBD as well,[105] potentially due to the dysregulation of T cell-mediated immune processes.[105]

ORAI1 and STIM1 form a complex that is vital to maintain cytoplasmic-endoplasmic reticulum calcium homeostasis of cells and is particularly important in the context of Ca^{2+}-dependent T-cell activation.[106] Patients with deficiency in ORAI1 or STIM1 present with variable expression of CID that is characterized by severe T-cell activation defects, with GI manifestations previously reported in ORAI1 deficiency. These findings illustrate that impaired calcium signaling can result in gut inflammation through reduced number of T_{reg} cells and/or aberrant T-cell thymic selection.[107]

Patients suffering from DNA repair defects have been sporadically reported to present with IBD. This is an interesting observation as a previously reported mouse model of IBD arising in knockout of DNA repair genes has been published.[108] However, there is currently insufficient reports of the prevalence of IBD in DNA repair defects. Among these are reports on a patient suffering from Bloom syndrome with ulcerative colitis[109] and VEO-IBD patient with mutation in *ZBTB24*.[110] To date, both these report lack direct conclusion about the molecular mechanism, but highlighted chromosomal stability as one of the influential factors of IBD pathogenesis. One could hypothesize that as DNA repair is important in the context of T- and B-cell maturation through V(D)J recombination, the development of IBD may be pinpointed toward lack of immune regulation.

Defects of telomere maintenance, exemplified by mutation in *DKC1* and *RTEL1* underlie dyskeratosis congenital myelodysplasia which can present with IBD.[111-114] In these cases, manifestation of GI inflammation can be one of the first presenting symptoms as reviewed by Jonassaint et al.[115] They proposed that the onset of GI inflammation is due to defective epithelial barrier function as they found that these patients present with extensive apoptosis in the intestinal mucosa, potentially resulting in the breach of the epithelium and unprecedented activation of the gut immune system. However, it is likely that the T-cell deficiency has an additional pathogenic role in the onset of bowel inflammation in these diseases.

The underlying causes of the systemic autoimmune disease in ITCH deficiency caused by defects in the *ITCH* gene are still elusive. ITCH deficiency is characterized by dysmorphic features, failure to thrive, hepatomegaly, splenomegaly, and delayed motor development,[116] similar to the phenotype of *Itch*$^{-/-}$ mice.[117] To date, two out of ten patients with ITCH deficiency have been described as developing autoimmune enteropathy and chronic diarrhea, with lymphocytic inflammation of the lamina propria.[116] As a ubiquitin ligase, ITCH attaches ubiquitin to substrate proteins and marks them for lysosomal degradation.[118] Ubiquitination is a key component of multiple signaling cascades of the immune system, including TCR downregulation. The exact molecular mechanism behind the systemic autoimmune disease in ITCH deficiency are unclear; however, it might be due to similar mechanics as dysfunction of other E3 ligases Cbl-b and GRAIL, which catalyze the final step of ubiquitin attachment, that can lead to indiscriminate T-cell activation and loss of tolerance to self-antigens.[119,120]

2.6 | Defects in Tregs or IL10 signaling

The discovery of biallelic LOF mutations in the *IL-10 receptor* genes presenting with bowel inflammation as the main phenotype have highlighted the pivotal role of IL-10, and IL-10 in T_{reg} cell function especially in the gut. Defects in the IL-10 receptor genes *IL10RA* and *IL10RB* and *IL-10* itself lead to early-onset enterocolitis involving hyperinflammatory immune responses in the intestine due to abrogated interleukin-10-induced signaling and therefore improper function of regulatory T cells.[26] Similarly, immune defects abrogating proper T_{reg} function can lead to bowel inflammation as well. Immune dysregulation, polyendocrinopathy, and enteropathy (IPEX) is caused by mutations in the *FOXP3* gene, a master regulator of the development and function of Tregs. In IPEX, the lack of or mutant FOXP3 protein causes abnormal T_{reg} function, which causes systemic autoimmunity and severe enteropathy associated with eosinophilic inflammation.[121] Mutations in *CD25* encoding IL2RA, a protein constituting the high affinity IL-2 receptor results in an IPEX-like syndrome. The patients exhibited defective IL-10 expression from CD4 lymphocytes, highlighting the importance of IL-2 in IL-10 production, and the priming of T_{reg} for immunosuppressive functions.[122]

CTLA-4 is an essential effector component of T_{reg} cells that is required for their suppressive function.[123] Therefore, CTLA-4 is a critical inhibitory checkpoint of immune responses. The crucial role of the negative regulation by CLTA-4 is illustrated by the lethal autoimmunity developed by Ctla4-deficient mice.[124] CTLA-4 resides in intracellular vesicles on T_{reg} and is released and mobilized to the cell surface after TCR stimulation, where it works as an "off" switch when bound to either CD80 or CD86 on the surface of antigen-presenting cells.[125,126] CTLA-4 haploinsufficiency or impaired ligand binding results in a complex syndrome presenting with features of both autoimmunity and immunodeficiency.[127]

Patients with CTLA-4 haploinsufficiency develop autoimmune thrombocytopenias and abnormal lymphocytic infiltration of non-lymphoid organs, including the lungs, brain, and GI tract, resulting in enteropathy.[128] CTLA-4 haploinsufficiency has been observed to have incomplete penetrance. However, as the age of studied patients ranges from 7 to 40, currently healthy mutation carriers may develop disease later on in life. Indeed, autoimmune features (psoriasis, type 1 diabetes, and prolonged episodes of diarrhea) are evident in carriers previously classified as healthy. Patients with biallelic mutations in the *LRBA* gene present with a phenotype clinically resembling CHAI disease, but with recessive inheritance.[129] LRBA plays an immunoregulatory role in the expression, function, and trafficking of CTLA-4 from the intracellular vesicles to the cell surface. In fact, patients with *LRBA* mutations show CTLA-4 loss and immune dysregulation[125] and can present with VEO-IBD.[32,130]

A considerable fraction of patients with Wiskott-Aldrich syndrome (WAS) can develop IBD or IBD-like gastroenterocolitis. WASP is expressed in hematopoietic cells and plays essential roles in signal transduction, cell-cell interactions, cell movement, and cell division. The mechanisms driving gut abnormalities in patients with *WAS* mutations most likely have a broader etiology and, like for LAD1 patients, are not restricted to neutrophil defects. Wasp-deficient mice develop chronic colitis associated with colon crypt hyperplasia and the presence of mixed lymphocytic and neutrophilic infiltrate within the lamina propria.[131] Defects in T_{regs} and expansion of autoreactive B cells are likely the main drivers of IBD/IBD-like colitis in WAS patients. Impaired regulatory T cells may also affect the microbiota, leading to dysbiosis that may contribute to colitis development.[132]

The STAT family of transcription factors plays a critical role in mediating responses to cytokines, thereby influencing and initiating cell activation, survival, and proliferation.[133] Autosomal dominant GOF mutations in *STAT3* result in infantile-onset multisystem autoimmune disease. Common manifestations include insulin-dependent diabetes mellitus and autoimmune enteropathy, or celiac disease, and autoimmune hematologic disorders.[134] It is postulated that GOF mutations in *STAT3* lead to autoimmunity, and thereby autoimmune enterocolitis through impairing the development of regulatory T-cells and promoting the expansion and activation of Th17 cells.[135,136]

Lack of T_{regs} in a combined immunodeficiency due to *MALT1* mutations (compound heterozygous splice acceptor and de novo deletion) has been recently described in a male infant who developed generalized rash, intestinal inflammation, and severe infections including persistent cytomegalovirus.[137] MALT1 is a paracaspase with a central role in the activation of lymphocytes and other immune

cells including myeloid cells, mast cells, and NK cells. MALT1 activity is required not only for the immune response, but also for the development of natural T_{reg} cells that keep the immune response in check and is an essential regulator for NF-κB activation.[138] Its inhibition attenuated symptoms of dextran sodium sulfate-induced colitis in mice reducing activation of NF-κB and NLRP3 inflammasome in macrophages.[139] MALT1-deficient patients fail to generate memory and T_{reg} cells and develop hypogammaglobulinemia, due to impaired NF-κB signaling in lymphocytes resulting in immune dysregulation.

Integrity of the TCR/CD3 complex is vital for proper T-cell maturation and function. Mutations in T-cell surface glycoprotein CD3 gamma chain (CD3γ) abrogate the integrity of the complex and result in autoimmunity, accompanied by IBD, due to T-cell phenotypic and functional defects, especially in $T_{reg.}$[140] Therefore, it is postulated that the pathomechanism of IBD in CD3G deficiency stems from the dysregulation due to reduced T_{reg} function.[140]

Mutations in the *IL21* gene, a critical regulator of STAT1, STAT3, and STAT5 signaling[141] cause early-onset IBD and common variable immunodeficiency-like disease.[31] In the context of IL-21 deficiency, the IBD phenotype could be explained by the lack of anti-inflammatory action of I-L21 in inducing IL-10 production through a STAT3-mediated signaling axis. However, this might not be the only mechanism as IL-21R deficient patients have not been reported to develop IBD. More patients need to be identified prior to a conclusive genotype-phenotype correlation.[141]

2.7 | Systemic autoinflammatory diseases and IBD

Systemic autoinflammatory diseases denote a group of immune dysregulatory conditions that usually present in early childhood with fever and disease-specific patterns of inflammation. Studying the gene defects underlying the recurrent inflammatory episodes has revealed key immune pathways underlying persistent inflammation such as excessive IL-1 signaling, constitutive NF-κB activation, and chronic type I IFN signaling.[142] VEO-IBD has been described as an accompanying phenotype in a number of systemic autoinflammatory diseases. Many of the exact causal mechanisms are still postulated, but it is likely that molecular defects underlying IBD in these autoinflammatory conditions disrupt the delicate homeostasis of immune cells, epithelial cells, and the microbiota in the gut by chronically activating proinflammatory pathways and cell types. The importance of such intrinsic innate signaling systems such as IL-1β signaling in the pathogenesis of IBD is illustrated by the fact that inhibiting IL-1β signaling can induce complete or partial elevation of symptoms in patients, including the remission of the VEO-IBD phenotype.[143,144]

Mevalonate kinase deficiency due to pathogenic mutations in the *MVK* gene presents with hyper IgD syndrome (HIDS), as well as polyarthralgia or nonerosive arthritis of large joints, cervical lymphadenopathy, abdominal pain, vomiting, diarrhea, and variable skin lesions, including maculopapular, urticarial, nodular, and purpuric rashes.[143] LOF mutations in *MVK*, encoding a key enzyme in the cholesterol synthesis pathway, impair the enzymatic activity and lead to a shortage of farnesyl pyrophosphate and geranylgeranyl pyrophosphate, intermediates for isoprenoid synthesis and substrates used for protein prenylation.[145,146] Flares in HIDS are thought to be the result of uncontrolled release of IL-1β as a consequence of insufficient geranylgeranyl pyrophosphate generation.[147]

PLCG2 encodes phospholipase Cγ2 (PLCγ2), an enzyme responsible for ligand-mediated signaling in cells of the hematopoietic system through IP3 and DAG, and plays a key role in the regulation of immune responses. Patients with GOF mutations in *PLCG2* develop autoinflammation and PLCγ2-associated antibody deficiency and immune dysregulation (APLAID). APLAID presents with recurrent blistering skin lesions, bronchiolitis, arthralgia, ocular inflammation, enterocolitis, absence of autoantibodies, and mild immunodeficiency, with a decrease in circulating IgM and IgA antibodies, decreased numbers of class-switched memory B cells, and decreased numbers of Natural Killer T (NKT) cells.[148] The phenotype in APLAID is thought to be the consequence of the GOF mutations that create an extra phosphorylation site which enhances activation *PLCG2* by compromised (although not completely abrogated) autoinhibition of PLCG2 activity. Intriguingly, *PLCG2* genomic deletions in individuals present with a distinct inflammatory disease manifested by cold-induced urticaria and immune dysregulation including features of both immunodeficiency and autoimmunity, called PLAID. The PLAID-associated genomic deletions disrupt the cSH2 domain of PLCγ2, resulting in constitutive phospholipase activity. Despite the constitutively active enzymatic activity, PLAID patients have reduced PLCγ2-mediated signal transduction at physiologic temperatures most likely as a result of a negative feedback caused by constitutive activation.

Mutations in two of the genes encoding for the inflammasome components NLRC4 and MEFV can cause monogenic autoinflammatory diseases that can present with IBD. Recessive and postulated autosomal dominant mutations in *MEFV*, a gene encoding the intracellular sensor pyrin/marenostrin, cause familial Mediterranean fever (FMF). FMF flares include fever, generalized peritonitis, and less frequently nonerosive oligoarthritis, and can include colitis.[149,150] MEFV has been implicated in multiple cellular and vital immune functions such as the assembly, intracellular danger sensing and induction of inflammation by the inflammasome, intracellular danger signal sensing, apoptosis, and autophagy in granulocytes and monocytes.[151] Although the concrete link between the FMF phenotype with colitis and MEFV is still to be understood, it is clear that mutations in *MEFV* result in the enhanced and extended inflammatory response to some of the innocuous factors that are tolerated well and handled efficiently by the normal immune system. Activating heterozygous mutations in *NLRC4* have been reported to cause recurrent fevers and severe systemic inflammation, similar to macrophage activation syndrome (MAS). To date, three of 4 reported patients developed enterocolitis.[152,153] NLRC4, a member of cytoplasmic NOD-like receptors, is involved in detection of pathogen-associated molecular patterns and initiate inflammatory responses by recruiting and proteolytically activating caspase-1 within the inflammasome upon stimulation. Mutant NLRC4 causes constitutive IL-1 and IL-18 family cytokine production, macrophage activation, and increased cell

death. Patient macrophages are polarized toward pyroptosis and exhibit abnormal staining for inflammasome components.

Heterozygous germline mutations in *TNFAIP3* cause a Behçet's-like disease, characterized by early-onset systemic inflammation, arthralgia/arthritis, oral/genital ulcers, and ocular inflammation described in six unrelated families.[154] *TNFAIP3* encodes the NF-κB regulatory protein A20 which is a potent inhibitor of the NF-κB signaling pathway via its deubiquitinase activity. *TNFAIP3* mutant patient-derived lymphocytes show increased degradation of IκBα and nuclear translocation of the NF-κB p65 subunit, together with increased expression of NF-κB-mediated proinflammatory cytokines. In these lymphocytes, TNF stimulation leads to defective removal of Lys63-linked ubiquitin from TRAF6, NEMO, and RIP1.[154]

LOF mutations in the gene encoding CASP8, a protease that initiates apoptosis and regulates immune responses have been described very recently to cause infant-onset IBD.[155] Previously, patients with *CASP8* mutations have been shown to present with autoimmune lymphoproliferative syndrome-like (ALPS) like disorder.[156] In contrast, the novel report shows patients with previously undocumented mutations in *CASP8* presenting with severe VEO-IBD as the main clinical manifestation. The patient lymphocytes exhibited defective T- and B-cell maturation proliferation and activation, as well as impaired inflammasome activation and defective epithelial cell death responses. These findings highlight the critical role of CASP8 in non-apoptotic functions, especially in maintaining intestinal immune homeostasis.

Abnormal nucleic acids generated during viral replication is one of the main triggers for antiviral immunity. Concomitantly, mutations disrupting nucleic acid metabolism can lead to autoinflammatory disorders. SKIV2L is an RNA helicase and is an important negative regulator of the RIG-I-like receptor (RLR)-mediated antiviral response. Mutations in *SKIV2L* cause THES, characterized by chronic diarrhea, liver disease, hair abnormalities, and high mortality in early childhood due to severe infection or liver cirrhosis.[157,158] It has been shown that the unfolded protein response (UPR), which generates endogenous RLR ligands through IRE-1 endonuclease cleavage of cellular RNAs, triggers type I interferon (IFN) production in SKIV2L-depleted cells.[159] Intriguingly, THES can be caused by mutations in *TTC37*[86,87] where, in contrast to SKIV2L, in vitro assays do not propose a role in interferon signaling. This suggests that most of the features of THES are most likely the consequence of a loss of cytosolic RNA exosome function in RNA turnover, instead of aberrant interferon response that is apparently specific to SKIV2L deficiency.

Intronic mutations in *DNA Polymerase Alpha 1* (*POLA1*) cause X-linker reticulate pigmentary disorder including early-onset IBD. *POLA1* encodes the catalytic subunit of DNA polymerase and is vital component of the DNA replication machinery. The polymerase A complex synthesizes RNA:DNA primers which initiate the production of Okazaki fragments. Mutations in *POLA1* affect the expression of DNA polymerase-α, leading to aberrant synthesis of RNA:DNA primers in cells, thereby inducing type 1 IFN.[160,161]

Patients with mutations in *ADAM17* present with early-onset pustular dermatitis, short and broken hair, paronychia, frequent cutaneous bacterial infections, cardiomyopathy, and early-onset diarrhea.[162] In a study of two related patients, patient-derived PBMCs showed high levels of lipopolysaccharide-induced production of interleukin-1β and interleukin-6 but impaired release of TNF-α.[162] ADAM17 plays a role in the processing of other cell surface proteins, including a TNF receptor, the L-selectin adhesion molecule, and transforming growth factor-alpha (TGF-α).[163] Although direct links between the patient's phenotype and ADAM17 defects is still elusive, lack of TNF-α is considered partly responsible for the increased susceptibility to infection and development of cardiomyopathy, and as *Adam17* knockout mice present with impaired epithelial cell maturation in multiple organs, the lack of proper epithelial barrier could be postulated to stem the IBD phenotype.

2.8 | Complement deficiencies

The complement system is made up of a large number of distinct plasma proteins and autologous cell surface proteins that react with one another to mainly opsonize pathogens and induce a series of inflammatory responses, initiating the adaptive inflammatory response. Deficiencies in complement proteins mostly manifest as recurrent bacterial infections due to defective bacterial clearance and autoimmunity such as systemic lupus erythematosus. However, multiple cases of complement deficiency presenting with IBD or IBD-like symptoms have been sporadically reported,[164] pointing to a possible role of complement pathway in IBD pathogenesis. The potential pathomechanism of IBD pathogenesis in complement deficiencies has been hypothetically directed toward defective bacterial clearance and potential defective epithelial defense against complement attack. In this case, the interplay between the microbiota and immune system is further highlighted as the manifestation of bowel inflammation is present in all patients.

The identification of MASP2 deficiency highlighted the potentially vital role of proper activation of the complement system in colitis.[165] In one patient, homozygous mutation in the *MASP2* gene caused defective activation of the complement system through the mannan-binding lectin (MBL) pathway, and resulted in a presentation of ulcerative colitis and later on erythema multiforme bullosum. Numerous polymorphisms in *MASP2* that causes lack of MBL pathway activation have been identified,[166] but no further reports of IBD have been described. Therefore, MASP2 might be a modulator of IBD pathogenesis and that requires further triggers to result in an IBD presentation.

Ficolin 3 deficiency was first reported in a patient with immunodeficiency and recurrent infections, clinical manifestations that are in line with complementopathies. In a report by Shlapbach et al[167], 2 patients with congenital FCN3 deficiency suffered from severe, potentially fatal necrotizing enterocolitis that they postulate was due to defective control of intestinal microbiota leading to local inflammation.

In 2017, we and others have identified biallelic LOF mutations in *CD55* encoding for the protein decay accelerating factor (DAF) in patients with severe early-onset protein-losing enteropathy.[168] CD55

is a complement regulatory binding protein present on autologous cells that acts to prevent the activation of the complement cascade on cell surfaces. It does so by binding to C3b and C4b, two complement convertases and silences their activity. To date, a total of 18 patients have been described in the literature to have mutations in *CD55* affected with protein-losing enteropathy and of these, 6 develop bowel inflammation with histologically proven lymphocytic infiltrates in the mucosa or mucosal ulcers. However, the extent of the inflammation is not as severe as in other EO-IBD patients. The origin of the inflammation is still unclear, but we propose 2 potential pathomechanisms: dysregulation of immunoregulatory T cells, similar to observation made in mouse models on the role of CD55 on T_{reg} homeostasis,[169] and epithelial and/or endothelial barrier damage due to complement activation. Interestingly, some patients develop thrombotic events, a clinical manifestation of many complementopathies. Patients responded well to the eculizumab treatment,[170] with immediate effects seen in the GI protein loss clinical manifestation. However, more data need to be obtained to see if eculizumab proves to be efficacious in relieving bowel inflammation in CD55-deficient patients.

2.9 | Other gene defects

IBD or an IBD-like phenotype have been described in diseases with no well-defined plausible mechanisms, or in diseases where well-defined molecular mechanisms exist but the underlying cause of IBD is still elusive.

Defects in HPS1, HPS4, HPS6 genes that underlie Hermansky-Pudlak syndrome (HPS), can present with colitis.[171,172] Patients presents with the triad of oculocutaneous tyrosinase-positive albinism, prolonged bleeding time secondary to platelet storage pool defect and ceroid depositions within the reticuloendothelial system. Reportedly, some patients develop GI complications related to chronic granulomatous colitis, enterocolitis, and extensive granulomatous perianal disease. Although some evidence suggests that an abnormality of lysosomal function may be responsible for the development of the disease, the underlying molecular mechanisms are still unclear. More intriguingly, mutations in *HPS3*, *HPS5*, and *HPS7* cause HPS that do not present with IBD.

PEPD encodes a member of the peptidase family with an important role in recycling of proline and might be rate limiting for the production of collagen.[173] Individuals with mutations in PEPD develop prolidase deficiency, characterized by lack of peptidase activity, skin ulcers, mental retardation, and recurrent infections. Patients may have splenomegaly, and in some cases, hepatosplenomegaly. Diarrhea, vomiting, and dehydration may also occur.[174,175] Pathogenic mutations in *PEPD* lead to reduction or loss of prolidase activity which may contribute to the multifactorial clinical presentation. Since phenotype, age of onset, and clinical course of prolidase deficiency are very variable even within the same family, and the number of molecularly characterized patients is very small, it is still difficult to define a genotype-phenotype relationship for this disease.[173]

Complex dysregulation of transforming growth factor beta as a result of autosomal dominant mutations in *TGFBR1* and *TGFBR2* (Loeys-Dietz syndrome) cause a syndrome with a variety of phenotypes including skeletal involvement, arterial abnormalities and immunological abnormalities, IBD, and encelopathy.[176] Recently, biallelic LOF mutations in the *TGFB1* gene encoding TGF-β1 have been described in patients with central nervous system disease including epilepsy, brain atrophy, and posterior leukoencephalopathy, and severe VEO-IBD.[177] The mutations in *TGFB1* seemingly impaired the bioavailability of TGF-β1. Although the exact mechanisms of how impaired TGF-β signaling leads to IBD is yet to be determined, these findings suggest a pivotal role in of TGF-β immune function, especially in intestinal immune homeostasis.

Defective adaptation to hyperosmotic stress in lymphocytes recently emerged as one of the novel mechanisms underlying IEI and IBD. A single male with de novo Nuclear Factor of Activated T Cells 5 (NFAT5) haploinsufficiency presented with autoimmune enterocolopathy, unexplained infections, and bowel inflammation. Further examination revealed IgG subclass deficiency, impaired antigen-induced lymphocyte proliferation, reduced cytokine production by $CD8^+$ T lymphocytes, and decreased numbers of NK cells.[178] NFAT5 is a transcription factor protein that is activated in response to osmotic stress. In NFAT5-deficient patients, regulation of immune cell function and cellular adaptation to hyperosmotic stress is abrogated, leading to the phenotype.

Dysregulation of mitochondrial integrity and increase in cellular stress have been recently identified as a cause of severe T-, B-, and NK-cell lymphopenia presenting with VEO-IBD. Two patients, one with homozygous and one with compound heterozygous mutations in ankyrin repeat and zinc-finger domain-containing 1 (*ANKZF1*) developed severe bowel inflammation, severe ulcerative skin lesions, and T-, B-, and NK-cell lymphopenia. The suspected causal gene, *ANKZF1* has a role in mitochondrial response to cellular stress. As a consequence of mutations in *ANKZF1*, mitochondrial respiration is impaired resulting in increased apoptosis in patient lymphocytes.[179]

3 | GENOMICS AND ITS INFLUENCE ON THERAPEUTIC GUIDELINES FOR VEO-IBD PATIENTS

VEO-IBD patients make interesting clinical cases as this group of rare diseases often comes without a clear-cut clinical decision-making scheme as they often present with multi-organ involvement that requires intervention from different clinicians. Treatment of VEO-IBD patients does not differ from adult-onset IBD patients in principle, in that the end result is to induce and maintain remission. These patients receive a standard care therapy, which frequently involves a combination aminosalicylates, corticosteroids, immunomodulators, antibiotics, and/or biologics. These medications aim to control intestinal inflammation by dampening the immune system. However, due to the heterogeneous clinical response of VEO-IBD patients to

immunomodulatory drugs, it is often difficult to prescribe a clinical guideline for treatment.

In the more severe cases, bowel resections may be performed to reduce inflammatory regions in the GI tract.

The identification of underlying genetic causes of the disease can highly influence the clinical decision making for patients with a mutation in known disease-causing genes. For instance, hematopoetic stem cell transplantation is currently the only curative therapy for patients with IL-10R deficiency[26] and has been shown to result in a positive clinical outcome in some patients with LRBA deficiency.[180] Treatment of CTLA-4 haploinsufficiency and LRBA are prime examples of genome-informed precision medicine, where treatment with Abatacept (CTLA-4-Ig) has proved to be successful in alleviating the infiltrative and autoimmune disease.[125,181]

In the case of a genetic mutation in a gene that affects both the immune and epithelial barrier (for example TTC7A deficiency), HSCT did not correct for the epithelial-intrinsic defect and enteral tolerance.[182] This further highlights the importance of identifying underlying genetic cause of VEO-IBD to reduce treatment-related mortality. More research needs to be performed in order to elucidate the roles of gene defects in cell types in which they were not implicated before.

4 | BOWEL INFLAMMATION AND THE MICROBIOME

While the link between gut inflammation and gut dysbiosis is not a novel concept, the development of culture-independent techniques like next-generation sequencing and metagenomics exploded the field of microbiome-related studies. These techniques enabled the global assessment of the gut microbiota more accurately and in a more sophisticated manner.[183,184] The largest and perhaps the most ambitious initiative that has emerged in the last decade to study the changes of the human microbiome in health and disease is the NIH sponsored Human Microbiome Project (HMP).[185] It has resulted in the publication of 5177 microbial taxonomic profiles from a population of 242 healthy adults and serves as a comprehensive database for research in this field.[186] This project was followed up by the second phase that, in addition to phylogenic composition, aimed to analyze functional omic data including transcriptome, proteome, and metabolome. Such multi-omic approaches with simultaneous analysis of host and microbiome proteins and metabolites aimed to better our understanding of the biology of the microbiome and sophisticated molecular mechanisms of host-microbiota interaction.[187] Such integrative analysis is the key feature of the future microbiome research.[188,189]

While microbiota from some body sites (for example skin) is easily accessible, the GI tract is much more challenging to sample and describe. The complex structural and functional features of the human GI tract is reflected by the differences in abundance and composition of bacteria and their dynamic variations along the intestine make human microbiome studies complex.[190] The excitement in studying the gut microbiome is not only driven by the fact that it is perhaps the most abundant and complex microbial community of the human body, but that it has also been associated with the development of wide spectrum of diseases. Indeed, numerous studies, including those that use integrative analysis of human gut microbiome and metabolome, have associated the gut microbiota with the promotion of health and development of IBD, obesity-related inflammatory disorders, allergic diseases, and infectious diseases.[191] Although the correlation between gut dysbiosis and IBD is well appreciated, the role of microbiome perturbations in disease development is not yet clearly defined.[192]

The role of the immune system in the preservation of healthy gut microflora is highlighted by the studies of IEI, showing that diverse pathomechanisms may underlie development of gut inflammation in immunocompromised patients. Studies of both adult and pediatric IBD showed decreased diversity of microflora in patients with CD and UC, increased numbers of mucosa-associated aerobic and facultative-anaerobic bacteria in colonic biopsies and perturbations in two most abundant fila—Firmicutes and Bacteroidetes.[187,193-196] While microbiome perturbations in IBD can have a complex etiology, dysbiosis in patients with VEO-IBD or IEI is driven primarily by the gene defects. A study of the gut microbiome in CVID patients showed significant differences in bacterial composition with dysbiosis and low alpha diversity characteristic of the patients with IBD.[197] Interestingly, while elevated dysbiosis index was a characteristic of patients "with infections only" and "with complications" subgroups, the latter had also reduced alpha diversity of the gut microbiota. Patients with enteropathy within the "with complications" subgroup did not show any significant differences in gut microbiota. The lack of obvious differences in microbiomes between patients with or without gut pathologies in this study is difficult to explain. No systematic studies to date involving genetically characterized VEO-IBD and IEI in patients have been conducted, and it is not clear whether these patients might develop gene defect-specific perturbations in the gut microbiome. Given the diversity of molecular pathomechanisms underlying IBD in patients with immune defects, one could speculate that their effect on the microflora might be quite different.

To date, a variety of mechanisms explaining how changes in gut microflora may impact host immune system have been described. This topic has had substantial advancements, highlighting some exciting examples of bacteria-derived metabolite involvement. Playing a pivotal role in maintaining organismal homeostasis and stable physiology, microbiota produce, degrade, and modulate a large number of small molecules—metabolites, complementing the host metabolic capacities. Another important function of this bacteria-modified metabolic network is communication with the host.[198] Even in a healthy state of intact gut epithelial integrity, many bacterial metabolites are absorbed, drain into the portal vein and can be detected in the periphery if they are not metabolized in the liver. Three main mechanisms of how bacterial metabolites impact the immune system have been described: (a) through binding to the specific cell surface receptors, (b) inflammasome-forming intracellular receptors, and (c) antigen presentation.

Short-chain fatty acids (SCFAs), tryptophan metabolites, and retinoic acid (RA) are the most illustrious examples of metabolites that are involved in various aspects of immune cell regulation, development, and differentiation of activation-specific G-protein-coupled-receptors. Downstream signaling through these receptors is responsible for T_{reg} expansion and differentiation, decrease of proinflammatory Th17 cells, changes in neutrophil, and lymphocytes chemotaxis, and hematopoiesis of dendritic cells from bone marrow. SCFA such as butyrate and propionate are known to act as histone deacetylase (HDAC) inhibitors. Butyrate suppresses proinflammatory effectors in lamina propria macrophages and differentiation of dendritic cells from bone marrow stem cells via HDAC inhibition, resulting in hyporesponsivness to commensals. In addition, SCFAs also regulate cytokine expression in T cells and generation of regulatory T_{reg} through HDAC inhibition.[199]

Some commensal microorganisms like *Lactobacilli* use tryptophan as an energy source to produce ligands of the aryl hydrocarbon receptor (AhR), such as the metabolite indole-3-aldehyde. AhR is a ligand-activated transcription factor critically important to the organogenesis of intestinal lymphoid follicles (ILFs). AhR is also expressed by immune cells, including RORγt$^+$ group 3 innate lymphoid cells (ILC3s) that are involved in ILF genesis, and AhR expression on ILC3s is functionally required for their expansion. AhR-induced IL-22 production by ILCs drives the secretion of the anti-microbial peptides lipocalin-2, S100A8, and S100A9, which protect from pathogenic infection by *Candida albicans*. In addition to its role in the function of ILCs, AhR was also found to be necessary for the maintenance of the epithelial barrier and the homeostasis of intraepithelial lymphocytes (IELs).[198]

Retinoic acid (RA) signaling has been shown to be important in the myeloid compartment. Specific subsets of intestinal DCs and macrophages constitutively produce RA and induce T_{reg} development through RA receptors. In addition, signaling downstream RA receptors induce expression of gut homing receptors on activated T and B cells and enhanced induction of immunoglobulin A (IgA) by B cells.[200]

The role of vitamins in maintaining T_{regs}, as well as a number of lymphocytes and NK-cell activity, has also been established.[198] In this case, the effect is mediated through specific receptors broadly expressed on various subsets of immune cells.

The modulation of inflammasome signaling by bacteria-derived metabolites is another distinct mechanism involved in modulation of host immunity. Recent studies implicated several low-molecular-weight compounds associated with metabolism, not immunity, in regulation of NLRP3 and NLRP6 activation.[201] In a recent study, it has been shown that microbial metabolites taurine, histamine, and spermine modulate NLRP6 inflammasome signaling, secretion of IL-18, and production of anti-microbial peptides shaping the host-microbiome interface.[202]

The discovery of bacteria-specific vitamin B metabolites recognized as antigens by mucosa-associated invariant T (MAIT) cells revealed yet another mechanism of host-microbiome interaction and provides an important hint as to how our immune system may sense and control the microbiome.[203] Protective role of MAIT cells upon bacterial infection and their role in autoimmune diseases such as multiple sclerosis and IBD makes these cell attractive targets for clinical interventions. Despite a huge interest in these unique T-cell subsets, their role in disease pathogenesis is still not clear complicating their therapeutic implementation. In addition, only few bacteria-derived molecules have been identified to date with agonistic or antagonistic effect on MAIT cells. Interestingly, a novel heterogeneous population of T cells has been recently identified. These cells recognize endogenous metabolites of unknown structure presented by MHC class I-related molecule 1 (MR1) the same molecule that present bacterial metabolites to MAIT cells.[204] The spectrum of stimulatory antigens and molecular mechanism of antigen presentation to MAIT and other MR1-restricted T cells are still the subject of active research.

Increasing numbers of studies with mouse models of colitis show protective effect of microbiota transfer. Fecal microbiota transplantation has been described as safe and promising treatment for IBD, with unexplained variable efficacy.[205]

Studies have shown that colitis in Nlrp12-deficient mice can be reversed equally by treatment with antibodies targeting inflammatory cytokines and by the administration of beneficial commensal isolates. Such contributions of the microbiome to the development of gut inflammation reveal a feed-forward loop in which a genetic defect promotes dysbiosis that further contributes to the development of gut inflammation.[206] Studies in Il10 and Nlrp6-deficient mice show that adaptive and innate immunity defects may have different contributions to the development of spontaneous colitis, inflammation, and colonization by a specific pathobionts.[207] Overall, studies of gut microbiome in respect to metabolite composition and dynamics provide a basis for the targeted metabolomic intervention and treatment or prevention of dysbiosis-driven diseases. This approach is exemplified by the well-documented beneficial effects of short-chain fatty acid butyrate, synthesized from non-absorbed carbohydrate by gut microbiota. A variety of approaches, including high-fiber diet, butyrate-producing bacteria, or coated tablets are currently in use for the butyrate-based treatment of IBD. Classical pro- and prebiotic-based therapies exhibit limited efficacy due to certain caveats such as colonization resistance and inter-individual variation in microbial composition. Recently, a novel personalized therapeutic approach based on supplementation of the host with metabolites downstream of the microbiome which can act directly on host-related metabolic pathways has been suggested.[208] The nature and efficacy of such potentially bioactive metabolites to be used for the therapy require further exploration.

5 | ORGANOIDS

Although single gene defects affecting major immune pathways have been investigated in detail, little is known about how mutations of VEO-IBD-associated genes are involved in epithelial barrier function or the homeostasis of the host and microbiome. As the interplay

between immune cells, gut epithelial barriers, and the gut microbiota represents a central axis in the onset of VEO-IBD, it is necessary to study the diverse genetic influences of EO-IBD-associated genes in these three players. Gut-derived organoid technology has been at the forefront of advancing our understanding of gut homeostasis, in particular in studying the elusive biology of the gut epithelium. This technology has allowed us to take a reductionist approach in studying specifically gut epithelial derived from any area of the gut and has been shown to be crucial to further our understanding of the biology of some VEO-IBD genes. For example, the role of TTC7A in controlling the polarity of the gut epithelium has been shown in organoids derived from patients with TTC7A deficiency.[37] Patient-derived organoids provide a potential wealth of resources for personalized medicine as it allows us to perform drug screens in the context of patients' genetic background. This was shown in the case of patients with different mutations in the *CFTR* gene, where different genetic backgrounds show varying responses to clinically available drugs, and that the phenotype seen in the organoids correlates with patients' clinical course on various drugs.[209] In the future, generation of biobanks with patient-derived gut organoids with genetically defined backgrounds will allow us to advance personalized medicine for this heterogenous group of rare diseases.

6 | THE FUTURE OF IBD GENETICS

Next-generation sequencing (NGS) has become widely used since 2008 to investigate the genetics of IBD. Most studies that aim to elucidate the genetic component of adult IBD focus on GWAS. However, despite all efforts and recent advances, the genotype-molecular mechanism-phenotype link is still missing for many of the significant GWAS loci. It has been observed that although one expects GWAS signals to cluster in disease relevant pathways and genes, association signals for complex traits tend to be spread across most of the genome including in the vicinity of numerous genes without a clear connection to disease. An "omnigenic" view of diseases proposes that in contrast to Mendelian diseases, which are often caused by high impact mutations in protein-coding regions in a few genes, complex traits and diseases are mainly driven by lower impact variants that affect a multitude of genes and pathways often outside of the primary pathways and genes involved in the respective diseases. Given the interconnected nature of cellular systems, these lower impact variations in diverse pathways converge and create the disease phenotype.[210] The assessment of such combinatorial effect requires complex models of molecular networks and integration of multiple datasets.

7 | WES, PANELS, AND WGS

High-coverage exome or targeted exome (also known as panel) sequencing are routinely used to identify genetic aberrations causing early-onset bowel inflammation. Due to its high accuracy and moderate costs, panel sequencing is preferentially used as a screening method in many institutes. A combinatorial method is to use exome or genome sequencing, but prioritize "virtual panels"—a selected list of genes—for an initial screening and later extend the scope of the analysis to novel genes. When relying on a targeted panel approach, the design of the targeted panel is a crucial step toward appropriate gene discovery and diagnosis. On one hand, by building the panel by using only genes whose link to the disease phenotype is clear, one might miss important new genes and diagnosis. Conversely, by including candidate genes (whose association to the phenotype is yet unknown), there is an increased chance of identifying variants of unknown significance. While WES conveys an advantage of looking at all known coding sequence and can be combined with a panel approach, incidental findings and variants of unknown significance are much higher.

The percentage of patients with genetically diagnosed VEO-IBD varies between centers and cohorts and ranges from 5%-31% depending the composition of the cohort.[211-216] The reasons behind the missing heritability are multifactorial. Firstly, we can hypothesize that not all VEO-IBD patients have a monogenic defect, but rather develop bowel inflammation a result of multiple genetic aberrations. In these cases, proving causality is a challenge. Secondly, inherent technical difficulties of exome sequencing, namely the challenge to detect structural variants such as insertions and deletions, copy number variations, inversions, and deletions could result in missing crucial genetic diagnoses. Naturally, as exome sequencing excludes all but coding sequences from the scope of analysis, genetic aberrations in critical regulatory regions are not detected.

Whole genome sequencing promises an unprecedented depth of study, including non-coding regions and the analysis of copy number variations, including deletions.[217,218] Moreover, there is increasing evidence that genome sequencing might be more powerful at detecting exonic variants than whole exome sequencing.[219] An example of applying low coverage exome sequencing to find an associative gene is the identification of ADC7Y[220] in a study of 4280 patients at low coverage.

RNA sequencing can be used to complement genomic technologies and is routinely used for muscular disorders where it has significantly improved the rate of successful genetic diagnosis.[221] RNA sequencing can be used to investigate transcriptomic differences between IBD and the general population, investigating alternate transcripts arising from alternative splicing and variants affecting mRNA abundance.[222]

8 | VARIANT PRIORITIZATION AND INTERPRETATION

Along with the advent of WES and WGS, accurate variant prioritization has become even more critical to successful genetic diagnosis. WES and WGS can unravel multiple potentially causal variants and variants of unknown significance. An average exome typically contains around 30 000 variants compared to the reference genome

and around 20 genes that are completely inactivated.[223] In line with this observation, analysis of WES data focusing on rare, missense variants, and frameshift insertions and deletions often yields several (typically 20-50) rare variants of unknown significance in a single patient. This poses a tremendous challenge to uniquely identify the causative gene variant. In the quest for the causative variants, variant interpretation, a process of connecting individual variants to disease phenotypes is essential to both reporting results to clinicians and patients and is also crucial to novel variant discovery and downstream functional research. When working under the hypothesis of a Mendelian disease, the expected causal mutations are both rare in the healthy population and severe enough to abrogate protein expression, function, or hinder initiation of vital signaling cascades.

Population-scale variant resources such as Exome Aggregation Consortium (ExAC),[224] the genome Aggregation Database (gnoMAD),[224] and the 1000 Genomes project[225] allow for filtering against population allele frequencies. Some caution should be used when relying these resources as they can contain data from yet unknown patients with various symptoms, potentially with symptoms similar to the disease of interest.

Pathogenicity prediction tools are increasingly used to assess the deleteriousness of protein-coding variants. Some examples include protein-based metrics such as Polymorphism Phenotyping 2 (PolyPhen2),[226] Sorting Intolerant From Tolerant (SIFT),[227] conservation-based tools like Genomic Evolutionary Rate Profiling (GEPR),[228] and integrative methods such as Combined Annotation Dependent Depletion (CADD).[229] These tools aim to decipher the consequence of genetic variants on protein structure and function. However, these algorithms tend to exploit a single information type (conservation for example) or are restricted in scope (focusing only on missense changes, disregarding deletions and insertions). Even when different information types are used in combination (like in CADD), a lack of information is available to assess the widespread signaling and cellular perturbations of genetic aberrations and these tools are often unable to assess the effect of non-coding, regulatory variants. Resources such as the ENCODE have facilitated the understanding of the functional and regulatory elements in the human genome, but the interpretation of non-coding variants remains a challenge.[230]

Although filtering according to allele frequency and pathogenicity prediction provides valuable information on potential causal variants, it has become clear that the detailed characterization of the individual molecular components alone (genes, proteins, metabolites, etc.) does not suffice to truly understand the nature of (patho-) physiological states and how to modulate them. Biomolecules do not act in isolation, but rather within an intricate and tightly coordinated machinery of complex interactions, such as protein-protein, gene regulatory, or signaling interactions. Drawing parallels, disease phenotypes are results of a complex interplay between multiple factors. Systems biology-based approaches have been increasingly developed and used to combine diverse types of information to assess potential causal variants and prioritize candidate variants and candidate genes. Currently, many prioritization methods and online interpretation tools exist with different approaches to data interpretation and analysis. These approaches include algorithms that use protein-protein interactions (PPI) networks,[231,232] functional similarity networks built utilizing pathway involvement,[233] structural variation data,[234] exploit thorough functional annotation,[235] or a combination of PPI based approach and phenotype similarity.[236] In addition to available web applications and open-source tools, professional organizations offer data interpretation and variant prioritization services.

While these algorithms have been developed to exploit diverse principles in order to prioritize variants or predict novel disease genes, there is currently no published example showcasing their validity. Algorithms and tools often fall short due to the lack of specificity or overlook important factors in a particular disease. In cases such as IBD, where the phenotype is a result of the complex interplay of multiple factors, a more tailor-made solution could be more powerful to predict novel disease genes. An attractive avenue of prediction tools is to converge toward a contextualization-heavy, integrative, disease-specific variant annotation tool. In contrast to clear-cut prioritization, this approach emphasizes annotating all potential candidates with various types of information and combines information to relate individual gene defects to known pathways and phenotypes. This approach is exemplified by the identification of causal variants in *TRIM22*, a protein which is linked to NOD2 signaling by multiple molecular networks.[71]

9 | DATA SHARING, UNIFIED NOMENCLATURE, ONTOLOGIES

9.1 | Toward efficient data sharing and facilitating collaborations

Research on rare diseases such as VEO-IBD often relies on a few small pedigrees presenting with heterogeneous phenotypes. The small number of available patients and material coupled with the sheer number of potentially causal variants from NGS makes identification of causative gene defects a classical hunt for the "needle in a haystack." In the quest for the "needle," finding an additional patient carrying the same mutation strengthens the case of causality immensely.

Unfortunately, current variant and patient matching is hindered by the existence of small, independent datasets within the individual research groups and organizations. With no clear communication and channel for exchange, matching researchers up based on shared phenotypes and/or genotypes is simply a case of serendipity. This gap in communication and exchange is one of the main determinants of the pace of gene discovery. In order to facilitate the pace for identification of novel key players of IBD, bridging the gap between islands of research needs to be a priority.

Efforts to facilitate data and material exchange promise to bridge gaps between hospitals, specialized centers, and laboratories. Matchmaker Exchange, a project launched in 2013 addresses this crucial challenge, to facilitate the matching of cases

with similar phenotypic and genotypic profiles, using standardized programming interfaces.[237] Similar to Matchmaker Exchange, GeneMatcher, a project dedicated to enable connections between clinicians and researchers from around the world, to help unsolved exomes.[238]

9.2 | Increasing need of high-quality metadata

In recent years, large-scale computation methods have been initiated to investigate the etiology of IBD. In these projects, researches often rely on public databases to provide them with data. Access to accurate genetic data is facilitated by resources such as Decipher,[239] HGMD,[240] OMIM[241]and ClinVar[242] that aim to aggregate clinical data. These resources and public data repositories are still incomplete and need to be queried manually or with specifically set-up local bioinformatics pipeline. These efforts are welcome steps toward efficient data access, but some issues with redundancy remain and variable quality and quantity of data that is still missing from these resources. A robust, unified database could be an approach worth considering.

It is becoming increasingly clear that beyond efficient access to genomic and variant information, there is a need for accurate metadata to describe clinical information not only in a genetic manner, but also phenotypically. Annotation of patients with accurate phenotype data, as well as the annotation of genes with pathways and molecular mechanisms requires standardized and objective language. Therefore, it is crucial to have a unified nomenclature and resource of disease-causing genes annotated with the corresponding physical, molecular, and cellular phenotypes. Along with annotating patients and genes with the correct disease and ontology, intra-institute and laboratory collaborations would benefit immensely from precise and objective descriptions of phenotypic, molecular, and genetic abnormalities.

Human Phenotype Ontology (HPO) is a phenotype vocabulary initially published in 2008.[243,244] It is a tool that enables accurate phenotyping which further facilitates efficient data and patient exchange. HPO is being increasingly adapted into everyday use as the standard to describe phenotypic abnormalities. Gene ontology (GO) on the other hand, is a computational representation of the function and localization of genes and gene products on the molecular level. Currently, the GO project has developed and constantly revised over 40 000 biological concepts and annotations.[245,246] GO provides a nomenclature to annotate gene defects with detailed molecular and mechanistic information in a unified manner. Disease ontologies aim to provide standardized, consistent, and objective descriptions of human disease terms, phenotype characteristics, and related medical vocabulary disease concepts, as well as hierarchical relationships between the disease entities themselves. Efforts are currently ongoing to translate and include diseases into ontologies such as Orphanet,[247] Disgenet[248], or Disease Ontology.[249,250] In addition to providing a unified nomenclature that allows clinicians and researchers to characterize patients better, the inherent network structure of ontologies such as HPO and GO allows for pairwise distance (similarity) between two terms. Consequently, pairwise similarity measures can be used to carry out complex comparisons such as testing the similarity between two patients annotated by different terms.[251] Current ontologies have been useful in fulfilling current gap, but they are not complete. Numerous diseases and disease-gene association are not documented, and the ontology structures are incomplete.

10 | BEYOND GENETICS: IBDOMICS—SYSTEMS BIOLOGY AND INTEGRATIVE METHODS

One of the limitations of current approaches toward uncovering the key players in IBD pathology is that they are looking at individual contributors separately. In a case of IBD, which arises as a result of the combinatorial effect in multiple key players such as genetics, environmental factors, microbial perturbations, epigenetics, and lifestyle factors, understanding the disease cannot be tackled by studying each pathogenic component in isolation, without considering the interaction among the different "omes." Integration of the different "omes" requires intelligently designed data integration and processing. This integrative approach, recently introduced as the IBD interactome or IBDome, calls for new concepts and tools to implement a systems biology approach toward unraveling the processes behind IBD.[252] These tools should rely on unbiased data-driven investigation and include strategies to reveal key drivers and pinpoint central players of inflammation and enable development of targeted therapies.

Querying and integrating multiple "omes" with bioinformatics methods enables the integration of genomic, epigenomic, transcriptomic, proteomic, metabolomics, and microbiomic data to construct a comprehensive molecular map of IBD. Although seamless data integration is yet to be available, methods to integrate various types of information and use it toward identifying key players of IBD are already underway.

Recently, Peters et al[253] used individual networks constructed from molecular data generated from intestinal samples isolated from three populations of patients with IBD at different stages of disease, including two adult and pediatric cohorts of IBD. As a result, they developed a predictive model of the immune component of IBD that informs causal relationships among loci previously linked to IBD through GWAS using functional and regulatory annotations that relate to the cells, tissues, and pathophysiology of IBD. This network revealed potential key drivers of IBD pathogenesis. Among the key drivers were numerous known VEO-IBD genes, but also inborn errors of immunity and new candidates that were showed to have a role in inflammatory immune response.[253]

11 | CONCLUSION AND FUTURE PERSPECTIVES

IBD is a complex, multifactorial condition with an onset brought on by a multitude of factors that can also present in rare, Mendelian fashion. While the direct link between the more common, adult version

of IBD and rare, Mendelian EO-IBD is still elusive, we have gained tremendous understanding of different key players in immune regulation and cellular mechanisms required for immune homeostasis in the gut over the last years. The considerable fraction of inborn errors of immunity presenting with an IBD-like phenotype highlight that all patients presenting with VEO-IBD should be subjected to detailed genetic and immunological examination and investigation, as IBD can be an early sign of an underlying immunodeficiency. Recent advances in genomic technologies, organoid systems, as well as our increasing understanding and modeling of the interplay between the gut microbiota, immune cells, epithelial cells, and the environment promise to shed new light on to the complex molecular network behind IBD pathology. This novel understanding could allow for more efficiently identification of patient subgroups, and therefore increasingly direct treatment strategies toward personalized medicine.

ACKNOWLEDGEMENTS

This work was supported by a DOC Fellowship of the Austrian Academy of Sciences (24486) to R.C.A. and Austrian Science Fund (FWF) grant I2250-B28 to K. B.

CONFLICT OF INTEREST

The authors declare that there are no conflict of interests.

ORCID

Julia Pazmandi https://orcid.org/0000-0003-3411-9598
Artem Kalinichenko https://orcid.org/0000-0002-7020-7256
Rico Chandra Ardy https://orcid.org/0000-0003-4975-2177
Kaan Boztug https://orcid.org/0000-0001-8387-9185

REFERENCES

1. Hill DA, Artis D. Intestinal bacteria and the regulation of immune cell homeostasis. *Annu Rev Immunol*. 2010;28:623-667.
2. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol*. 2011;9:356-368.
3. Peterson LW, Artis D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Publ Gr*. 2014;14:141-153.
4. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat Rev Immunol*. 2008;8:458-466.
5. Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. *Nat Rev Immunol*. 2003;3:521-533.
6. Abraham C, Cho JH. Inflammatory bowel disease. *N Engl J Med*. 2009;361:2066-2078.
7. Torres J, Colombel J-F. Genetics and phenotypes in inflammatory bowel disease. *Lancet (London, England)*. 2016;387:98-100.
8. Jostins L, Ripke S, Weersma RK, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012a;491:119-124.
9. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47:979-986.
10. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49:256-261.
11. Verstockt B, Smith KG, Lee JC. Genome-wide association studies in Crohn's disease: past, present and future. *Clin Transl Immunol*. 2018;7:e1001.
12. Strober W, Watanabe T. NOD2, an intracellular innate immune sensor involved in host defense and Crohn's disease. *Mucosal Immunol*. 2011a;4:484-495.
13. Murthy A, Li Y, Peng I, et al. A Crohn's disease variant in Atg16 l1 enhances its degradation by caspase 3. *Nature*. 2014;506:456-462.
14. Zhang Y, Tian L, Sleiman P, Ghosh S, Hakonarson H, International IBD Genetics Consortium. Bayesian analysis of genome-wide inflammatory bowel disease data sets reveals new risk loci. *Eur J Hum Genet*. 2018;26:265-274.
15. Li YR, Li J, Zhao SD, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med*. 2015;21:1018-1027.
16. Ellinghaus D, Jostins L, Spain SL, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*. 2016;48:510-518.
17. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet*. 2013;14:661-673.
18. Maurano MT, Humbert R, Rynes E, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195.
19. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317-330.
20. Farh KK-H, Marson A, Zhu J, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518:337-343.
21. Ye CJ, Feng T, Kwon H-K, et al. Intersection of population variation and autoimmunity genetics in human T cell activation. *Science*. 2014;345:1254665.
22. Fairfax BP, Humburg P, Makino S, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*. 2014;343:1246949.
23. Peters JE, Lyons PA, Lee JC, et al. Insight into genotype-phenotype associations through eQTL mapping in multiple cell types in health and immune-mediated disease. *PLoS Genet*. 2016;12:e1005908.
24. Cho JH, Abraham C. Inflammatory bowel disease genetics: Nod2. *Annu Rev Med*. 2007;58:401-416.
25. Uhlig HH, Powrie F. Translating immunology into therapeutic concepts for inflammatory bowel disease. *Annu Rev Immunol*. 2018;36:755-781.
26. Glocker E-O, Kotlarz D, Boztug K, et al. Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N Engl J Med*. 2009a;361:2033-2045.
27. Uhlig HH, Schwerd T, Koletzko S, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. *Gastroenterology*. 2014a;147:990-1007 e3.
28. Vatn MH. Natural history and complications of IBD. *Curr Gastroenterol Rep*. 2009;11:481-487.
29. Speckmann C, Ehl S. XIAP deficiency is a mendelian cause of late-onset IBD. *Gut*. 2014;63:1031-1032.
30. Avitzur Y, Guo C, Mastropaolo LA, et al. Mutations in tetratricopeptide repeat domain 7A result in a severe form of very early onset inflammatory bowel disease. *Gastroenterology*. 2014;146:1028-1039.

31. Salzer E, Kansu A, Sic H, et al. Early-onset inflammatory bowel disease and common variable immunodeficiency-like disease caused by IL-21 deficiency. *J Allergy Clin Immunol*. 2014;133:1651-1659 e12.
32. Serwas NK, Kansu A, Santos-Valente E, et al. Atypical manifestation of LRBA deficiency with predominant IBD-like phenotype. *Inflamm Bowel Dis*. 2015;21:40-47.
33. Müller T, Rasool I, Heinz-Erian P, et al. Congenital secretory diarrhoea caused by activating germline mutations in GUCY2C. *Gut*. 2016;65:1306-1313.
34. Fréour T, Jarry A, Bach-Ngohou K, et al. TACE inhibition amplifies TNF-alpha-mediated colonic epithelial barrier disruption. *Int J Mol Med*. 2009;23:41-48.
35. Bousfiha A, Jeddane L, Picard C, et al. The 2017 IUIS phenotypic classification for primary immunodeficiencies. *J Clin Immunol*. 2018;38:129-143.
36. Uhlig HH, Schwerd T. From genes to mechanisms. *Inflamm Bowel Dis*. 2016;22:202-212.
37. Bigorgne AE, Farin HF, Lemoine R, et al. TTC7A mutations disrupt intestinal epithelial apicobasal polarity. *J Clin Invest*. 2014;124:328-337.
38. Woutsas S, Aytekin C, Salzer E, et al. Hypomorphic mutation in TTC7A causes combined immunodeficiency with mild structural intestinal defects. *Blood*. 2015;125:1674-1676.
39. Kern J, Herz C, Haan E, et al. Chronic colitis due to an epithelial barrier defect: the role of kindlin-1 isoforms. *J Pathol*. 2007;213:462-470.
40. Freeman EB, Köglmeier J, Martinez AE, et al. Gastrointestinal complications of epidermolysis bullosa in children. *Br J Dermatol*. 2008;158:1308-1314.
41. Sadler E, Klausegger A, Muss W, et al. Novel KIND1 gene mutation in kindler syndrome with severe gastrointestinal tract involvement. *Arch Dermatol*. 2006;142:1619-1624.
42. Ussar S, Moser M, Widmaier M, et al. Loss of kindlin-1 causes skin atrophy and lethal neonatal intestinal epithelial dysfunction. *PLoS Genet*. 2008;4:e1000289.
43. Wertheim-Tysarowska K, Sobczyńska-Tomaszewska A, Kowalewski C, et al. The COL7A1 mutation database. *Hum Mutat*. 2012;33:327-331.
44. Fiskerstrand T, Arshad N, Haukanes BI, et al. Familial diarrhea syndrome caused by an activating *GUCY2C* mutation. *N Engl J Med*. 2012;366:1586-1595.
45. Janecke AR, Heinz-Erian P, Yin J, et al. Reduced sodium/proton exchanger NHE3 activity causes congenital sodium diarrhea. *Hum Mol Genet*. 2015;24:6614-6623.
46. Afrache H, Gouret P, Ainouche S, Pontarotti P, Olive D. The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response. *Immunogenetics*. 2012;64:781-794.
47. Umeno J, Hisamatsu T, Esaki M, et al. A hereditary enteropathy caused by mutations in the SLCO2A1 gene, encoding a prostaglandin transporter. *PLoS Genet*. 2015;11:e1005581.
48. Uchida K, Nakajima A, Ushijima K, et al. Pediatric-onset chronic nonspecific multiple ulcers of small intestine: a nationwide survey and genetic study in Japan. *J Pediatr Gastroenterol Nutr*. 2017;64:565-568.
49. Zhang Z, Xia W, He J, et al. Exome sequencing identifies SLCO2A1 mutations as a cause of primary hypertrophic osteoarthropathy. *Am J Hum Genet*. 2012;90:125-132.
50. Busch J, Frank V, Bachmann N, et al. Mutations in the prostaglandin transporter SLCO2A1 cause primary hypertrophic osteoarthropathy with digital clubbing. *J Invest Dermatol*. 2012;132:2473-2476.
51. Brooke MA, Longhurst HJ, Plagno V, et al. Cryptogenic multifocal ulcerating stenosing enteritis associated with homozygous deletion mutations in cytosolic phospholipase A2-α. *Gut*. 2014;63:96-104.
52. Henter J-I, Horne A, Aricó M, et al. HLH-2004: diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. *Pediatr Blood Cancer*. 2007;48:124-131.
53. Meeths M, Entesarian M, Al-Herz W, et al. Spectrum of clinical presentations in familial hemophagocytic lymphohistiocytosis type 5 patients with mutations in STXBP2. *Blood*. 2010;116:2635-2643.
54. Riento K, Jäntti J, Jansson S, et al. A sec1-related vesicle-transport protein that is expressed predominantly in epithelial cells. *Eur J Biochem*. 1996;239:638-646.
55. Tellam JT, McIntosh S, James DE. Molecular identification of two novel Munc-18 isoforms expressed in non-neuronal tissues. *J Biol Chem*. 1995;270:5857-5863.
56. Fournier BM, Parkos CA. The role of neutrophils during intestinal inflammation. *Mucosal Immunol*. 2012;5:354-366.
57. Arnold DE, Heimall JR. A review of chronic granulomatous disease. *Adv Ther*. 2017;34:2543-2557.
58. Rada B, Leto T. Oxidative innate immune defenses by Nox/Duox Family NADPH oxidases. *Contrib Microbiol*. 2008;15:164-187.
59. Hayes P, Dhillon S, O'Neill K, et al. Defects in nicotinamide-adenine dinucleotide phosphate oxidase genes NOX1 and DUOX2 in very early onset inflammatory bowel disease. *Cell Mol Gastroenterol Hepatol*. 2015;1:489-502.
60. Tse HM, Thayer TC, Steele C, et al. NADPH oxidase deficiency regulates Th lineage commitment and modulates autoimmunity. *J Immunol*. 2010;185:5247-5258.
61. Meissner F, Seger RA, Moshous D, Fischer A, Reichenbach J, Zychlinsky A. Inflammasome activation in NADPH oxidase defective mononuclear phagocytes from patients with chronic granulomatous disease. *Blood*. 2010;116:1570-1573.
62. Visser G, Rake J-P, Fernandes J, et al. Neutropenia, neutrophil dysfunction, and inflammatory bowel disease in glycogen storage disease type Ib: results of the European Study on Glycogen Storage Disease Type I. *J Pediatr*. 2000;137:187-191.
63. Yamaguchi T, Ihara K, Matsumoto T, et al. Inflammatory bowel disease-like colitis in glycogen storage disease type 1b. *Inflamm Bowel Dis*. 2001;7:128-132.
64. Bégin P, Patey N, Mueller P, et al. Inflammatory bowel disease and T cell lymphopenia in G6PC3 deficiency. *J Clin Immunol*. 2013;33:520-525.
65. Hayee B, Antonopoulos A, Murphy EJ, et al. G6PC3 mutations are associated with a major defect of glycosylation: a novel mechanism for neutrophil dysfunction. *Glycobiology*. 2011;21:914-924.
66. Arnaout MA. Leukocyte adhesion molecules deficiency: its structural basis, pathophysiology and implications for modulating the inflammatory response. *Immunol Rev*. 1990;114:145-180.
67. D'Agata ID, Paradis K, Chad Z, Bonny Y, Seidman E. Leucocyte adhesion deficiency presenting as a chronic ileocolitis. *Gut*. 1996;39:605-608.
68. Uzel G, Kleiner DE, Kuhns DB, Holland SM. Dysfunctional LAD-1 neutrophils and colitis. *Gastroenterology*. 2001;121:958-964.
69. Wouters CH, Maes A, Foley KP, Bertin J, Rose CD. Blau syndrome, the prototypic auto-inflammatory granulomatous disease. *Pediatr Rheumatol Online J*. 2014;12:33.
70. Schwerd T, Pandey S, Yang HT, et al. Impaired antibacterial autophagy links granulomatous intestinal inflammation in Niemann-Pick disease type C1 and XIAP deficiency with NOD2 variants in Crohn's disease. *Gut*. 2017;66:1060-1073.
71. Booth C, Gilmour KC, Veys P, 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;117:53-62.
72. Filipovich AH, Zhang K, Snow AL, Marsh RA. X-linked lymphoproliferative syndromes: brothers or distant cousins? *Blood*. 2010;116:3398-3408.

73. Sharifi R, Sinclair JC, Gilmour KC, et al. SAP mediates specific cytotoxic T-cell functions in X-linked lymphoproliferative disease. *Blood*. 2004;103:3821-3827.
74. Sayós J, Martín M, Chen A, et al. Cell surface receptors Ly-9 and CD84 recruit the X-linked lymphoproliferative disease gene product SAP. *Blood*. 2001;97:3867-3874.
75. Nichols KE, Hom J, Gong S-Y, et al. Regulation of NKT cell development by SAP, the protein defective in XLP. *Nat Med*. 2005;11:340-345.
76. Rigaud S, Fondanèche M-C, Lambert N, et al. XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome. *Nature*. 2006;444:110-114.
77. Horn PC, Schuster V. Gastritis and colitis can be associated with XLP-1 (SAP deficiency). *Pediatr Int*. 2012;54:964.
78. Li Q, Lee CH, Peters LA, et al. Variants in TRIM22 that affect NOD2 signaling are associated with very-early-onset inflammatory bowel disease. *Gastroenterology*. 2016;150:1196-1207.
79. Parlato M, Charbit-Henrion F, Pan J, et al. Human ALPI deficiency causes inflammatory bowel disease and highlights a key mechanism of gut homeostasis. *EMBO Mol Med*. 2018;10:e8483.
80. Agarwal S, Mayer L. Pathogenesis and treatment of gastrointestinal disease in antibody deficiency syndromes. *J Allergy Clin Immunol*. 2009a;124:658-664.
81. Conley ME, Dobbs AK, Quintana AM, et al. Agammaglobulinemia and absent B lineage cells in a patient lacking the p85α subunit of PI3K. *J Exp Med*. 2012;209:463-470.
82. Elgizouli M, Lowe DM, Speckmann C, et al. Activating PI3Kδ mutations in a cohort of 669 patients with primary immunodeficiency. *Clin Exp Immunol*. 2016;183:221-229.
83. Heindl M, Händel N, Ngeow J, et al. Autoimmunity, intestinal lymphoid hyperplasia, and defects in mucosal B-cell homeostasis in patients with PTEN hamartoma tumor syndrome. *Gastroenterology*. 2012;142:1093-1096 e6.
84. Driessen GJ, IJspeert H, Wentink M, et al. Increased PI3K/Akt activity and deregulated humoral immune response in human PTEN deficiency. *J Allergy Clin Immunol*. 2016;138:1744-1747 e5.
85. Kinnear C, Glanzmann B, Banda E, et al. Exome sequencing identifies a novel TTC37 mutation in the first reported case of Trichohepatoenteric syndrome (THE-S) in South Africa. *BMC Med Genet*. 2017;18:26.
86. Hartley JL, Zachos NC, Dawood B, et al. Mutations in TTC37 cause trichohepatoenteric syndrome (phenotypic diarrhea of infancy). *Gastroenterology*. 2010a;138:2388-2398 e2.
87. Rider NL, Boisson B, Jyonouchi S, et al. Novel TTC37 mutations in a patient with immunodeficiency without diarrhea: extending the phenotype of trichohepatoenteric syndrome. *Front Pediatr*. 2015a;3:2.
88. Frans G, Moens L, Schaballie H, et al. Homozygous N-terminal missense mutation in TRNT1 leads to progressive B-cell immunodeficiency in adulthood. *J Allergy Clin Immunol*. 2017;139:360-363 e6.
89. Warnatz K, Bossaller L, Salzer U, et al. Human ICOS deficiency abrogates the germinal center reaction and provides a monogenic model for common variable immunodeficiency. *Blood*. 2006;107:3045-3052.
90. Notarangelo LD. Functional T cell immunodeficiencies (with T cells present). *Annu Rev Immunol*. 2013;31:195-225.
91. Liston A, Enders A, Siggs OM. Unravelling the association of partial T-cell immunodeficiency and immune dysregulation. *Nat Rev Immunol*. 2008;8:545-558.
92. Felgentreff K, Perez-Becker R, Speckmann C, et al. Clinical and immunological manifestations of patients with atypical severe combined immunodeficiency. *Clin Immunol*. 2011;141:73-82.
93. Chou J, Hanna-Wakim R, Tirosh I, et al. A novel homozygous mutation in recombination activating gene 2 in 2 relatives with different clinical phenotypes: Omenn syndrome and hyper-IgM syndrome. *J Allergy Clin Immunol*. 2012;130:1414-1416.
94. Villa A, Santagata S, Bozzi F, et al. Partial V(D)J recombination activity leads to Omenn syndrome. *Cell*. 1998;93:885-896.
95. Ege M, Ma Y, Manfras B, et al. Omenn syndrome due to ARTEMIS mutations. *Blood*. 2005;105:4179-4186.
96. Villa A, Notarangelo LD, Roifman CM. Omenn syndrome: inflammation in leaky severe combined immunodeficiency. *J Allergy Clin Immunol*. 2008;122:1082-1086.
97. Santagata S, Gomez CA, Sobacchi C, et al. N-terminal RAG1 frameshift mutations in Omenn's syndrome: internal methionine usage leads to partial V(D)J recombination activity and reveals a fundamental role in vivo for the N-terminal domains. *Proc Natl Acad Sci*. 2000;97:14572-14577.
98. Flinn AM, Gennery AR. Adenosine deaminase deficiency: a review. *Orphanet J Rare Dis*. 2018;13:65.
99. Apasov SG, Blackburn MR, Kellems RE, Smith PT, Sitkovsky MV. Adenosine deaminase deficiency increases thymic apoptosis and causes defective T cell receptor signaling. *J Clin Invest*. 2001;108:131-141.
100. Aldrich MB, Chen W, Blackburn MR, Martinez-Valdez H, Datta SK, Kellems RE. Impaired germinal center maturation in adenosine deaminase deficiency. *J Immunol*. 2003;171:5562-5570.
101. Takeshita T, Asao H, Ohtani K, et al. Cloning of the gamma chain of the human IL-2 receptor. *Science*. 1992;257:379-382.
102. Schmalstieg FC, Leonard WJ, Noguchi M, et al. Missense mutation in exon 7 of the common gamma chain gene causes a moderate form of X-linked combined immunodeficiency. *J Clin Invest*. 1995;95:1169-1173.
103. Noguchi M, Yi H, Rosenblatt HM, et al. Interleukin-2 receptor gamma chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell*. 1993;73:147-157.
104. Dobbs K, Domínguez Conde C, Zhang S-Y, et al. Inherited DOCK2 deficiency in patients with early-onset invasive infections. *N Engl J Med*. 2015;372:2409-2422.
105. Chan AC, Kadlecek TA, Elder ME, et al. ZAP-70 deficiency in an autosomal recessive form of severe combined immunodeficiency. *Science*. 1994a;264:1599-1601.
106. Feske S. Calcium signalling in lymphocyte activation and disease. *Nat Rev Immunol*. 2007;7:690-702.
107. Feske S. Immunodeficiency due to defects in store-operated calcium entry. *Ann N Y Acad Sci*. 2011;1238:74-90.
108. Prochazkova J, Sakaguchi S, Owusu M, et al. DNA repair cofactors ATMIN and NBS1 are required to suppress T cell activation. *PLoS Genet*. 2015;11:e1005645.
109. Wang J, German J, Ashby K, French SW. Ulcerative colitis complicated by dysplasia-adenoma-carcinoma in a man with Bloom's syndrome. *J Clin Gastroenterol*. 1999;28:380-382.
110. Conrad MA, Dawany N, Sullivan KE, Devoto M, Kelsen JR. Novel ZBTB24 mutation associated with immunodeficiency, centromere instability, and facial anomalies type-2 syndrome identified in a patient with very early onset inflammatory bowel disease. *Inflamm Bowel Dis*. 2017;23:2252-2255.
111. Ballew BJ, Yeager M, Jacobs K, et al. Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in Dyskeratosis congenita. *Hum Genet*. 2013;132:473-480.
112. Sznajer Y, Baumann C, David A, et al. Further delineation of the congenital form of X-linked dyskeratosis congenita (Hoyeraal-Hreidarsson syndrome). *Eur J Pediatr*. 2003;162:863-867.
113. Borggraefe I, Koletzko S, Arenz T, et al. Severe variant of X-linked dyskeratosis congenita (hoyeraal-hreidarsson syndrome) causes significant enterocolitis in early infancy. *J Pediatr Gastroenterol Nutr*. 2009;49:359-363.
114. Knight SW, Heiss NS, Vulliamy TJ, et al. Unexplained aplastic anaemia, immunodeficiency, and cerebellar hypoplasia (Hoyeraal-Hreidarsson syndrome) due to mutations in the dyskeratosis congenita gene, DKC1. *Br J Haematol*. 1999;107:335-339.

115. Jonassaint NL, Guo N, Califano JA, Montgomery EA, Armanios M. The gastrointestinal manifestations of telomere-mediated disease. *Aging Cell*. 2013;12:319-323.
116. Lohr NJ, Molleston JP, Strauss KA, et al. Human ITCH E3 ubiquitin ligase deficiency causes syndromic multisystem autoimmune disease. *Am J Hum Genet*. 2010a;86:447-453.
117. Matesic LE, Haines DC, Copeland NG, Jenkins NA. Itch genetically interacts with Notch1 in a mouse autoimmune disease model. *Hum Mol Genet*. 2006;15:3485-3497.
118. Lorick KL, Yang Y, Jensen JP, Iwai K, Weissman AM. Studies of the ubiquitin proteasome system. *Curr Protoc Cell Biol*. 2006;31:15.9.1-15.9.85.
119. Anandasabapathy N, Ford GS, Bloom D, et al. GRAIL: an E3 ubiquitin ligase that inhibits cytokine gene transcription is expressed in anergic CD4 + T cells. *Immunity*. 2003;18:535-547.
120. Bachmaier K, Krawczyk C, Kozieradzki I, et al. Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor Cbl-b. *Nature*. 2000;403:211-216.
121. Torgerson TR, Ochs HD. Immune dysregulation, polyendocrinopathy, enteropathy, X-linked: forkhead box protein 3 mutations and lack of regulatory T cells. *J Allergy Clin Immunol*. 2007;120:744-750.
122. Caudy AA, Reddy ST, Chatila T, Atkinson JP, Verbsky JW. CD25 deficiency causes an immune dysregulation, polyendocrinopathy, enteropathy, X-linked–like syndrome, and defective IL-10 expression from CD4 lymphocytes. *J Allergy Clin Immunol*. 2007;119:482-487.
123. Friedline RH, Brown DS, Nguyen H, et al. CD4 $^{+}$ regulatory T cells require CTLA-4 for the maintenance of systemic tolerance. *J Exp Med*. 2009;206:421-434.
124. Tivol EA, Borriello F, Schweitzer AN, Lynch WP, Bluestone JA, Sharpe AH. Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity*. 1995;3:541-547.
125. Lo B, Zhang K, Lu W, et al. Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy. *Science*. 2015a;349:436-440.
126. Krummel MF, Allison JP. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *J Exp Med*. 1995;182:459-465.
127. Schubert D, Bode C, Kenefeck R, et al. Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat Med*. 2014;20:1410-1416.
128. Kuehn HS, Ouyang W, Lo B, et al. Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. *Science*. 2014;345:1623-1627.
129. Lopez-Herrera G, Tampella G, Pan-Hammarström Q, et al. Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. *Am J Hum Genet*. 2012;90:986-1001.
130. Lo B, Fritz JM, Su HC, Uzel G, Jordan MB, Lenardo MJ. Blood spotlight CHAI and LATAIE: new genetic diseases of CTLA-4 checkpoint insufficiency. *Blood*. 2016;128:1037-1042.
131. Snapper SB, Rosen FS, Mizoguchi E, et al. Wiskott-Aldrich syndrome protein-deficient mice reveal a role for WASP in T but not B cell activation. *Immunity*. 1998;9:81-91.
132. Ohya T, Yanagimachi M, Iwasawa K, et al. Childhood-onset inflammatory bowel diseases associated with mutation of Wiskott-Aldrich syndrome protein gene. *World J Gastroenterol*. 2017;23:8544-8552.
133. Ivashkiv LB. Jak-STAT signaling pathways in cells of the immune system. *Rev Immunogenet*. 2000;2:220-230.
134. Flanagan SE, Haapaniemi E, Russell MA, et al. Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nat Genet*. 2014;46:812-814.
135. Yang XO, Panopoulos AD, Nurieva R, et al. STAT3 regulates cytokine-mediated generation of inflammatory helper T cells. *J Biol Chem*. 2007;282:9358-9363.
136. Harris TJ, Grosso JF, Yen H-R, et al. Cutting edge: an in vivo requirement for STAT3 signaling in TH17 development and TH17-dependent autoimmunity. *J Immunol*. 2007;179:4313-4317.
137. Punwani D, Wang H, Chan AY, et al. Combined immunodeficiency due to MALT1 mutations, treated by hematopoietic cell transplantation. *J Clin Immunol*. 2015;35:135-146.
138. Jaworski M, Thome M. The paracaspase MALT1: biological function and potential for therapeutic inhibition. *Cell Mol Life Sci*. 2016;73:459-473.
139. Liu W, Guo W, Hang N, et al. MALT1 inhibitors prevent the development of DSS-induced experimental colitis in mice via inhibiting NF-κB and NLRP3 inflammasome activation. *Oncotarget*. 2016;7:30536-30549.
140. Rowe JH, Delmonte OM, Keles S, et al. Patients with *CD3G* mutations reveal a role for human CD3γ in T_{reg} diversity and suppressive function. *Blood*. 2018a;131:2335-2344.
141. Kotlarz D, Ziętara N, Uzel G, et al. Loss-of-function mutations in the IL-21 receptor gene cause a primary immunodeficiency syndrome. *J Exp Med*. 2013a;210:433-443.
142. de Jesus AA, Canna SW, Liu Y, Goldbach-Mansky R. Molecular mechanisms in genetically defined autoinflammatory diseases: disorders of amplified danger signaling. *Annu Rev Immunol*. 2015;33:823-874.
143. Bader-Meunier B, Florkin B, Sibilia J, et al. Mevalonate kinase deficiency: a survey of 50 patients. *Pediatrics*. 2011;128:e152-e159.
144. Galeotti C, Meinzer U, Quartier P, et al. Efficacy of interleukin-1-targeting drugs in mevalonate kinase deficiency. *Rheumatology*. 2012;51:1855-1859.
145. Frenkel J, Houten SM, Waterham HR, et al. Mevalonate kinase deficiency and Dutch type periodic fever. *Clin Exp Rheumatol*. 2000;18:525-532.
146. Drenth JPH, Cuisset L, Grateau G, et al. Mutations in the gene encoding mevalonate kinase cause hyper-IgD and periodic fever syndrome. *Nat Genet*. 1999;22:178-181.
147. Mandey SHL, Kuijk LM, Frenkel J, Waterham HR. A role for geranylgeranylation in interleukin-1β secretion. *Arthritis Rheum*. 2006;54:3690-3695.
148. Zhou Q, Lee G-S, Brady J, et al. A hypermorphic missense mutation in PLCG2, encoding phospholipase Cγ2, causes a dominantly inherited autoinflammatory disease with immunodeficiency. *Am J Hum Genet*. 2012;91:713-720.
149. Egritas O, Dalgic B. Infantile colitis as a novel presentation of familial mediterranean fever responding to colchicine therapy. *J Pediatr Gastroenterol Nutr*. 2011;53:102-105.
150. Sari S, Egritas O, Dalgic B. The familial Mediterranean fever (MEFV) gene may be a modifier factor of inflammatory bowel disease in infancy. *Eur J Pediatr*. 2008;167:391-393.
151. Manukyan G, Aminov R. Update on pyrin functions and mechanisms of familial mediterranean fever. *Front Microbiol*. 2016;7:456.
152. Romberg N, Al Moussawi K, Nelson-Williams C, et al. Mutation of NLRC4 causes a syndrome of enterocolitis and autoinflammation. *Nat Genet*. 2014;46:1135-1139.
153. Canna SW, de Jesus AA, Gouni S, et al. An activating NLRC4 inflammasome mutation causes autoinflammation with recurrent macrophage activation syndrome. *Nat Genet*. 2014;46:1140-1146.
154. Zhou Q, Wang H, Schwartz DM, et al. Loss-of-function mutations in TNFAIP3 leading to A20 haploinsufficiency cause an early-onset autoinflammatory disease. *Nat Genet*. 2016;48:67-73.
155. Lehle AS, Farin HF, Marquardt B, et al. Intestinal inflammation and dysregulated immunity in patients with inherited caspase-8 deficiency. *Gastroenterology*. 2018. https://doi.org/10.1053/j.gastro.2018.09.041.
156. Chun HJ, Zheng L, Ahmad M, et al. Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature*. 2002;419:395-399.

157. Fabre A, Charroux B, Martinez-Vinson C, et al. SKIV2L mutations cause syndromic diarrhea, or trichohepatoenteric syndrome. *Am J Hum Genet*. 2012;90:689-692.
158. Hiejima E, Yasumi T, Nakase H, et al. Tricho-hepato-enteric syndrome with novel SKIV2L gene mutations: a case report. *Medicine (Baltimore)*. 2017;96:e8601.
159. Eckard SC, Rice GI, Fabre A, et al. The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nat Immunol*. 2014;15:839-845.
160. Jaeckle Santos LJ, Xing C, Barnes RB, et al. Refined mapping of X-linked reticulate pigmentary disorder and sequencing of candidate genes. *Hum Genet*. 2008;123:469-476.
161. Starokadomskyy P, Gemelli T, Rios JJ, et al. DNA polymerase-α regulates the activation of type I interferons through cytosolic RNA:DNA synthesis. *Nat Immunol*. 2016;17:495-504.
162. Blaydon DC, Biancheri P, Di W-L, et al. Inflammatory skin and bowel disease linked to *ADAM17* deletion. *N Engl J Med*. 2011a;365:1502-1508.
163. Peschon JJ, Slack JL, Reddy P, et al. An essential role for ectodomain shedding in mammalian development. *Science*. 1998;282:1281-1284.
164. Jain U, Otley AR, Van Limbergen J, Stadnyk AW. The complement system in inflammatory bowel disease. *Inflamm Bowel Dis*. 2014;20:1628-1637.
165. Stengaard-Pedersen K, Thiel S, Gadjeva M, et al. Inherited deficiency of mannan-binding lectin–associated serine protease 2. *N Engl J Med*. 2003;349:554-560.
166. Thiel S, Steffensen R, Christensen IJ, et al. Deficiency of mannan-binding lectin associated serine protease-2 due to missense polymorphisms. *Genes Immun*. 2007;8:154-163.
167. Schlapbach LJ, Thiel S, Kessler U, Ammann RA, Aebi C, Jensenius JC. Congenital H-ficolin deficiency in premature infants with severe necrotising enterocolitis. *Gut*. 2011;60:1438-1439.
168. Ozen A, Comrie WA, Ardy RC, et al. CD55 deficiency, early-onset protein-losing enteropathy, and thrombosis. *N Engl J Med*. 2017;377:52-61.
169. Sutavani RV, Bradley RG, Ramage JM, Jackson AM, Durrant LG, Spendlove I. CD55 costimulation induces differentiation of a discrete T regulatory type 1 cell population with a stable phenotype. *J Immunol*. 2013;191:5895-5903.
170. Kurolap A, Eshach-Adiv O, Hershkovitz T, et al. Loss of CD55 in eculizumab-responsive protein-losing enteropathy. *N Engl J Med*. 2017;377:87-89.
171. Erzin Y, Cosgun S, Dobrucali A, Tasyurekli M, Erdamar S, Tuncer M. Complicated granulomatous colitis in a patient with Hermansky-Pudlak syndrome, successfully treated with infliximab. *Acta Gastroenterol Belg*. 2000;69:213-216.
172. Hazzan D, Seward S, Stock H, et al. Crohn's-like colitis, enterocolitis and perianal disease in Hermansky-Pudlak syndrome. *Color Dis*. 2006;8:539-543.
173. Lupi A, Tenni R, Rossi A, Cetta G, Forlino A. Human prolidase and prolidase deficiency: an overview on the characterization of the enzyme involved in proline recycling and on the effects of its mutations. *Amino Acids*. 2008;35:739-752.
174. Ledoux P, Scriver C, Hechtman P. Four novel PEPD alleles causing prolidase deficiency. *Am J Hum Genet*. 1994;54:1014-1021.
175. Lupi A, Riso A, Torre S, et al. Characterization of a new PEPD allele causing prolidase deficiency in two unrelated patients: natural-occurrent mutations as a tool to investigate structure?function relationship. *J Hum Genet*. 2004;49:500.
176. Naviglio S, Arrigo S, Martelossi S, et al. Severe inflammatory bowel disease associated with congenital alteration of transforming growth factor beta signaling. *J Crohn's Colitis*. 2014;8:770-774.
177. Kotlarz D, Marquardt B, Barøy T, et al. Human TGF-β1 deficiency causes severe inflammatory bowel disease and encephalopathy. *Nat Genet*. 2018;50:344-348.
178. Boland BS, Widjaja CE, Banno A, et al. Immunodeficiency and autoimmune enterocolopathy linked to NFAT5 haploinsufficiency. *J Immunol*. 2015;194:2551-2560.
179. van Haaften-Visser DY, Harakalova M, Mocholi E, et al. Ankyrin repeat and zinc-finger domain-containing 1 mutations are associated with infantile-onset inflammatory bowel disease. *J Biol Chem*. 2017;292:7904-7920.
180. Hauck F, Voss R, Urban C, Seidel MG. Intrinsic and extrinsic causes of malignancies in patients with primary immunodeficiency disorders. *J Allergy Clin Immunol*. 2018;141:59-68 e4.
181. Zhang Y, Su HC, Lenardo MJ. Genomics is rapidly advancing precision medicine for immunological disorders. *Nat Immunol*. 2015;16:1001-1004.
182. Kammermeier J, Lucchini G, Pai S-Y, et al. Stem cell transplantation for tetratricopeptide repeat domain 7A deficiency: long-term follow-up. *Blood*. 2016;128:1306-1308.
183. Weinstock GM. Genomic approaches to studying the human microbiota. *Nature*. 2012;489:250-256.
184. Zoetendal EG, Rajilic-Stojanovic M, de Vos WM. High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. *Gut*. 2008;57:1605-1615.
185. NIH HMP Working Group TNHW, Peterson J, Garges S, et al. The NIH human microbiome project. *Genome Res*. 2009;19:2317-2323.
186. Human Microbiome Project Consortium BA, Nelson KE, Pop M, et al. A framework for human microbiome research. *Nature*. 2012;486:215-221.
187. McHardy IH, Goudarzi M, Tong M, et al. Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome*. 2013;1:17.
188. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18:83.
189. Sun YV, Hu Y-J. Integrative analysis of multi-omics data for discovery and functional studies of complex human diseases. *Adv Genet*. 2016;93:147-190.
190. Marchesi JR, Adams DH, Fava F, et al. The gut microbiota and host health: a new clinical frontier. *Gut*. 2016;65:330-339.
191. Zeng MY, Inohara N, Nuñez G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol*. 2017;10:18-26.
192. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol*. 2017;14:573.
193. Conte MP, Schippa S, Zamboni I, et al. Gut-associated bacterial microbiota in paediatric patients with inflammatory bowel disease. *Gut*. 2006;55:1760-1767.
194. Schwiertz A, Jacobi M, Frick J-S, Richter M, Rusch K, Köhler H. Microbiota in pediatric inflammatory bowel disease. *J Pediatr*. 2010;157:240-244 e1.
195. Hansen R, Russell RK, Reiff C, et al. Microbiota of De-Novo Pediatric IBD: increased faecalibacterium prausnitzii and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am J Gastroenterol*. 2012;107:1913-1922.
196. Morgan XC, Tickle TL, Sokol H, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol*. 2012;13:R79.
197. Jørgensen SF, Trøseid M, Kummen M, et al. Altered gut microbiota profile in common variable immunodeficiency associates with levels of lipopolysaccharide and markers of systemic immune activation. *Mucosal Immunol*. 2016;9:1455-1465.
198. Levy M, Thaiss CA, Elinav E. Metabolites: messengers between the microbiota and the immune system. *Genes Dev*. 2016;30:1589-1597.

199. Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell*. 2016;165:1332-1345.
200. Manicassamy S, Pulendran B. Retinoic acid-dependent regulation of immune responses by dendritic cells and macrophages. *Semin Immunol*. 2009;21:22-27.
201. Próchnicki T, Latz E. Inflammasomes on the crossroads of innate immune recognition and metabolic control. *Cell Metab*. 2017;26:71-93.
202. Levy M, Thaiss CA, Zeevi D, et al. Microbiota-modulated metabolites shape the intestinal microenvironment by regulating NLRP6 inflammasome signaling. *Cell*. 2015;163:1428-1443.
203. Kjer-Nielsen L, Patel O, Corbett AJ, et al. MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature*. 2012;491:717-723.
204. Lepore M, Kalinichenko A, Calogero S, et al. Functionally diverse human T cells recognize non-microbial antigens presented by MR1. *Elife*. 2017;6:1-22.
205. Colman RJ, Rubin DT. Fecal microbiota transplantation as therapy for inflammatory bowel disease: a systematic review and meta-analysis. *J Crohn's Colitis*. 2014;8:1569-1581.
206. Chen L, Wilson JE, Koenigsknecht MJ, et al. NLRP12 attenuates colon inflammation by maintaining colonic microbial diversity and promoting protective commensal bacterial growth. *Nat Immunol*. 2017;18:541-551.
207. Seregin SS, Golovchenko N, Schaf B, et al. NLRP6 protects $Il10^{-/-}$ mice from colitis by limiting colonization of *Akkermansia muciniphila*. *Cell Rep*. 2017;19:733-745.
208. Suez J, Elinav E. The path towards microbiome-based metabolite treatment. *Nat Microbiol*. 2017;2:17075.
209. Dekkers JF, Berkers G, Kruisselbrink E, et al. Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis. *Sci Transl Med*. 2016;8:344ra84.
210. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169:1177-1186.
211. Kammermeier J, Drury S, James CT, et al. Targeted gene panel sequencing in children with very early onset inflammatory bowel disease–evaluation and prospective analysis. *J Med Genet*. 2014;51:748-755.
212. Kammermeier J, Dziubak R, Pescarin M, et al. Phenotypic and genotypic characterisation of inflammatory bowel disease presenting before the age of 2 years. *J Crohn's Colitis*. 2017;11:60-69.
213. Kelsen JR, Dawany N, Moran CJ, et al. Exome sequencing analysis reveals variants in primary immunodeficiency genes in patients with very early onset inflammatory bowel disease. *Gastroenterology*. 2015;149:1415-1424.
214. Ashton JJ, Andreoletti G, Coelho T, et al. Identification of variants in genes associated with single-gene inflammatory bowel disease by whole-exome sequencing. *Inflamm Bowel Dis*. 2016;22:2317-2327.
215. Ostrowski J, Paziewska A, Lazowska I, et al. Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Sci Rep*. 2016;6:39831.
216. Oh SH, Baek J, Kim KM, et al. Is whole exome sequencing clinically practical in the management of pediatric Crohn's disease? *Gut Liv*. 2015;9:767.
217. Petersen B-S, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18:14.
218. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet*. 2015;47:717-726.
219. Belkadi A, Bolze A, Itan Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*. 2015;112:5473-5478.
220. Luo Y, de Lange KM, Jostins L, et al. Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. *Nat Genet*. 2017;49:186-192.
221. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9:eaal5209.
222. Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun*. 2017;8:15824.
223. MacArthur DG, Balasubramanian S, Frankish A, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012;335:823-828.
224. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285-291.
225. Gibbs RA, Boerwinkle E, Doddapaneni H, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74.
226. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7:248-249.
227. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31:3812-3814.
228. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15:901-913.
229. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310-315.
230. Rosenbloom KR, Dreszer TR, Long JC, et al. ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res*. 2012;40:D912-D917.
231. Köhler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet*. 2008;82:949-958.
232. Ghiassian SD, Menche J, Barabási A-L. A DIseAse MOdule Detection (DIAMOnD) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput Biol*. 2015;11:e1004120.
233. Gorenshteyn D, Zaslavsky E, Fribourg M, et al. Interactive big data resource to elucidate human immune pathways and diseases. *Immunity*. 2015;43:605-614.
234. Meyer MJ, Beltrán JF, Liang S, et al. Interactome INSIDER: a structural interactome browser for genomic studies. *Nat Methods*. 2018;15:107-114.
235. Tranchevent L-C, Ardeshirdavani A, ElShal S, et al. Candidate gene prioritization with Endeavour. *Nucleic Acids Res*. 2016;44:W117-W121.
236. Smedley D, Jacobsen JOB, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015;10:2004-2015.
237. Philippakis AA, Azzariti DR, Beltran S, et al. The matchmaker exchange: a platform for rare disease gene discovery. *Hum Mutat*. 2015;36:915-921.
238. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*. 2015;36:928-930.
239. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am J Hum Genet*. 2009;84:524-533.
240. Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. The Human Gene Mutation Database: building a comprehensive

mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014;133:1-9.
241. Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum Mutat*. 2011;32:564-567.
242. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42:D980-D985.
243. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83:610-615.
244. Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res*. 2017;45:D865-D876.
245. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25-29.
246. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. 2017;45:D331-D338.
247. Singh J. The portal for rare diseases and orphan drugs. *J Pharmacol Pharmacother*. 2013;4:168-169.
248. Piñero J, Bravo À, Queralt-Rosinach N, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2017;45:D833-D839.
249. Kibbe WA, Arze C, Felix V, et al. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res*. 2015;43:D1071-D1078.
250. Schriml LM, Arze C, Nadendla S, et al. Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res*. 2012;40:D940-D946.
251. Oellrich A, Collier N, Groza T, et al. The digital revolution in phenotyping. *Brief Bioinform*. 2016;17:819-830.
252. de Souza HSP, Fiocchi C, Iliopoulos D. The IBD interactome: an integrated view of aetiology, pathogenesis and therapy. *Nat Rev Gastroenterol Hepatol*. 2017;14:739-749.
253. Peters LA, Perrigoue J, Mortha A, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Gene*. 2017a;49:1437-1449.

How to cite this article: Pazmandi J, Kalinichenko A, Ardy RC, Boztug K. Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms. *Immunol Rev*. 2019;287:162–185. https://doi.org/10.1111/imr.12726

1.3. Network medicine and network-based methods

Traditionally, scientific investigation has followed a reductionist approach, that analyzed genes, proteins, molecules and reactions individually, focusing on their local effects. This approach has proven to be valuable, and has culminated in a catalogue of genes, proteins and other molecules (Lander et al. 2001; Venter et al. 2001). More recently, technological advances resulted in the development of high-throughput technologies and high quality data on our genome, transcriptome and epigenome. The increase of available data provides a catalogue of molecular building blocks of cells that has never been more complete. The traditionally used single-gene and pathway focused approaches, however, are no longer sufficient to delineate and explore the consequences of pathobiological processes of diseases and therapeutic approaches in this catalog of vast data.

Indeed, the molecules within the cell do not carry out their function in isolation but through interactions with other molecules. The health of an organism is influenced by a multitude of intertwined processes and perturbations of these interactions. Similar to social and technological systems, biological systems are linked by a multitude of interconnected relationships that are organized by basic principles. As outlined in the introduction above, on a cellular level, an interaction between an antigen presenting cell and a lymphocyte can elicit a cascade of immune responses that send the body into a global inflammation. On a molecular level, single interacting receptor and ligand molecules can induce a signaling event that leads to an activation of pathway, a cell, and an organ downstream. These biological interactions, chains and links span many orders of magnitude in space and time, collectively forming the basis of life. Just like drawing the layout of a city in a map, we can map out these biological interactions as different networks within networks (Prulj 2019). The molecular interactions and networks that govern our cells have been extensively mapped out in the last two decades. Although these networks are still incomplete, they have been increasingly used as tools for addressing fundamental questions in health and disease.

1.3.1. Introduction to network medicine

The connectivity of molecules and building blocks of cells implies that the impact of a genetic lesion is not only restricted to the protein encoded by a specific gene, but that it can spread along the interconnected molecular network. Indeed, in contrast to the classical “one-gene, one-

function, one-phenotype" approach that gained popularity in the 1940s (Beadle and Tatum 1941) it is increasingly clear that the effect of a particular genetic malformation does not only impact the function of the direct gene product, but also influences a multitude of other interactions between other, often less obvious processes (Barabási, Gulbahce, and Loscalzo 2011).

Network medicine, first introduced in 2007 by Albert-László Barabási (Barabási 2007), applies network science to analyze, identify and understand human diseases (Barabási, Gulbahce, and Loscalzo 2011a). It aims to give an all-encompassing, systematic view of health and disease through the analysis of various biological networks. Diseases can be mapped to biological networks by their causal or associated mutations, by transcriptional signatures or other means. The interaction partners and the network context therefore determine the phenotypic impact of genetic lesions or signatures. This allows us to better understand the complexity of diseases, and in turn disease phenotypes can be better mapped out and understood with biological networks. Network-based approaches to human disease have numerous biological and clinical applications. Our improved understanding of the complexity of the interconnected molecular networks underlying disease entities can help us understand disease progression and identify vital disease-related pathways and genes. This could offer better targets for drug development (Z.-C. Li et al. 2016; Yildirim et al. 2007), drug repurposing (Cheng et al. 2018; J. Li et al. 2016) and the discovery of more accurate biomarkers (J. Zhang et al. 2010; L.-X. Wang, Li, and Chen 2018). Due to their ability to integrate large-scale datasets, network-based methods hold the potential to aid with disease classification as well as pave the way to truly personalized therapeutic approaches and personalized medicine.

1.3.1.1. Biological networks and their applications

As more and more key players of cellular processes are identified through high-throughput studies, our repertoire of available data for building biological networks that model cellular systems grows as well. This chapter discusses the main biological network types and their most common applications.

1.3.1.1.1. Protein-protein interaction networks

Proteins are one of the basic building blocks and acting agents within and around cells. Physical interactions between proteins are essential for normal function of the cell, and aberrations in such interactions or the proteins within the interactions have particularly strong effects. Physical interactions of proteins can be mapped out onto protein interaction networks. The systematic charting of these physical protein-protein interactions is an ongoing process, and although the mapping is not yet complete, a significant subsection of these interactions have been identified already. Two main experimental techniques have been used to establish such networks: yeast two-hybrid (Y2H) assays that map out precise binary interactions (Rolland et al. 2014) , and binding affinity purification coupled to mass spectrometry, where a bait protein and its interactions are captured by beads and large complex protein structures can be identified (Huttlin et al. 2015, 2017). Additional to these two methods, other protein capturing techniques such as co-immunoprecipitation, X-ray crystallography or nuclear magnetic resonance (Menche et al. 2015), or computational prediction of protein interactions from amino acid sequences (Ofran and Rost 2003; Gallet et al. 2000; Yan, Dobbs, and Honavar 2004; Deng et al. 2002), gene fusions (Marcotte and Marcotte 2002), or phylogenic trees (Pellegrini et al. 1999). Each of these methods has advantages and disadvantages in terms of their comprehensiveness, noise and biases towards certain interactions (Gillis, Ballouz, and Pavlidis 2014; Rolland et al. 2014).

Various online repositories exist that collect protein-protein interaction data such as the IntAct maintained by EMBL-EBI (Kerrien et al. 2012), the Human Integrated Protein Protein Interaction rEference (HIPPIE) (Alanis-Lobato, Andrade-Navarro, and Schaefer 2016), or the Search Tool for Recurring Instances of Neighboring Genes (STRING) (Snel et al. 2000). Protein interaction networks have various applications and have been utilized for research on rare diseases specifically. They have been used to investigate basic principles of cellular function (Jeong et al. 2001), the efficacy of drugs, and in particular for predicting novel disease genes (Vanunu 2009; Oti et al. 2006), a topic that will be discussed in detail below.

1.3.1.1.2. Metabolic networks

Metabolism is the collection of chemical processes within a cell, and metabolic networks are one of the most extensively studied biological networks. In these maps, metabolites serve as nodes

and they are linked by chemical reactions. There are numerous databases of metabolic networks and maps such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (M. Kanehisa and Goto 2000; Minoru Kanehisa et al. 2008), the Human Recon 2.2 (Swainston et al. 2016) and the Edinburgh human metabolic network (Ma et al. 2007). Metabolic networks are informative to investigate diseases that arise as a result of over-amplification of certain metabolic pathways, such as in type 2 diabetes mellitus (Zelezniak et al. 2010).

1.3.1.1.3. Gene regulatory networks

Gene expression is a dynamic process that requires tight regulation. The most basic gene regulatory network models consist of nodes that represent transcription factors and their target DNA regulatory elements (Hecker et al. 2009). These networks are built on experimentally verified genetic regulatory interaction data, from databases such as JASPAR (Sandelin et al. 2004) and TRANSFAC (Matys et al. 2003). In addition to transcription factors and regulatory elements, gene expression is also regulated through interactions between RNAs, or DNA and RNA. Databases such as TargetScan (Edris 2011), PicTar (Martín et al. 2018), microRNA, miRWalk (Sticht et al. 2018) store computationally predicted interactions between these elements, while TarBase (Vergoulis et al. 2012) and miRecords (Xiao et al. 2008) contain experimentally verified interaction data.

1.3.1.1.4. Non physical interaction networks: coexpression and genetic interaction networks, co-perturbation networks

Our understanding of gene regulatory networks remains incomplete as they are highly context-dependent and involve the interaction of a large number of different molecules. A more straight-forward quantity that can serve as a proxy to study regulation of genetic material is coexpression. Two genes are coexpressed if their expression levels correlate under various conditions, such as in a disease, over time, or across certain stimuli. The expression of genes can be assessed genome-wide with RNAseq (B. Zhang and Horvath 2005; De Smet and Marchal 2010). The GTEx consortium (Aguet et al., n.d.) stores curated expression data across various cell and tissue types. Coexpression networks are context specific, and connect two genes if they are coexpressed in a certain context. In contrast to gene regulatory networks, they are undirected and do not imply a

causal relationship between connected genes. Coexpression networks can be used to pinpoint functionally related genes that are controlled by similar transcriptional regulation, or members of the same pathway (Weirauch 2011). In a disease context, coexpression networks have been used for the analysis of inflammatory bowel disease, autism spectrum disorder, Alzheimer's disease, and cancer (Parikshak et al. 2016; B. Zhang et al. 2013; Peters et al. 2017; S. Zhang et al. 2017; J. Zhang et al. 2010; L.-X. Wang, Li, and Chen 2018).

Genetic interaction networks map out phenotypes that arise from a simultaneous mutation of two genes. Specifically, a negative genetic interaction between two genes implies lethality if both genes are mutated but no lethality if only one of the genes is mutated. A positive genetic interaction arises if a mutation in one of the interacting genes rescues a lethal mutation in the other interacting gene (Costanzo et al. 2016). Large yeast-based screens have been used to identify such interactions (Costanzo et al. 2016; Tong et al. 2004). Genetic interaction networks have been used to investigate genetic therapies such as in FA (Moder et al. 2017), or for cancers to identify targets for chemotherapy (Srivas et al. 2016). Co-perturbation networks encapsulate information from various perturbation screens such as RNAi screens (Kim and Rossi 2008), CRISPR screens (Doench et al. 2016), or drug screens (Kubicek et al. 2012; Bansal et al. 2014; Markt et al. 2012). In these networks nodes represent genes and edges stand for correlations of two genes in response to the perturbation studies. These networks have also been used to investigate molecular mechanisms of drugs (F. Zhang et al. 2013; Noh, Shoemaker, and Gunawan 2018), predict drug targets (Isik et al. 2015), or pathway activity (Schubert et al. 2018; Dorel et al. 2018; Molinelli et al. 2013).

Disease networks represent linkedness of diseases that can occur on several scales: on the molecular level (by linking diseases through shared genetic origin), phenotypic level (by connecting diseases if they share phenotypes) or on the population level (by adding edges between diseases based on their co-occurrence, or comorbidity). The first published comprehensive map of the linkedness of human diseases was termed the "diseaseome" (Goh et al. 2007), based on disease-associations from the OMIM (Amberger et al. 2018) database. The diseaseome revealed that most diseases are not isolated entities, but fall into highly connected clusters with common molecular roots. Disease networks based on phenotypic similarity have been published (Zhou et al. 2014), as well as population-level disease networks based on comorbidity (Chmiel, Klimek, and Thurner 2014; Hidalgo et al. 2009).

1.3.1.1.5. Interactomes

The term “interactome” is loosely defined and usually refers to networks that contain different types of interactions, in order to represent the entirety of molecular interactions within and across cells. The interactions within interactomes can be generally categorized into direct physical interactions, and indirect functional interactions. Sources of these interactions can be:

(1) Literature-curated interactions from small-scale experiments,
(2) Interactions from large-scale interaction mapping efforts,
(3) Computationally predicted interactions.

Interactomes are often the go-to network to use for disease analyses. For this, various interactomes have been curated and published so far that have shown utility to unravel disease-related traits (Menche et al. 2015; Luck et al. 2020).

1.3.2. Interactomes as maps, network-based properties

Although built on an array of diverse data, biological networks and interactomes have been shown to have certain common features. There are three features of networks that form the basis of viewing molecular networks as maps of cellular function (Figure 5).

(1) On an individual node basis, the position of a node is linked to its importance in the specific biological system represented in the network.
(2) On a larger level, the connectivity of a group of nodes in a network can be associated with shared biological function.
(3) Taking a step further back, the distance between individual groups of nodes can indicate their functional relatedness.

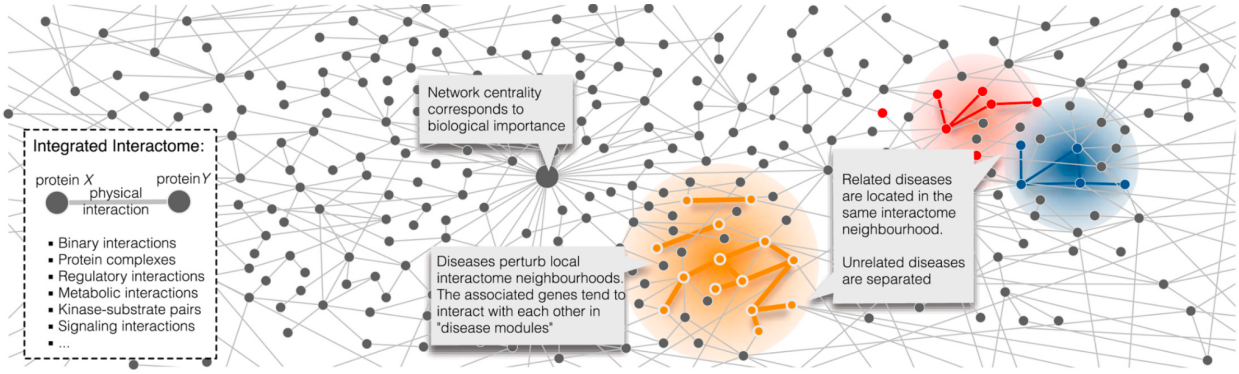


Figure 5. Biological networks as maps of cellular function. Network centrality, disease modules, network neighborhoods and their relative positions are showcased. Reprinted by permission from Elsevier: Current Opinion in Systems Biology, “Interactome-based approaches to human disease” (Caldera et al. 2017).

The next chapter of the introduction discusses some of the most commonly used network properties.

1.3.2.1. General network properties

Biological networks can be described as collections of *nodes N* or vertices, their interactions/connections as *edges E* or links. Networks can be *undirected*: here, two nodes are connected by an undirected link if there is an interaction between them. In *directed* networks each interaction has a source and a target, such as in gene regulatory networks: “gene A inhibits gene B”. Protein-protein interaction networks are traditionally undirected. Most often these links are *unweighted* (all links are equal in importance) and indicate a yes or no relationship. In contrast, in *weighted networks* not all links are equal and each link carries an additional property. Examples for weighted networks are metabolic networks or coexpression networks. *Bipartite networks* are built from two different types of nodes, such that interactions only occur between nodes of different types. Examples of these are disease-gene networks (Goh et al. 2007; Zhou et al. 2014).
The total number of nodes *(N)* determines *network size*, the total number of links is denoted as *(E)*. Most networks are sparse, i.e., only a small number of all possible links are present. The number of links a node has is termed its *degree (k)*. *Hubs* in a network are nodes with a high number of connections. Hubs have been shown to show particular biological importance as they can be essential genes that are vital for the survival of cell lines (Blomen et al. 2015). The *degree distribution* of a network is the frequency distribution of all degrees over the whole network. In

scale free networks, the degree distribution follows a power law, that is the fraction of (k) nodes in the network having k degrees goes for large values of k as

$$P(k) \sim k^{-y}$$

where y is a parameter with a value typically ranging from 2 < y < 3 (Onnela et al. 2007; Choromański, Matuszak, and Miękisz 2013). Interactomes are often scale free. Scale free networks exhibit the *"small world effect" (Cohen and Havlin 2003),* a phenomenon whereby hubs in the network connect distinct parts of the network, shortening the average distance between nodes. A *path* in the network is a sequence of links that connect two nodes (A and B) in the network. The number of steps gives the length of a path. *The shortest path length* determines the network-based *distance d* as the minimal number of links connecting two nodes A and B. The *diameter* of a network is the longest of all shortest paths between any two nodes in the network.

A *subgraph* of a network is a subset of particular nodes in a network with all of the edges linking the nodes. A *connected component* of a particular network or subgraph is a subgraph with no loose nodes, i.e., there is a path connecting every node to every other node in the connected component. The *largest connected component* is the connected component of a subgraph or network with the highest number of nodes.

1.3.2.2. Centralities

The *centrality* of a node measures its topological importance in the network. There are different ways of quantifying centrality, all focusing on different attributes of importance. Here, the four main centrality measures are discussed. The conceptually simplest centrality measure is *degree centrality,* that is defined as the number of links or connections of a node in a network. Hubs in networks have the highest degree centrality. *Closeness centrality* of a node is the average length of the shortest paths between the node and all other nodes in the graph. A node with high closeness centrality is therefore a node that is close to all other nodes (Bavelas 1950; Sabidussi 1966). *Betweenness centrality,* first introduced as a measure for quantifying the control of a human on the communication between other humans in a social network (Freeman 1977), measures the number of shortest paths going through a node in a given network. *Eigenvector centrality*, or *eigencentrality* is a measure of influence of a particular node over a network. Eigenvector centrality assigns centrality scores to all nodes in a network based on their degree,

and is calculated based on a node's connections to high-scoring nodes (Newman 2008). The *Katz* (Katz 1953) and *PageRank* centralities are variations of eigenvector centrality. It has been shown that not only do centrality measures pinpoint topologically important nodes in networks, but that they can also be used to identify biologically important nodes, as cancer driver genes have been shown to be central in various networks (Piñero et al. 2016).

1.3.2.3. Distance measures

As the relative position of nodes and groups of nodes on a network is of particular importance from a biological perspective, the quantification of distance on networks is of interest in a multitude of research endeavors.

Shortest distance, or the *shortest path* between two nodes on a network is defined as the connecting path between the two nodes with the minimum number of edges. The *average distance* between two groups of nodes on a network is calculated by averaging the pairwise shortest distances between each node-node pair in the two sets of nodes. The *minimum distance* between two sets of nodes on a network is defined as the shortest of all pairwise shortest distances between the two sets.

Centre distance measure of two sets of nodes is defined as the shortest path between the nodes with highest closeness centralities in the two node sets. *Network separation* calculates the sum of mean distances between the two sets of nodes on the network using the average distance measure and subtracts it from the average shortest distance within the two sets of nodes (Menche et al. 2015).

1.3.2.4. Connectivity measures, modules

Network modules denote a group of nodes in a network with dense connections among themselves. Nodes within a specific module are often functionally related (Spirin and Mirny 2003; Hartwell et al. 1999; Barabási and Oltvai 2004), belong to the same pathway, or are coexpressed (Huttlin et al. 2017; Rolland et al. 2014).

Disease modules are usually defined as a set of molecular components and their interactions associated with a certain disease. Although network modules have been identified in groups of disease-related proteins (Feldman, Rzhetsky, and Vitkup 2008), the connectivity pattern of disease-associated nodes has unique properties as compared to the dense aggregation of nodes

in a network linked to a specific biological function (Ghiassian, Menche, and Barabási 2015). This observation led to the hypothesis that *dys*function is usually distributed among loosely connected functional modules on the interactome. The identification of disease modules in a network therefore requires slightly different approaches than common community detection methods that are used to identify functionally related groups of nodes (Fortunato 2010).

The existence of connected disease modules over the past decade served as the basis hypothesis of interactome-based approaches to human disease. Commonly, disease modules were defined as the connected subgraph of disease-associated proteins within the interactome (Barabási, Gulbahce, and Loscalzo 2011a). Since this discovery, several other methods have been proposed for the identification of disease-related modules on networks, as well as for the analysis of the connectedness and relationships between disease modules (Menche et al. 2015; Ghiassian, Menche, and Barabási 2015). The flagship discovery that disease neighborhoods associated with complex traits are significantly linked and reside in specific network neighborhoods (Menche et al. 2015) has paved the way for a broad application of interactome-based methods for disease discovery.

1.3.3 Network medicine for diseases

As outlined above, recent years have shown ample evidence that on interactomes, connectedness and function are intimately linked. This observation has made interactomes attractive models in elucidating fundamental questions in the research on a wide-variety of diseases. Interactomes have been applied to predict putative function and disease-relatedness of nodes, to elucidate disease-disease relationships and to investigate the efficacy and identify novel therapeutic approaches for diseases (Barabási, Gulbahce, and Loscalzo 2011; Caldera et al. 2017). This chapter discusses the various applications of network medicine to elucidate various questions regarding disease pathobiology.

1.3.3.1. Disease gene identification

Multiple interactome-based disease-module and disease gene identification methods have been proposed in recent years. These methods commonly explore the network neighborhood of

previously identified disease-associated genes, so called “seed genes” to identify novel disease-associated entities (X. Wang, Gulbahce, and Yu 2011).

a) *Path-based approaches* use shortest path-based measurements to rank the putative candidates on the network in relation to known disease genes (George et al. 2006; Dezső et al. 2009).
b) *Dynamical approaches* pinpoint novel candidate genes using dynamic propagation such as diffusion methods around previously identified disease modules. These diffusion methods include random walk-based approaches that expand around a set of seed genes following their links (Krauthammer et al. 2004; Vanunu 2009; Vandin, Upfal, and Raphael 2011; Smedley et al. 2014), ranking those nodes higher that have been visited more frequently by the walker.
c) *Connectivity-based methods* rely on the connectivity of candidate nodes to the already identified disease module (Guney and Oliva 2012; X.-D. Wang et al. 2014; Ghiassian, Menche, and Barabási 2015).

1.3.3.2. Disease-disease relationships

Much like individual nodes in a network, due to the interconnectedness of biological systems, diseases cannot be fully understood as isolated entities. Shedding light on the molecular connections that link diseases can help us uncover the molecular links of pathobiological phenotypes and disease comorbidity (Hu, Thomas, and Brunak 2016; Zhou et al. 2014).
Interactomes have been used to identify the shared genetic architecture of diseases, revealing that more than 800 diseases have at least one genetic link to another disease (Zhou et al. 2014). The relationship between two diseases can be represented as the tendency of their disease-modules to overlap in an interactome-based framework. A study of 44,551 disease pairs identified that the degree of this disease-module overlap is indicative of the pathobiological similarity of diseases on a phenotypical, gene expression and comorbidity level (Menche et al. 2015). Disease classification, which is historically based on clinicopathological traits of diseases and often grouped according to organ system, can also benefit from interactome-based methods that rely on the molecular similarity of disease entities (Chan and Loscalzo 2012). In addition to interactomes, metabolic pathways (Lee et al. 2008), phenotype similarity (Zhou et al. 2014), disease ontologies (van Driel et al. 2006), and comorbidity data (Hidalgo et al. 2009; Klimek, Aichberger, and Thurner 2016) have been used to build disease-disease networks.

1.3.3.3. Drug efficacy

From an interactome-based perspective, the effect of drugs - similarly to diseases - can be understood as local perturbations in the network neighborhood of the drugs and their targets. Network pharmacology therefore has applied network-medicine concepts detailed above to investigate the effect of drugs (Hopkins 2008; Csermely et al. 2013) on biological systems. Through these studies, it has been shown that most drugs have more targets than previously though, that these targets are highly connected (Yildirim et al. 2007; Keiser et al. 2009), and that most drugs either only target a small subset of a particular disease module, or its adjacent network neighborhood (Guney et al. 2016).
Furthermore it has been shown that drugs whose target network neighborhood is closer to a particular disease module have been found to be more effective in the clinic (Z.-C. Li et al. 2016). This discovery has prompted the use of interactome based methods for predicting or prioritizing novel drugs and drug targets for diseases based on their proximity on the interactome (Z.-C. Li et al. 2016; Csermely et al. 2013), as well as drug repurposing by identification of diseases with shared molecular background that may be treated by similar therapeutic approaches (J. Li et al. 2016).

1.3.3.4. Specific interactomes

The majority of interactome-based methods rely on global interactomes that contain an aggregate of interactions between molecules that have been identified under different experimental and biological conditions. While these interactomes provide informative landscapes for investigating general principles of disease pathogenesis and cellular organization, it has also become clear that context-specific interactomes that represent certain cellular states, cell types or tissues are needed to answer more specific questions, especially for diseases with particular tissue or cell specific representations.

Context specific expression data can be overlaid on global interactomes to approximate cell type, tissue or disease specificity (Fagerberg et al. 2014; Melé et al. 2015; Yeger-Lotem and Sharan 2015). These specific interactomes are usually smaller than the global network and contain less links. A general observation emerging from the use of specific interactomes is that housekeeping

genes - genes that are widely expressed across different tissue types and states - have been found to form a core interactome (Bossi and Lehner 2009; W. Liu et al. 2014; Barshir et al. 2014). It has been shown that the connectedness of disease modules over different tissue-specific networks is indicative of their presentation in the respective tissues (Kitsak et al. 2016). Moreover, tissue and cell specific networks have been shown to improve disease-gene prioritisation (Barshir et al. 2014; Magger et al. 2012; M. Li et al. 2014).

2 Aims of the thesis

Rare diseases of the immune system are a heterogeneous group of diseases that, collectively, affect a considerable fraction of the population, pose a significant demand on the healthcare system, and require tailored scientific approaches. While systems-biology and network-based tools have been applied to various diseases to answer a multitude of questions, e.g., to identify key drivers in diseases, a general outline on how to use them to address a specific group of rare disorders is still lacking. These systems-based methods are needed to objectively assess characteristics of diseases, and are useful not only to describe these diseases in an accurate fashion but also to provide a general overview of represented phenotypes, and to quantify the relationships between the rare diseases themselves and other diseases beyond. The aim of this thesis is to provide a general overview of the challenges in research on rare diseases of the immune system and showcase how tools of systems biology, in particular machine learning on clinical datasets and network-based methods, can be used to address the gaps regarding the available knowledge.

In the first research manuscript, presented in results section 3.1 the thesis introduces our effort on expanding and optimizing the available phenotypic representation of IEI using HPO, in the manuscript titled “Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity”. We developed a method with the goal to systematically revise and expand the IEI-relevant phenotype information, by working on the completion of the HPO ontology in an expert-driven framework. In addition, we aimed to assess the efficacy of our approach using a real-life based diagnostic challenge.

Finally, in the second research article included in the results section 3.2. titled “AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation”, we used systems-biology methods to chart the molecular landscape of rare monogenic autoimmune and autoinflammatory diseases. Our goal was to construct a novel, network-based framework that could be leveraged to unravel objective pathobiological properties of monogenic autoimmunity/autoinflammation, to identify molecular subclusters of these diseases and to identify novel targetable therapeutic pathways.

3 Results

3.1. Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity

Matthias Haimel PhD*, Julia Pazmandi MSc*, Raúl Jiménez Heredia MSc, Jasmin Dmytrus PhD, Sevgi Köstel Bal M.D.,PhD, Samaneh Zoghi PhD, Paul van Daele M.D., Tracy A. Briggs PhD, Carine Wouters M.D., Brigitte Bader-Meunier M.D., Florence A. Aeschlimann M.D., Roberta Caorsi M.D., Despina Eleftheriou M.D., Esther Hoppenreijs M.D., Elisabeth Salzer M.D.,PhD, Shahrzad Bakhtiar M.D., Beata Derfalvi M.D., Francesco Saettini M.D., Maaike A. A. Kusters M.D.,PhD , Reem Elfeky M.D., Johannes Trück M.D., Jacques G. Rivière M.D., Mirjam van der Burg PhD , Marco Gattorno M.D., Markus G. Seidel M.D., Siobhan Burns M.D., Klaus Warnatz M.D., Fabian Hauck M.D.,PhD , Paul Brogan M.D., Kimberly C. Gilmour PhD, Catharina Schuetz M.D., Anna Simon M.D.,PhD, Christoph Bock PhD , Sophie Hambleton PhD , Esther de Vries PhD , Peter Robinson M.D., Marielle van Gijn PhD †#, Kaan Boztug M.D.†#. 2021. The Journal of Allergy and Clinical Immunology. Published:May 11,2021. DOI: https://doi.org/10.1016/j.jaci.2021.04.033 * and †, these authors contributed equally # to whom correspondence should be addressed.

We have initiated an international collaboration of experts on immune mediated disorders in order to review, revise and complete the HPO for IEI. This manuscript showcases our results regarding the revision of the HPO tree and the reannotation of IEI based on four major disease groups. First, we have used a hands-on approach to expand the existing HPO tree with novel terms relevant for IEI. Next, we used an ontology-guided machine learning approach, coupled with expert review to reannotate IEI with HPO terms. We show that both of these approaches resulted in a significant gain in available HPO terms per disease, and a more complete phenotypic annotation of diseases that has improved disease-disease and patient-disease similarity matching.

The introduction of our initiative, working structure and methods, as well as our results and their validation using a cohort of IEI patients can be found in this publication. In addition, the supplementary materials and methods provide further technical details on our methodology on the computational side.

Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity

Matthias Haimel PhD [1-3]*, Julia Pazmandi MSc [1-3]*, Raúl Jiménez Heredia MSc [1-3], Jasmin Dmytrus MSc [1-3], Sevgi Köstel Bal M.D.,PhD [1-3], Samaneh Zoghi PhD [1-3], Paul van Daele M.D.[4], Tracy A. Briggs PhD [5,6], Carine Wouters M.D.[7,8], Brigitte Bader-Meunier M.D.[9,10], Florence A. Aeschlimann M.D.[9,10], Roberta Caorsi M.D.[11], Despina Eleftheriou M.D.[12,13], Esther Hoppenreijs M.D.[14], Elisabeth Salzer M.D.,PhD [1-3], Shahrzad Bakhtiar M.D.[15], Beata Derfalvi M.D.[16], Francesco Saettini M.D.[17], Maaike A. A. Kusters M.D.,PhD [12,13], Reem Elfeky M.D.[12,13], Johannes Trück M.D.,Phil [18], Jacques G. Rivière M.D.[19,20], Mirjam van der Burg PhD [21,22], Marco Gattorno M.D.[11], Markus G. Seidel M.D.[23], Siobhan Burns M.D.[24], Klaus Warnatz M.D.[25,26], Fabian Hauck M.D.,PhD [27,28], Paul Brogan M.D.[12,13], Kimberly C. Gilmour PhD [13], Catharina Schuetz M.D.[29], Anna Simon M.D.,PhD [30], Christoph Bock PhD [1,3,31], Sophie Hambleton PhD [32], Esther de Vries M.D., PhD [33,34], Peter Robinson M.D.[35], Marielle van Gijn PhD [36]†#, Kaan Boztug M.D.[1-3,37]†#

* and †, these authors contributed equally

to whom correspondence should be addressed:

Kaan Boztug, Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD) and St. Anna Children's Cancer Research Institute (CCRI), Zimmermannplatz 10, A-1090 Vienna, kaan.boztug@rud.lbg.ac.at, Phone: +43 1-40470-4080, Fax: +43-1-40170-7280

Marielle Van Gijn, Department of Genetics, University Medical Center Groningen, Antonius Deusinglaan 1, 9713AV Groningen, Netherlands, m.e.van.gijn@umcg.nl, +31-55256416

[1]Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Vienna, Austria
[2]St. Anna Children's Cancer Research Institute, Vienna, Austria
[3]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria
[4]Department of Clinical Immunology, Erasmus University Medical Center, Rotterdam, The Netherlands
[5]NW Genomic Laboratory Hub, Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Manchester, United Kingdom
[6]Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, United Kingdom.
[7]Department of Microbiology and Immunology, Immunobiology, KU Leuven, Leuven, Belgium

[8]Department of Pediatrics, Division of Pediatric Rheumatology, University Hospitals Leuven, Leuven, Belgium
[9]Pediatric Immuno-Hematology and Rheumatology Unit, Necker Hospital for Sick Children - AP-HP, Paris, France, EU.
[10]Reference Center for Rheumatic, Autoimmune and Systemic Diseases in Children (RAISE), Paris, France
[11]Center for Autoinflammatory diseases and Immunodeficiency, IRCCS Istituto Giannina Gaslini, Genova, Italy
[12]University College London Great Ormond Street Institute of Child Health, London, United Kingdom
[13]Department of immunology, Great Ormond Street (GOS) Hospital for Children NHS Foundation Trust, London, United Kingdom
[14]Department of Paediatric Rheumatology, Radboud University Medical Centre, Nijmegen, The Netherlands
[15]Department for Children and Adolescents, Division for Stem Cell Transplantation, Immunology and Intensive Care Unit,Goethe University, Frankfurt, Germany
[16]Deptment of Pediatrics, Division of Immunology, Dalhousie University/IWK Health Centre Halifax, Nova Scotia, Canada
[17]Pediatric Hematology Department, Fondazione MBBM, University of Milano Bicocca, via Pergolesi 33, 20900, Monza, Italy
[18]Division of Immunology, University Children's Hospital Zurich, Switzerland
[19]Pediatric Infectious Diseases and Immunodeficiencies Unit, Vall d'Hebron Research Institute, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain
[20]Jeffrey Model Foundation Excellence Center, Barcelona, Spain
[21]Department of Immunology, University Medical Center Rotterdam, Rotterdam, The Netherlands
[22]Laboratory for Pediatric Immunology, Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands
[23]Research Unit for Pediatric Hematology and Immunology, Division of Pediatric Hemato-Oncology, Department of Pediatrics and Adolescent Medicine, Medical University Graz, Graz, Austria
[24]Department Immunology, UCL Institute of Immunity & Transplantation, Department of immunology, Royal Free Hospital NHS Foundation Trust, Pond Street, London, NW3 2QG, UK
[25]Division of Immunodeficiency, Department of Rheumatology and Clinical Immunology, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[26]Center for Chronic Immunodeficiency (CCI), Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[27]Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany
[28]Munich Centre for Rare Diseases (M-ZSE^{LMU}), University Hospital, Ludwig-Maximilians-Universität München, Munich, Germany
[29]Department of Pediatrics, Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden, Germany
[30]Radboudumc Expertise Centre for Immunodeficiency and Autoinflammation (REIA), department of Internal Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

[31]Institute of Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Vienna, Austria
[32]Immunity and Inflammation Theme, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom
[33]Tranzo, Tilburg University, Tilburg, The Netherlands
[34]Laboratory for Medical Microbiology and Immunology, Elisabeth-Tweesteden Hospital, Tilburg, The Netherlands
[35]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06032, USA
[36]Department of Genetics,University Medical Center Groningen, Groningen, The Netherlands
[37]Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria

Competing interests: The authors declare no conflict of interests.

Funding: The work was supported by the European Research Council (ERC Consolidator Grant 820074 "iDysChart" to K.B. Additional financial support for the workshops was granted by the Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD), the European Reference Network on Rare Primary Immunodeficiency, Autoinflammatory and Autoimmune diseases (ERN-RITA), and the European Society for Immunodeficiencies (ESID).

Author contributions: MH, JP, MVG, KB: study design and manuscript writing. MH, JP: data acquisition, coordination of working groups. MH, JP analysis and interpretation of data. SH, clinical cohort data extraction. KB, MVG: Study supervision. All co-authors participated in the meetings and revision of terms. The manuscript was reviewed, edited and approved by all co-authors.

Abstract

BACKGROUND: Accurate, detailed and standardized phenotypic descriptions are essential to support diagnostic interpretation of genetic variants and to discover new diseases. The Human Phenotype Ontology (HPO), extensively used in rare disease research, provides a rich collection of vocabulary with standardized phenotypic descriptions in a hierarchical structure. However, to date the use of HPO has not yet been widely implemented in the field of inborn errors of immunity (IEIs), mainly due to a lack of comprehensive IEI-related terms.

OBJECTIVES: We sought to systematically review available terms in HPO for the depiction of IEIs, to expand HPO yielding more comprehensive sets of terms, and to reannotate IEIs with HPO terms to provide accurate, standardized phenotypic descriptions.

METHODS: We initiated a collaboration involving expert clinicians, geneticists, researchers working on IEIs and bioinformaticians. Multiple branches of the HPO tree were restructured and extended based on expert review. Our ontology-guided machine learning coupled with a two-tier expert review was applied to reannotate defined subgroups of IEIs.

RESULTS: We revised and expanded four main branches of the HPO tree. Here, we reannotated 73 diseases from four IUIS-defined IEI disease subgroups with HPO terms. We achieved a 4.7-fold increase in number of phenotypic terms per disease. Given the new HPO annotations, we demonstrated improved ability to computationally match selected IEI cases to their known diagnosis, and improved phenotype-driven disease classification.

CONCLUSION: Our targeted expansion and reannotation presents enhanced precision of disease annotation, will enable superior HPO-based IEI characterization and hence benefit both IEI diagnostic and research activities.

Key message

HPO is a robust resource for supporting IEI diagnostics and genetics with adequate ontology breadth and disease annotation depth.

Capsule Summary

Our newly formed expert consortium systematically reviewed and expanded existing HPO terms of IEIs and reannotated IEIs with HPO terms. This will support diagnostic pipelines and analysis of variants from next-generation sequencing.

Key words

HPO; ontology; phenotype; rare diseases; inborn errors of immunity; immune deficiencies; disease classification; diagnostic support; patient matching; genetic analysis

Abbreviations

ALPS - Autoimmune Lymphoproliferative Syndrome
CVID – Common Variable Immunodeficiency Disorders
EBV – Epstein-Barr Virus
EHR - Electronic Health Record
ERN-RITA - European Reference Network on Rare Primary Immunodeficiency; Autoinflammatory and Autoimmune diseases
ESID - European Society for Immunodeficiencies
HLH - Hemophagocytic Lymphohistiocytosis
HPO - Human Phenotype Ontology
IEI - Inborn Errors of Immunity
IUIS – International Union of Immunological Societies
ISSAID - International Society of Systemic Autoinflammatory Diseases
LBI-RUD - Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases
OMIM – Online Mendelian Inheritance in Man
PAD – Primary Antibody Deficiencies
SCID – Severe Combined Immunodeficiency
TRAPS - Tumor necrosis factor receptor-associated periodic syndrome
UDNI - Undiagnosed Diseases Network International
UDP and UDN - Undiagnosed Disease Program and Network

Introduction

Rare and undiagnosed diseases pose challenges for affected patients, clinicians and researchers working to improve diagnostic and therapeutic approaches. Because of the rarity, clinicians often only see a few patients with specific rare phenotypes throughout their careers, leading to considerable diagnostic delay (*1*). Genetic research on rare diseases often relies on single pedigrees or a few patients, leaving many patients undiagnosed (*1*). Compiling a cohort of patients - so-called patient matching - is often crucial to gain insight into the phenotypic spectrum, natural/clinical history of the disease, and adequate monitoring and treatment strategies. The rare disease community has recognized these challenges and established tools enabling efficient data sharing across institutions and borders, including genetic data exchange through the Matchmaker Exchange platform (*2*) to solve undiagnosed exomes and genomes (*3*). These platforms however are highly dependent on accurately phenotyped and categorized patients and standardized disease classifications.

To date, several nomenclatures and reference systems for diseases have been developed (*4,5*). In parallel, ontologies were established to provide a more systematic, hierarchical classification of diseases (*6,7*). However, these nomenclatures group patients by disease label and do not describe the underlying phenotypic features. Consequently, clinical features, laboratory measurements, anatomical and functional phenotypes of patients are often described with variable quality and specificity, which hampers patient matching, diagnostic efficiency, genetic variant prioritization in diagnostic pipelines and global data exchange.

Given these challenges and the need for accurate, standardized phenotyping, the Human Phenotype Ontology (HPO) system was conceptualized and published with initial terminology in 2008 (*8,9*). To date, HPO provides the most comprehensive deep phenotyping resource for rare diseases for clinicians, researchers, bioinformaticians and electronic health record (EHR) systems in the world. HPO is used in many projects including the 100,000 Genomes Project, the NIH Undiagnosed Disease Program and Network (UDP and UDN), the Undiagnosed Diseases Network International (UDNI), RD-CONNECT, and SOLVE-RD (*1,10-13*). HPO is a community-based tool and is increasingly adapted as the standard to describe phenotypic abnormalities for everyday use (*14*). Each term in HPO describes a distinct phenotypic feature (e.g. lymphadenopathy, HP:0002716) and the HPO tree structure allows similarity measures between patient phenotypes. HPO contains over 200,000 phenotypic annotations for hereditary diseases, of which 2,120 are considered rare diseases. Inborn errors of immunity (IEIs) form a subgroup of these rare diseases. Clinical experts in IEI agree that a major barrier to the adoption

of HPO terminology has not been used widely for IEIs partly due to the lack of disease specific HPO terms for IEI patients *(15)*. Adequate depiction of the complex clinical and immunological phenotypes of IEI disease entities with HPO terms would allow discrimination between heterogeneous groups of IEIs. Illustrating the lack of terms, in 2017 HPO contained more than 11,000 terms, out of which 5,000 terms have been applied to the musculoskeletal system, with only 1,000 terms related to IEIs (*9,15*). In addition, the phenotypic annotation of IEIs often includes results of specific immunological assays, which pose a challenge to accurately reflect in HPO terms (*15*). Because of the lack of specific HPO terms depicting results of laboratory assays, often a non-specific broader term is used for the annotation of IEIs. Therefore, HPOs are currently not specific enough to be used for genetic analysis and diagnostic aid for IEIs. In a study addressing the clinical efficacy of genetic testing in IEI, bioinformatics tools using existing HPO terms missed the disease causing gene in 37% of the patients with known monogenic disorders *(16)*. In this study, we set out to improve HPO terminology for IEIs by applying established bioinformatic methodologies coupled with expert review. The aims of this project were therefore to i) systematically review existing HPO terms for IEIs, ii) revise ontology structures, to iii) add missing terms, as well as iv) reannotate existing IEIs with HPO terms, to collectively enable systematic use of HPO by the IEI-community.

Materials and Methods

Spearheaded by the European Reference Network on Rare Primary Immunodeficiency, Autoinflammatory and Autoimmune diseases (ERN-RITA) and the European Society for Immunodeficiencies (ESID), we set up working groups comprising members of the participating immunodeficiency societies to revise and expand HPO terms for IEIs. Three workshops, numerous teleconferences and joint task forces took place over the span of 2 years, with over 30 participants including expert clinicians, geneticists, researchers working on IEIs and bioinformaticians. All participating clinicians and geneticists identified through ERN-RITA, ESID, and the International Society of Systemic Autoinflammatory Diseases (ISSAID) are established experts in their fields from different European countries and North America. Additional scientific support provided the indispensable bioinformatics expertise.

Establishment of working structure

A remote working structure (detailed in the Supplementary Methods) was launched to address gaps in the HPO tree and in the annotation of IEI diseases.

Expansion and restructuring of disease-related branches of the HPO tree

Disease-specific HPO restructuring was discussed within four working groups. Each group focused on a different HPO branch; the suggested changes were agreed on among all participants. Differences between centers and countries in the use of terms and definitions were highlighted during the face-to-face workshops. The results were summarized electronically in Excel documents or pictures and flipchart drawings by the main coordinators before being submitted to HPO. The full list of restructured tree elements and new submitted HPO terms is detailed in the Supplementary Document 1. Additionally, missing terms describing pulmonary and gastro-intestinal complications of primary antibody deficiency (PAD) were discussed during teleconferences and thereafter submitted to update the HPO ontology. A list of HPO resources can be found in Supplementary Materials and Methods.

Standardized reannotation of rare, genetically diagnosed diseases

A four-step process was developed for a standardized reannotation effort across working groups and to consistently annotate IEIs (spanning over 300 different diseases in Online Mendelian Inheritance in Man (OMIM)) with HPO terms (Fig 1). As IEIs represent a large and heterogenous group of rare diseases, we here decided to selectively focus on defined subgroups of IEI to test the feasibility and usefulness of such an endeavor. First, publications were collected by experts for each disease within the subgroups (minimum of two articles per disease), representing key phenotypic presentation(s) of the specific disease. In the second step, HPO terms were extracted from the provided publications for each disease using machine learning ((*17*), explained in detail in Supplementary Materials and Methods) and summarized into Excel documents. Third, a two-tier expert review evaluated the text mined terms, suggested additional terms if required and the responsible working group agreed (defined as at least 80% agreement amongst group experts) on the final HPO annotations for each disease. Fourth, the validated terms were submitted to HPO. Supplementary Document 2 contains the reannotated diseases and the list of reannotated terms for each disease is available in Supplementary Document 3.

Standardized reannotation of genetically undiagnosed diseases

The methods above were specifically designed for application in (very) rare diseases, where the number of patients and therefore the described phenotypic spectrum and clinical presentation is sparse. In case of diseases and disease groups where an adequate amount of patient and phenotype data was available, in addition to a True/False annotation, the frequency of each phenotypic item was assessed. The frequencies correspond to the following representation in patients: common = Frequent (79-30%); sometimes = Occasional (29-5%); rare = Very rare (<4-1%).

Patient cohort

We randomly selected 30 patients that harbored a genetic diagnosis in one of the reannotated diseases from a large pediatric referral center research database. Clinical summaries of these patients prior to genetic diagnosis were retrieved by an expert clinician. The clinical summaries were parsed and HPO terms were extracted using machine learning as in the Supplementary Methods.

HPO information content measures, and disease patient similarity measures

Information content of all HPO terms was assessed with the *R* package ontologyIndex v2.5 (*18*). The phenotypic similarity of diseases and patients before and after reannotation was compared using the *R* package ontologySimilarity v2.3 (*18*). The Euclidean distances between the diseases were computed based on similarity measures, clustered with hierarchical clustering and visualized with ggtree using the *R* packages ggtree (*19*) and ape v5.2 (*20*).

A detailed description including the data processing pipeline and tools are available in the Supplementary Materials and Methods.

Supplementary Materials

Supplementary Materials and Methods
Supplementary Document 1: HPO tree restructuring and the list of new terms
Supplementary Document 2: Summary of diseases reannotated
Supplementary Document 3: List of all terms per disease after reannotation
Supplementary Document 4: List of cases used for phenotype to diagnosis matching

Results

Systematic evaluation and expansion of the HPO structure and terms relevant to IEIs

Our approach has resulted in the restructuring of four main branches of the HPO tree, namely: i) abnormality of the immune system (HP:0002715) ii) abnormality of metabolism/homeostasis (HP:0001939) iii) abnormality of the integument (HP:0001574) and iv) abnormality of the cardiovascular system. (Fig 2A, Supplementary Document 1). Together, this revision prompted the replacement/restructuring of 67 terms, and the addition of 57 new terms to the HPO tree, among them "recurrent fever", "unusual infections", "IgG levels in blood" (Fig 2B, comprehensive list in Supplementary Documents 1 and 2).

Directed expansion of primary antibody deficiency (PAD) terms

Overall, the PAD working group focused on replacing broad and non-specific terms with terms that describe phenotypes in more detail and accuracy (example: 'partially absent total IgG/IgA/IgM in blood' and '(near) absent total IgG/IgA/IgM in blood' instead of 'hypogammaglobulinemia') Fig 2B. In addition, we proposed that the full detailed spectrum of specific antibody as well as IgG-subclass deficiencies was described by separate HPO terms. For example, we described individual terms related to 'decreased specific antibody response to vaccination in blood' divided according to the response to different types of vaccination (protein, protein-conjugated polysaccharide and unconjugated polysaccharide).

Standardized reannotation of rare, genetically diagnosed IEIs

We started by a systematic review of four disease categories of the IUIS classification of IEIs, as proof of concept: diseases affecting cellular and humoral immunity (IUIS Table 1), diseases of immune dysregulation (IUIS Table 4), autoinflammatory disorders (IUIS Table 7) and genetically undiagnosed predominantly antibody deficiencies (IUIS Table 3), detailed in Table 1 and Supplementary Document 3. As a first step, we assessed the already available HPO annotation for each disease in the v2019-06-03 HPO release (see Supplementary Materials and Methods). We found that 15% of diseases considered (11 of 73 diseases in total) did not have any associated HPO terms (Fig 3A). Overall, we found that on average 13.3 phenotype terms were available per disease (Fig 3B), later referred to as "existing terms".

The text mining and evaluation process was separated into four steps shown in Fig 3C. We have first focused the reannotation of 72 genetically diagnosed IEIs, and genetically

undiagnosed PADs. For genetically diagnosed IEIs, text mining was based on 162 expert-curated articles, on average 2.57 articles per disease (Fig 3D). This resulted in 4,517 extracted phenotype terms, 66.42 terms per disease (Fig 3E). Of these terms, 3,242 - or 71% per disease (47.67 out of 66.42) - were accepted as correctly attributed terms by the expert reviewers (Fig 3F). Expert suggestions added up to 529 additional HPO terms, in addition to the existing and text mined terms.

After reannotation, a mean of 63.1 terms were available for each disease, resulting in a 4.7-fold gain in the number of available annotations (Fig 3G). The mean information content as measured by the overall frequency of terms in each disease's annotations has increased from 6.17 to 8.3 (Fig 3H) after reannotation.

The new annotation of diseases consisted mainly of text mined terms (70.6%) (Fig 3I), followed by already existing terms (9.3%) and additional suggestions by experts (9.3%, adding a further 5.2 additional terms per disease) (Supplementary Document 3).

Standardized reannotation of genetically undiagnosed primary antibody deficiencies (PADs)

PADs form a heterogeneous group, and the majority of PADs do not (as yet) have a genetic diagnosis. We collected articles describing the heterogeneous PADs related to common variable immunodeficiency disorders (CVID), agammaglobulinemia, selective IgM deficiency, selective IgA deficiency, IgG-subclass deficiency, specific antibody deficiency and unclassified antibody deficiency subgroups. In total, 541 terms were text mined from these articles, many of these in more than one PAD subgroup, and 245 of these terms (45.2%) were annotated as correctly associated to the respective PAD subgroup by the expert reviewers (Fig 3J). Of these 245 terms, the experts annotated 16.3% as commonly found in PAD diseases, 48.97% as sometimes associated (albeit less commonly), and 34.7% as rarely associated with PAD (Fig 3K).

Patient-disease matching

We set out to showcase the efficacy of our reannotation effort by highlighting the potential diagnostic impact of optimized disease annotation. To do this, we have selected 30 clinical cases from a large immunology referral center research database (Supplementary Document 4). HPO terms were matched to patient phenotypes by experts from the clinical synopsis and the phenotypic similarity to all HPO-annotated diseases was calculated based on these selected

patient HPO terms (Fig 4A), as illustrated by one concrete clinical example of a patient with Tumor Necrosis Factor Receptor Associated Periodic Syndrome (TRAPS, Fig 4B). Overall, we show a significant 47% improvement in the specificity of patient phenotype matching to correct diagnosis (from 0.49 to 0.72, p value = 1.8e-07, Fig 4C), and a significantly better ranking of the correct clinical diagnosis across all possible diseases after reannotation: in the majority of cases, the correct diagnosis was in the top 10 of matched diseases (Fig 4D) after reannotation, and the rank of the correct diagnosis for individual patients was highly significantly improved, from a mean of 285 to 19 (14.9 fold improvement, p value = 9.1e-07, Fig 4E).

Phenotype-driven disease classification

We tested the efficacy of our approach in selecting biologically and clinically meaningful phenotypes by assessing the HPO-ontology based phenotypic similarity of diseases before and after reannotation. In particular, we assessed whether the similarity was greater within or between IUIS clinically defined groups. We found that the phenotype-driven disease classification after reannotation has resulted in a clustering more in concordance with the IUIS-based clinical classification (Fig 5A-B).

Discussion

Unified data standards, consistent classification and robustly verified clinical data are vital pillars supporting diagnostic pipelines and data-driven research. Although databases and vocabularies that aim to provide accurate phenotypic descriptions exist (*5-9*), there are still major gaps in the depiction of IEIs in these datasets. Here we used a cross-community collaboration to review, expand and improve the depiction of IEIs in HPO, and reannotate IEIs with HPO terms. We reviewed four separate branches of the HPO tree and submitted 57 new and expanded HPO terms, the majority of which are now included in the official HPO dataset. We introduced a semi-automated reannotation pipeline, that combines ontology-guided machine learning and a two-tier expert review to reannotate four main categories of IEIs. The basis of the ontology-guided machine learning was the expert curated list of articles (162 in total), that was submitted to the PanelApp (*21*) to serve as a public resource. The text mined phenotypes were subjected to expert review to confer face validity or refute the putative new HPO terms. IEIs and their current HPO terms covered by the working groups were scrutinized

in-depth, resulting in high-quality annotations. Overall, we have achieved a 4.7-fold gain in number of HPO terms annotating each disease. These annotations included unspecific (frequently annotated) as well as specific (less frequently annotated) HPO terms holding less and more information content respectively. Combined, the mean information content increased from 6.17 to 8.3.

Each reannotated disease showed an increase in information content and a quantitative gain in the number of available HPO terms. Through patient-disease matching and disease-similarity examples we illustrated that these gains and increases translated to significant qualitative improvement in patient-disease matching in an independent cohort of IEI patients (Figure 4), and phenotype-driven classification of IEIs that more closely resembles clinical consensus (Figure 5). Although neither of these measures are systematic assessments of global patient-disease matching and disease similarity comparisons, they highlight that there is considerable benefit by the revision of specific subclasses of diseases. Once a near complete HPO phenotype reannotation of almost all IEIs is available, it will be intriguing to assess how well patients with genetic diagnoses match reannotated OMIM diseases in a clinical setting, how patient matching to genetic diagnosis is transformed, and if these changes ultimately lead to an earlier diagnosis. Finally, once a detailed and accurate phenotypic description is available for all IEIs, identification phenotype-driven patient subgroups will be common practice, and a more objective entirely phenotype-driven classification and ontology of IEIs can become a reality.

Accurate phenotypic description of patients holds promise for diagnostic utility and for the discovery of novel diseases. Phenotype-driven genetic diagnostic tools now exist, but their full clinical potential is hampered by the lack of complete phenotypic descriptions for most types of IEIs. Phenotips (*22*) is a free and open source software for collecting and analyzing phenotypic information of patients with genetic disorders that is widely used in the rare disease community. Tools such as Exomiser use HPO terms to annotate and to prioritize potentially casual variants (*23*). New integrative ‘omics approaches and the analysis of large-scale data with artificial intelligence will allow us to go from a one-size-fits-all to a more personalized medicine, including in IEIs. We see the potential to integrate the richer phenotyping of previously undiagnosed groups of IEI patients with available sequencing data to accelerate disease gene discovery and at the same time increase the diagnostic rate in new patients (*24*).

Novel disease-gene or phenotype associations depends on sufficient numbers of cases as well as a control cohort of comparable quality. Cross-institute and cross-country collaborations for cohorts of undiagnosed, but well-phenotyped patients could shed light on novel disease-

causing genes not only of the immune system. Trusted and accepted data and information sharing platforms are already being developed (*13, 22*) to provide robust and sufficiently granular HPO terms as a standardized way of phenotyping patients. Electronic health records (EHR) (*25*) could facilitate the transfer of HPO terms by integrating with available sharing platforms. Capturing HPO annotations of novel rare diseases or cases is an ongoing challenge for a complete disease representation. Thus it is important that alongside of updating the official IUIS classification, HPO descriptions of disorders are curated once every several years. We suggest a community effort for such regular reviews of HPO regarding IEIs, such as a team of experts, part of big international groups of clinicians such as ESID or ERN RITA, the Clinical Immunology Society (CIS) or other similar organizations. Publication standards that require the submission of HPO annotations up-front would greatly improve this process.

Once phenotyped patients are available, robust and global approaches are accessible (*2*) to find phenotypic similar cases. These comparisons are performed by advanced machine learning algorithms. However, machine learning can also be a very powerful tool to automate the identification of relevant phenotype information in publications or clinical notes. We applied an ontology-guided machine learning tool to support the annotation of diseases and explored the full spectrum of terms – from very relevant to not relevant at all. The same process can be applied to unstructured clinical notes to accelerate in-depth annotation of patients. For patients with EHR (*25*), abnormal clinical values can automatically be translated into HPO codes (*26*) for a more precise diagnostic application and integrated with sharing platforms as mentioned before. The foundation of these comparisons is an ontology with a comprehensive set of term, which is widely used.

As there is currently no gold-standard on how to perform an expert-based review of ontologies. guidance on annotating diseases with HPO phenotypes can vary between diseases, disease classes and centers. IEIs are rare diseases, and often there are only a few patients described (sometimes only one kindred in case of ultra-rare diseases). Therefore, the depth of currently available published phenotypes is at times limited. The low number of patients and insufficient depth of available phenotypes brings up a question as to which diseases to include in phenotyping exercises of this nature. On the one hand, focusing on IEIs that are commonly accepted, with multiple patients diagnosed and well described by multiple researchers can increase the depth of phenotyping. However, this approach excludes at least 10% of IEIs (the ultra-rare diseases). On the other hand, an all-inclusive approach including every disease

systematically means that we rely on sparsely phenotyped patients and perhaps insufficient data for ultra-rare disorders. A warning of accuracy by indicating the frequency of each phenotype for diseases could soon be possible, with the addition of phenotype frequency to the HPO dataset, an expansion that is currently work in progress. This implies the need for a responsive system, capable of assimilating new phenotypic information as the pool of confidently diagnosed patients increases.

Our ongoing approach aims to address these gaps for IEIs and to provide an ontology that is practical, useful and as complete as possible. However, the existence of a well-built ontology and the awareness of clinicians and researchers itself does not guarantee a shift in the community to fully adapt a standardized phenotyping approach. Our approach raised awareness regarding the concept and importance of HPO amongst the IEI community. Moreover, the process made the participating clinicians aware of the available terms and highlighted where these were lacking. Moving forward, it is very important that official entities adopt HPO terms as the unified means of patient phenotyping. We hypothesize that as soon as the widely used registries such as the Undiagnosed Disease Network (*11*) or the IUIS (*27*) use HPO to refer to phenotypic annotation, this will propel the IEI field towards adopting HPO as the main nomenclature for phenotyping IEI patients. One promising move in this direction is the recent expansion of the ESID registry working definitions for the clinical diagnosis of IEIs (*28*), which derives HPO terms from OrphaNet using the ORDO Ontological Module (HOOM) platform (*29*), prompted by our HPO initiative.

In summary, our work reviewed and expanded the phenotypic depiction of multiple subclasses of IEIs, and to our knowledge, this initiative is the first endeavor of its kind with the aim of standardizing IEI phenotypes. Our semi-automated annotation-based approach is scalable to include all IEIs as illustrated herein. We propose our reannotation approach as a blueprint for systematic HPO (re)annotation for additional immunological and non-immunological diseases.

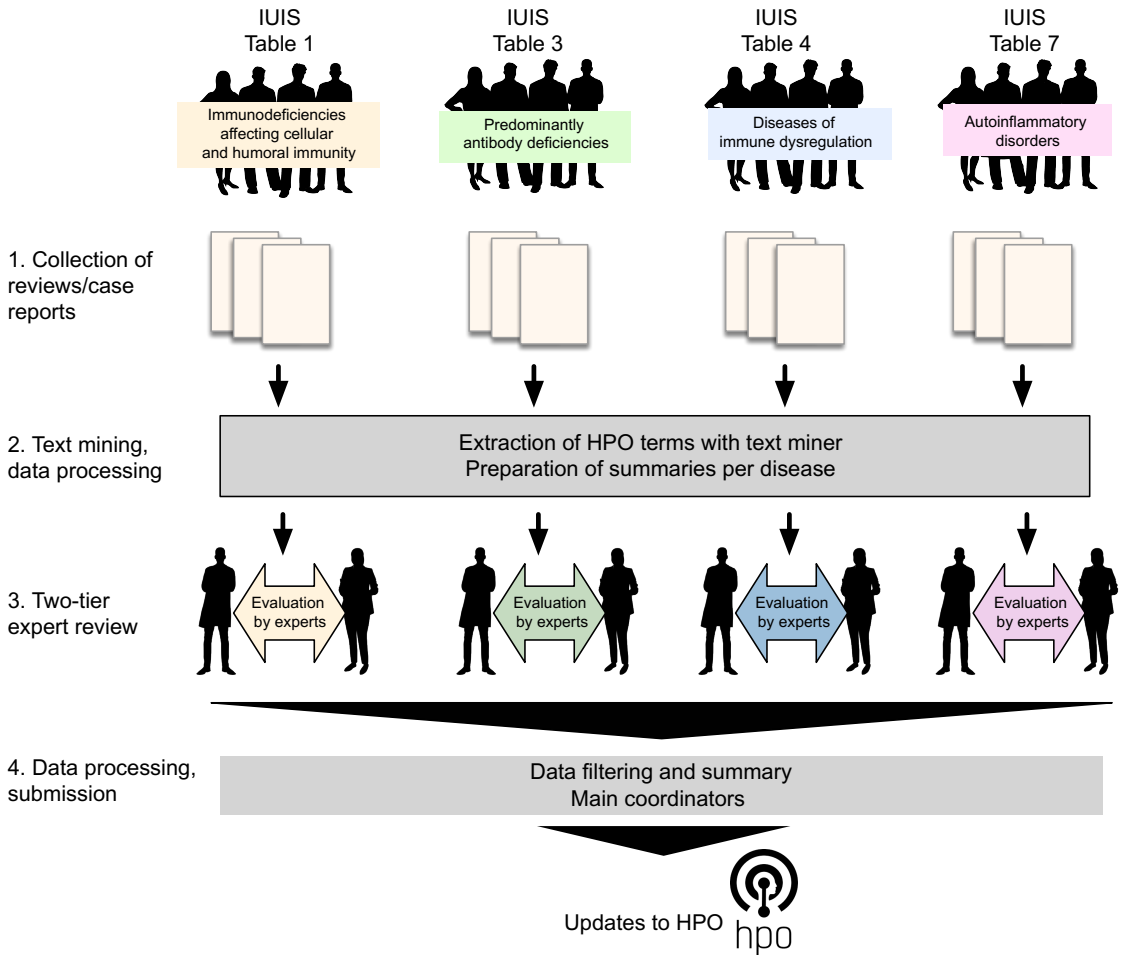


Fig 1: Pipeline for of standardized reannotation of IEI diseases. First, scientific publications were collected by experts for each disease within the subgroups. Second, HPO terms were extracted from the provided publications for each disease using machine learning and summarized into Excel documents. Third, a two-tier expert review evaluated the text mined terms, suggested additional terms if required and the responsible working group agreed on the final HPO annotations for each disease. Fourth, data were collated, and the agreed terms were submitted to HPO.

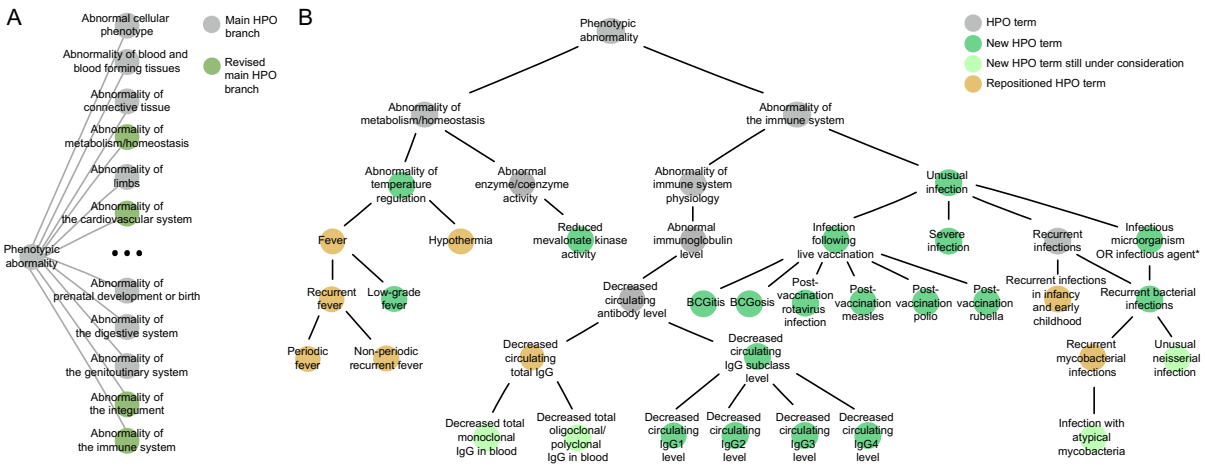
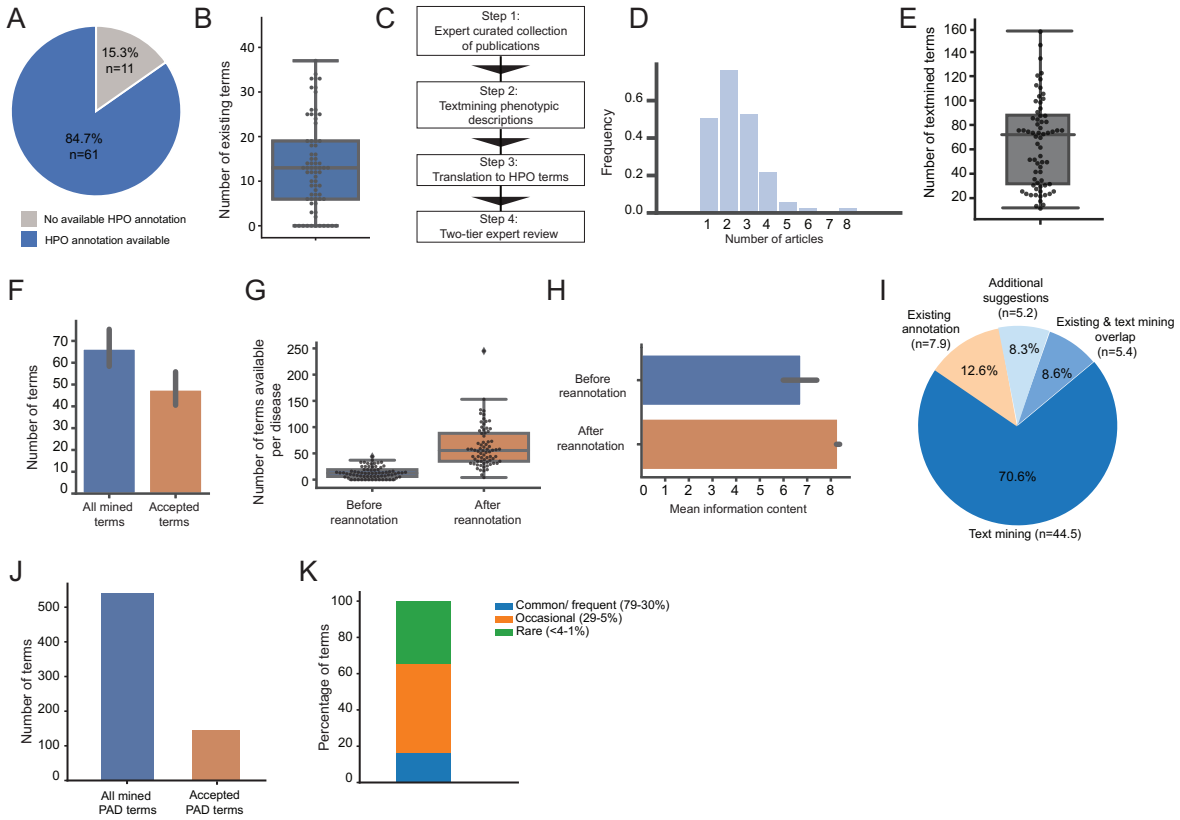


Fig 2: Revision and expansion of the HPO tree. A) Schematic representation of the restructuring of the HPO tree. Main branches of the HPO tree where restructuring was performed are marked with light green. B) “Abnormality of temperature”, “Abnormality of immunoglobulin level“ and “Unusual infections“ as examples of revised branches of the HPO tree. New additions and suggestions are marked with green, repositioned terms are marked with yellow.



513

514 **Fig 3: Result of disease reannotation.** A) HPO annotation availability in the subset of 72
515 diseases. B) Distribution of number of available HPO terms per disease. C) Distribution of the
516 number of articles used per disease for the reannotation pipeline. D) Number of mined terms
517 per disease. Each dot represents a disease. E) All mined vs all accepted terms. F) Number of
518 available terms per disease before and after reannotation. Each dot represents a disease. G)
519 Mean information content available per disease before and after reannotation. H) The aggregate
520 mean annotation per disease after reannotation. I) All text mined terms from PAD publications
521 J) Frequency distribution of different PAD terms according to the experts. HPO: Human
522 Phenotype Ontology; PAD: Primary Antibody Deficiencies.

523

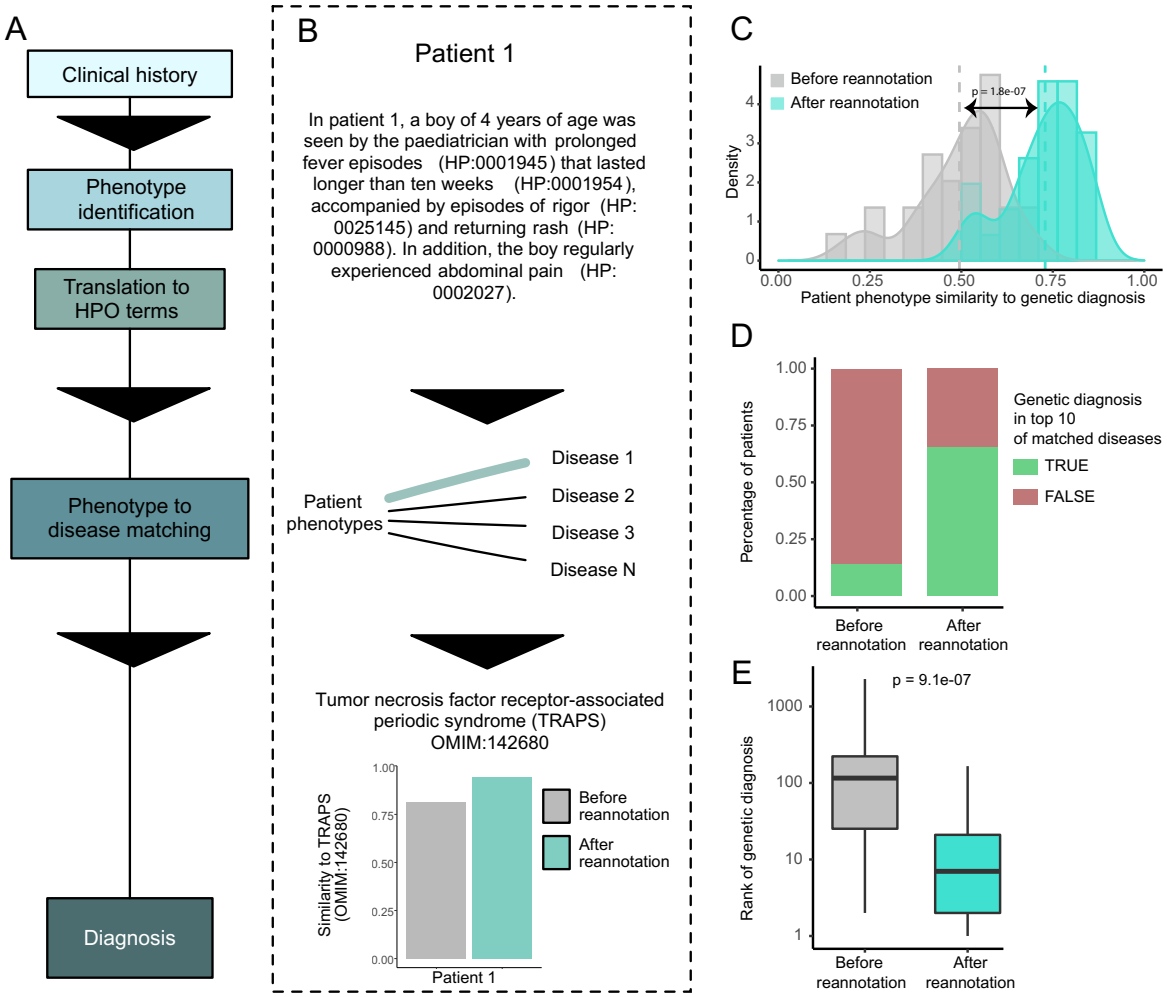


Fig 4: Patient-disease matching. A) Schematic overview of the different steps of patient-to-disease matching. First, the phenotypes were identified in a patient's clinical history. Second, these phenotypes were translated to HPO terms. Finally, patient phenotype to disease matching was measured by Lin similarity. B) Matching patient 1 to a diagnosis. C) Similarity of patients in patient cohort to genetic diagnosis before and after reannotation. D) The rank of correct clinical diagnosis more often is in the top 10 of matched diseases after reannotation. E) Improvement of ranks of clinical diagnosis before and after reannotation. Significance was assessed by Student t-test.

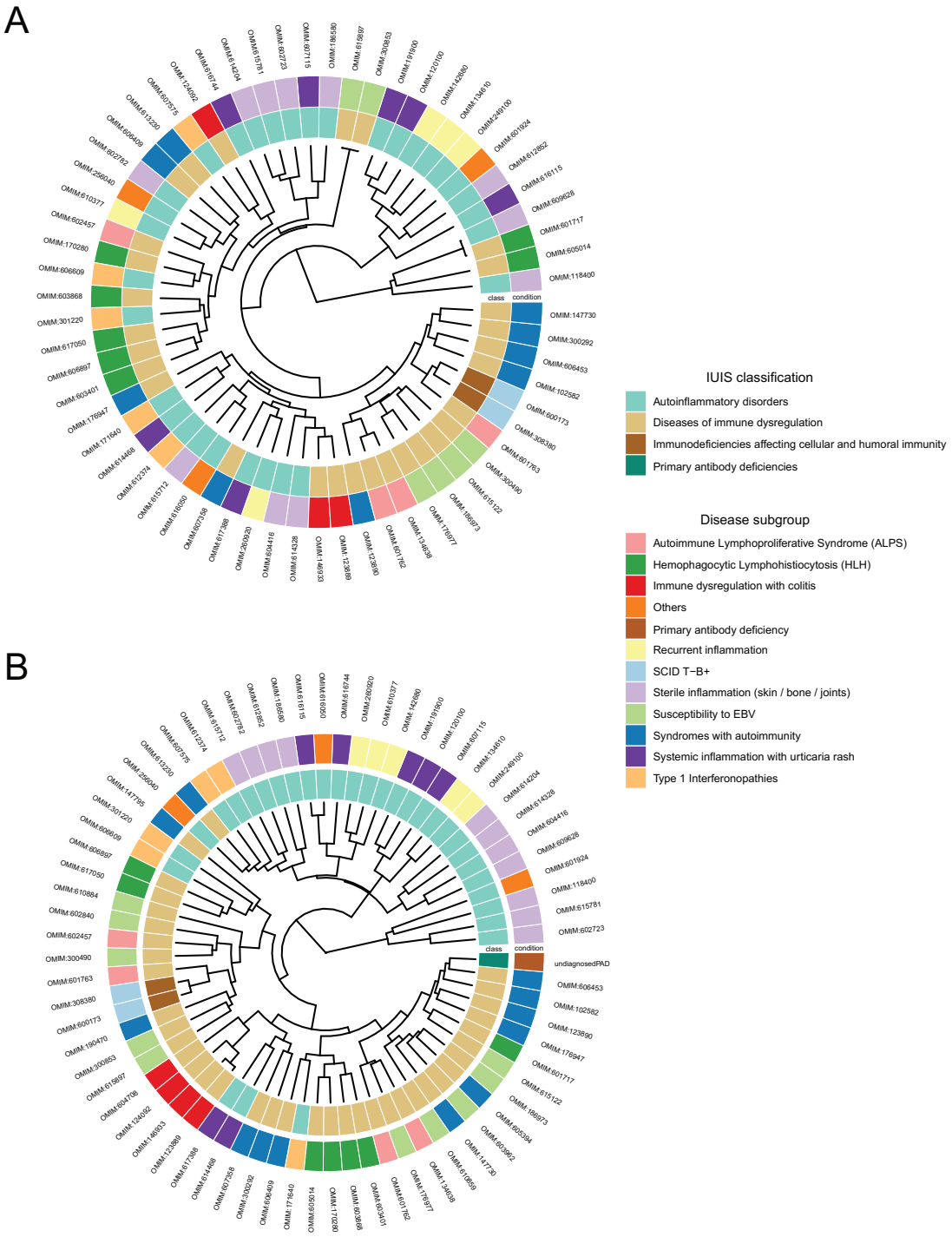


Fig 5: Phenotypic similarity of diseases before and after reannotation. Diseases are annotated with the IUIS disease group (inner circle), sub-group (outer circle) and OMIM identifier. A) Clustering of diseases based on phenotypic similarity before reannotation. B) Clustering of diseases based on phenotypic similarity after reannotation. HPO: Human Phenotype Ontology; IUIS: International Union of Immunological Societies, OMIM: Online Mendelian Inheritance in Men; IEI: Inborn Errors of Immunity; EBV: Epstein-Barr Virus

References

1. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, et al. The National Institutes of Health Undiagnosed Diseases Program: insights into rare diseases. *Genet Med*. 2012 Jan;14(1):51–9.
2. 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;36(10):915–21.
3. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat*. 2015 Oct;36(10):928–30.
4. Hernandez-Ibarburu G, Perez-Rey D, Alonso-Oset E, Alonso-Calvo R, de Schepper K, Meloni L, et al. ICD-10-CM extension with ICD-9 diagnosis codes to support integrated access to clinical legacy data. *Int J Med Inform*. 2019;129:189–97.
5. Amberger J, Bocchini C, Hamosh A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). *Hum Mutat*. 2011 May;32(5):564–7.
6. Pavan S, Rommel K, Mateo Marquina ME, Höhn S, Lanneau V, Rath A. Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. *PLoS ONE*. 2017;12(1):e0170365.
7. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res*. 2019 08;47(D1):D955–62.
8. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008 Nov;83(5):610–5.
9. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res*. 2017 04;45(D1):D865–76.
10. Ramoni RB, Mulvihill JJ, Adams DR, Allard P, Ashley EA, Bernstein JA, et al. The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease. *Am J Hum Genet*. 2017 02;100(2):185–92.
11. Taruscio D, Groft SC, Cederroth H, Melegh B, Lasko P, Kosaki K, et al. Undiagnosed Diseases Network International (UDNI): White paper for global actions to meet patient needs. *Mol Genet Metab*. 2015 Dec;116(4):223–5.
12. Gall T, Valkanas E, Bello C, Markello T, Adams C, Bone WP, et al. Defining Disease, Diagnosis, and Translational Medicine within a Homeostatic Perturbation Paradigm: The National Institutes of Health Undiagnosed Diseases Program Experience. *Front Med (Lausanne)*. 2017;4:62.
13. Thompson R, Johnston L, Taruscio D, Monaco L, Béroud C, Gut IG, et al. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. *J Gen Intern Med*. 2014 Aug;29 Suppl 3:S780-787.
14. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research*. 2019 Jan 8;47(D1):D1018–27.
15. Chinn IK, Chan AY, Chen K, Chou J, Dorsey MJ, Hajjar J, et al. Diagnostic interpretation of genetic studies in patients with primary immunodeficiency diseases: A working group report of the Primary Immunodeficiency Diseases Committee of the American Academy of Allergy, Asthma & Immunology. *Journal of Allergy and Clinical Immunology*. 2020 Jan;145(1):46–69.

16. Rae W, Ward D, Mattocks C, Pengelly RJ, Eren E, Patel SV, et al. Clinical efficacy of a next-generation sequencing gene panel for primary immunodeficiency diagnostics. *Clin Genet*. 2018 Mar;93(3):647–55.
17. Arbabi A, Adams DR, Fidler S, Brudno M. Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning. *JMIR Med Inform*. 2019 May 10;7(2):e12596.
18. Greene D, Richardson S, Turro E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics*. 2017 01;33(7):1104–6.
19. Yu G, Lam TT-Y, Zhu H, Guan Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol*. 2018 01;35(12):3041–3.
20. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 2019 Feb 1;35(3):526–8.
21. Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat Genet*. 2019;51(11):1560–5.
22. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat*. 2013 Aug;34(8):1057–65.
23. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015 Dec;10(12):2004–15.
24. Westbury SK, Turro E, Greene D, Lentaigne C, Kelly AM, Bariana TK, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Medicine*. 2015 Apr 9;7(1):36.
25. Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The Use of FHIR in Digital Health - A Review of the Scientific Literature. *Stud Health Technol Inform*. 2019 Sep 3;267:52–8.
26. Shefchek KA, Harris NL, Gargano M, Matentzoglu N, Unni D, Brush M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res*. 2020 08;48(D1):D704–15.
27. Bousfiha A, Jeddane L, Picard C, Ailal F, Bobby Gaspar H, Al-Herz W, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J Clin Immunol*. 2018;38(1):129–43.
28. Seidel MG, Kindle G, Gathmann B, Quinti I, Buckland M, van Montfrans J, et al. The European Society for Immunodeficiencies (ESID) Registry Working Definitions for the Clinical Diagnosis of Inborn Errors of Immunity. *J Allergy Clin Immunol Pract*. 2019 Aug;7(6):1763–70.
29. Gasteiger LM, Robinson PN, Pazmandi J, Boztug K, Seppänen MRJ, Seidel MG. Supplementation of the ESID registry working definitions for the clinical diagnosis of inborn errors of immunity with encoded human phenotype ontology (HPO) terms. *The* Journal of Allergy and Clinical Immunology: In Practice. 2020 May 1;8(5):1778.
30. Groza T, Köhler S, Doelken S, Collier N, Oellrich A, Smedley D, et al. Automatic concept recognition using the human phenotype ontology reference and test suite corpora. *Database (Oxford)*. 2015;2015.

Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity

Materials and Methods

Organization of working groups, working group participants

Working groups were established following the 2017 IUIS classification categories (*27*). Based on the participants expertise, the working groups addressed the following IUIS classification categories: diseases affecting cellular and humoral immunity (IUIS Table 1), predominantly antibody deficiencies (IUIS Table 3), diseases of immune dysregulation (IUIS Table 4), and autoinflammatory disorders (IUIS Table 7). Headed by a group lead, themselves an expert in the field of the specific diseases, each working group was accompanied by an additional member (coordinator) from the Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD) to facilitate coordinated communication and organization. Working groups had between 5 and 9 members with disease-group-specific standard operational procedures in place to facilitate the workflow. Coordinators initiated and communicated general agendas within each working group, implemented disease-group-specific tasks, hosted remote meetings and communicated results and subsequent actions via email.

Table 1: Organization and working group participants

Organization & coordination			Specific working groups			
Initiative leaders	**Main coordinators**	**Group coordinators**	**Diseases affecting innate and adaptive immunity (IUIS Table I)**	**Antibody deficiencies (IUIS Table III)**	**Diseases of immune dysregulation (IUIS Table IV)**	**Autoinflammatory diseases (IUIS Table VII)**
Kaan Boztug[1-3,37]	Matthias Haimel[1-3]	Raúl Jiménez Heredia[1-3]	Sophie Hambleton[29]	Esther de Vries[33,34]	Catharina Schuetz[30]	Anna Simon[31]
Marielle van Gijn[36]	Julia Pazmandi[1-3]	Jasmin Dmytrus[1-3]	Kimberly C. Gilmour[13]	Siobhan Burns[24]	Shahrzad Bakhtiar[15]	Paul Brogan[12,13]
Peter Robinson[35]		Sevgi Köstel Bal[1-3]	Fabian Hauck[27,28]	Johannes Trück[18]	Beata Derfalvi[16]	Marco Gattorno[11]
Christoph Bock[1,3,32]		Samaneh Zoghi[1-3]	Francesco Saettini[17]	Jacques Rivière[19,20]	Markus G. Seidel[23]	Paul van Daele[4]
		Julia Pazmandi[1-3]	Mirjam van der Burg[21,22]			Tracy A Briggs[5,6]
			Maaike A. A. Kusters[12,13]			Carine Wouters[7,8]
			Reem Elfeky[12,13]			Brigitte Bader-Meunier[9,10]
			Elisabeth Salzer[1-3]			Florence Aeschliman[9,10]
						Roberta Caorsi[11]
						Despina Eleftheriou[12,13]
						Esther Hoppenreijs[14]

Expansion and restructuring of disease related branches of the HPO tree

OMIM (*5*) (if unavailable, OrphaNet (*6*)) identifiers were used as starting points to refer to Inborn Errors of Immunity (IEI)s. For each disease, available Human Phenotype Ontology (HPO) terms and the ontology structures were extracted from the v2018-06-13 HPO disease annotation and ontology release (https://hpo.jax.org), and Excel documents were prepared summarizing the annotations per disease. Each document contained the currently available HPO terms associated with one disease, upstream terms of these current HPO terms organized in a tree structure. Both the correctness of available terms and the ontology structure associated with the terms was assessed. Disease-specific HPO restructuring was discussed within the working groups and the results debated among all participants. The suggested changes were summarized electronically in Excel documents or pictures of flipchart drawings by the main coordinators before being submitted to HPO. Additionally, missing terms describing pulmonary and gastro-intestinal complications of Primary Antibody Deficiency (PAD) were discussed during teleconferences and thereafter submitted to update the HPO ontology. All results of the restructuring are detailed in Supplementary Document 1&2.

Standardized re-annotation of rare, genetically diagnosed diseases

A standardized, semi-automated reannotation process was developed in order to consistency annotate all IEIs (over 300 different diseases in OMIM) with HPO terms across working groups.

In the first step, working groups collected a minimum of two articles in portable document format (PDF) that adequately illustrated the phenotypic spectrum of each disease. In the second step, the text was extracted from the PDF files using the content analysis tool Apache Tika (https://tika.apache.org/) from the python package tika (version 1.19). Text sections associated with HPO terms were identified by applying an ontology-guided machine learning tool, the Neural Concept Recognizer (NCR), with default settings as previously published (*17*), trained on the v2019-06-03 HPO ontology release. The NCR was selected due to the ability to work with HPO, utilize the hierarchy information and semantic similarity for improved identification of HPO terms, and the robust performance on a published PubMed article abstracts dataset with manual HPO annotations (*17,30*). The identified HPO terms and term frequencies were collected for each article and further summarized per disease. These disease specific summaries were prepared as Excel documents, where HPO terms were ranked by frequency across articles.

Highlighted HPO terms indicated already available annotations in the v2019-06-03 HPO disease annotation release on the same HPO branch (of this specific or more/less specific HPO term). These Excel documents were distributed to the working groups and evaluated with two-tier expert review. For the two-tier expert review, each disease was reviewed by at least two independent experts. The experts were asked to indicate if an HPO term was either a true phenotype present in the disease, or a false positive association. In addition to evaluating the HPO terms identified in the articles, the experts evaluated existing HPO terms from the HPO annotation release as well. Further phenotypes not identified by the previous two steps (identified in articles or available in the HPO annotation release) were suggested by experts as additional terms (HPO terms if available or free text) to cover the full phenotypic spectrum of the diseases.

In case of any disagreement in the evaluation, a consensus discussion between the two experts for that particular disease was scheduled. If after the second-tier overview between the two experts there was still no agreement, these were discussed by the whole group for overall consensus, defined as at least 80% agreement amongst the expert group. The consensus of the expert evaluations was collected in standardized Excel documents. These consensus Excel documents per disease were integrated by the main coordinators at LBI-RUD with all diseases across working groups. The full list of reannotated diseases available in Supplementary Document 3. The list of reannotated terms for each disease is available in Supplementary Document 4.

Standardized re-annotation of genetically undiagnosed diseases

Literature describing the phenotypic characteristics of the various PAD subtypes without a known monogenetic defect were collected and reviewed by the PAD-subgroup members. The ontology-guided machine learning tool was run as described above. Each HPO identifier (either identified in an article or available in the HPO annotation release) was annotated with: true/false/exclusion criteria. In case of a true phenotype, the observed frequency within patients was assessed and noted down as well. The frequencies correspond to the following representation in patients: Common = Frequent (79-30%); sometimes = Occasional (29-5%); rare = Very rare (<4-1%).

Preparation of disease annotation data for similarity measures

The list of existing annotations per disease (later referred as "HPO-disease-annotations") was obtained by extracting the HPO terms available per disease from the v2020-03-27 HPO release. Redundant terms (as defined as terms, where more specific terms are already linked and available to the disease) were removed from each disease annotation by applying the *minimal_set* function, based on the v2020-03-27 HPO release ontology structure. To obtain the reannotated set of annotations (detailed above, later referred to as "reannotated-disease-annotations"), the list of reviewed and evaluated disease annotations was extracted from the working groups. Redundant terms were removed with by applying the *minimal_set* function.

Disease-disease similarity measures

Similarity measures of diseases based on both disease annotations sets - "HPO-disease-annotations" and "reannotated-disease-annotations" - was carried out by the R package ontologyX (*18*), with default settings applying the Lin similarity measure. A phenotypic similarity matrix of disease similarity data was calculated for both sets of annotations using the *get_sim_grid* method with default parameters. Diseases were clustered based on this similarity matrix using euclidean distances. Hierarchical clustering was performed and visualized with *ggtree* using the *R* packages ggtree v1.14.6 (*19*) and ape v5.2 (*20*).

Patient-disease similarity measures

For each patient, HPO terms from clinical synopses were extracted with the Neural Concept Recognizer (*17*), then reviewed and expanded by an expert clinician. The semantic similarity of the extracted HPO terms per patient was compared to all diseases in both disease annotation sets, "HPO-disease-annotations" and "reannotated-disease-annotations" (see above), using the R package ontologyX (*18*), with default settings using the Lin similarity measure, by applying *get_profile_sim*. The statistical significance of the difference between cohort-wide similarity scores to genetic diagnosis using "HPO-disease-annotations" (therefore before reannotation) and "reannotated-disease-annotations" (therefore after reannotation) was assessed by a Student T-test using the R package ggpubr v0.2.999. Next, per patient, all diseases were ranked according to their similarity scores to the list of patient HPO terms. Diseases with the highest similarity received the lowest rank. The statistical significance of the difference of rank comparing the two annotations - "HPO-disease-annotations" (therefore before reannotation) and "reannotated-disease-annotations" (therefore after reannotation) - was assessed by Student

T-test using the R package ggpubr v 0.2.999. A detailed list of patients and their similarity to their genetic diagnosis can be found in Supplementary Table 4.

List of Human Phenotype Ontology Resources:

Data	**Accessible from**
Current version of the HPO ontology	https://hpo.jax.org/app/download/ontology
Current HPO annotation of diseases	https://hpo.jax.org/app/download/annotation
List of HPO ontology versions	https://github.com/obophenotype/human-phenotype-ontology/releases https://bioportal.bioontology.org/ontologies/HP
v2018-06-13 HPO ontology release	http://purl.obolibrary.org/obo/hp/releases/2018-06-13/hp.owl
v2019-06-03 HPO ontology release	http://purl.obolibrary.org/obo/hp/releases/2019-06-03/hp.owl
v2020-03-27 HPO ontology release	http://purl.obolibrary.org/obo/hp/releases/2020-03-27/hp.owl
HPO disease annotation archive file	https://archive.monarchinitiative.org/hpo-archive/20210126_jenkins_jobs.tar.gz
v2018-06-13 HPO annotation release	jobs/hpo.annotations/builds/1254/archive/misc (in HPO annotation archive file)
v2019-06-03 HPO annotation release	jobs/hpo.annotations/builds/1266/archive/misc (in HPO annotation archive file)
v2020-03-27 HPO annotation release	jobs/hpo.annotations/builds/1271/archive/misc (in HPO annotation archive file)
Current HPO disease annotations	https://ci.monarchinitiative.org/view/hpo/job/hpo.annotations/

List of Supplementary Documents:

Supplementary Materials and Methods

Supplementary Document 1: HPO tree restructuring and the list of new terms

Supplementary Document 2: Summary of diseases reannotated

Supplementary Document 3: List of all terms per disease after reannotation

Supplementary Document 4: List of cases used for phenotype to diagnosis matching

3.2. AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation.

Julia Pazmandi, Sevgi Köstel Bal, Felix Müller, Celine Sin, Christiane V. R. Hütter, Jörg Menche*#, Kaan Boztug*#. * these authors contributed equally, # to whom correspondence should be addressed.

The overarching ambition of this project was to develop a systems-level view of rare autoimmune and autoinflammatory diseases and showcase its utility for addressing a wide range of important biomedical questions. To achieve this, we developed a network-based framework for integrating all currently known monogenic immune defects underlying autoimmunity and autoinflammation and their molecular interactions. We show that autoimmune and autoinflammatory phenotypes do not separate on the interactome, and that gene defects that present with both phenotypes have a tendency to have a more diverse molecular outreach in their network neighborhoods. With this network-based framework, we identify the AutoCore, a markedly connected subnetwork on the interactome that is the functional core of rare monogenic, but also of complex polygenic autoimmune/autoinflammatory diseases. We show how the AutoCore connects monogenic and polygenic diseases with disease examples of SLE and IBD. We further use the AutoCore to define 19 phenotypically and therapeutically cohesive molecular disease-subclusters. Finally, we show that the disease clusters can relate to known targeted therapies, suggesting that network-based measures may, in the future, provide an additional layer of information for physicians when choosing novel therapeutic avenues or repurposing strategies.

This manuscript contains a general overview of the genetics of rare autoimmune and autoinflammatory diseases, as well as the introduction of our network-based methodology and the construction of the AutoCore. Subsequent analysis underpin the functional relevance of the AutoCore and interactome-based methods for rare diseases in general.

AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation

Julia Pazmandi[1,2,3], Sevgi Köstel Bal[1,2,4], Felix Müller[2,3], Celine Sin[2,3], Christiane V. R. Hütter[2,3], Jörg Menche[2,3,5,*], Kaan Boztug[1,2,4,6,7,*]

[1]Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases, Lazarettgasse 14, AKH BT 25.3, A-1090 Vienna, Austria

[2]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, AKH BT 25.3, A-1090 Vienna, Austria

[3]Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Campus Vienna BioCenter 5, A-1030 Vienna, Austria

[4]St. Anna Children's Cancer Research Institute (CCRI), Zimmermannplatz 10, A-1090 Vienna, Austria.

[5]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria

[6]St. Anna Children's Hospital, Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria

[7]Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria.

*to whom correspondence should be addressed:

Jörg Menche, Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Campus Vienna BioCenter 5, A-1030 Vienna, Austria joerg.menche@univie.ac.at, Phone: +43 1-4277-74370, Fax: +43-1-4277-874370

Kaan Boztug, Ludwig Boltzmann Institute for Rare and Undiagnosed Diseases (LBI-RUD) and St. Anna Children's Cancer Research Institute (CCRI), Zimmermannplatz 10, A-1090 Vienna, kaan.boztug@rud.lbg.ac.at, Phone: +43 1-40470-4080, Fax: +43-1-40170-7280

Keywords: network medicine, autoimmunity, autoinflammation, rare diseases

Abstract

Monogenic autoimmune/autoinflammatory diseases have advanced our understanding of core genes and pathways of the immune system. As they are studied largely in isolation, a unifying view of this important class of diseases is still lacking. Here, we present a network-based approach for integrating all currently known monogenic autoimmune/autoinflammatory diseases into a global network map of human immune dysregulation. We identify the AutoCore, a markedly connected interactome subnetwork as the set of genes and their molecular interactions essential for immune homeostasis. We show that autoimmunity and autoinflammation are intimately linked on a molecular level, that the monogenic AutoCore is the topological core of polygenic autoimmune/autoinflammatory diseases, and define 19 molecularly and phenotypically cohesive disease subclusters of monogenic autoimmunity/autoinflammation. We use the AutoCore to pinpoint novel therapeutically targetable pathways, illustrating the relevance of the AutoCore as a resource to distill the molecular landscape of autoimmunity/autoinflammation and quantify previously only descriptive clinical observations.

Introduction

Humans are constantly challenged by exposure to different pathogens. Evolution has installed sophisticated defense mechanisms to these threats leading to the formation of the human immune system as we know it today. The main challenge for the immune system therefore is to recognize self and distinguish it from foreign (in the form of pathogens or other antigens), to neutralize the latter. A considerable number of human diseases including rheumatoid arthritis[1,2], type 1 diabetes[3,4] or inflammatory bowel disease[5,6], are the result of misguided immune reactions leading to autoimmunity and/or autoinflammation[7]. These conditions are typically thought to be of multi-factorial origin, involving a complex interplay of genetic and environmental factors that remains poorly understood. Recently, also a number of rare, monogenic causes of autoimmunity and autoinflammation have been identified[8], enabled by deep sequencing and global efforts of phenotypic and genetic data sharing[9,10]. In contrast to polygenic autoimmune/autoinflammatory diseases, monogenic diseases offer much clearer genotype-phenotype relationships and thus enable mechanistic dissection of key pathways that are required for immune homeostasis. Their study has markedly enhanced our understanding of self-tolerance and immune pathways, provided targets for in-depth mechanistic studies and potential therapeutic intervention[11,12]. Notably, although individual components of these diseases have been investigated in detail, because of their rarity, heterogeneity, and the difficulty to efficiently model a considerable fraction of the diseases in animals[13,14], systematic studies that compare and integrate the key players of these phenotypes are still lacking. A systems-based disease paradigm[15], as recently proposed for rare diseases of innate immunity and some complex autoimmune diseases[16,17], could aid in the identification of disease subgroups, implicated molecular mechanisms, and elucidation of the relationship between monogenic and polygenic autoimmunity/autoinflammation. Network medicine provides such a paradigm, by integrating all disease-associated proteins and their molecular interactions into a disease module[18,19,20]. So far, network medicine methods have predominantly been applied to common, polygenic diseases[21–24], with only few applications for rare monogenic diseases[25,26]. Here, we use network approaches to identify the AutoCore, a connected, functionally relevant core of autoimmunity/autoinflammation on the interactome network of molecular interactions. We navigate the AutoCore to systematically compare polygenic and monogenic autoimmune/autoinflammatory diseases, to identify molecularly-, phenotypically- and clinically-informative disease subclusters, and to pinpoint novel targetable pathway candidates for therapeutic manipulation.

Results

Charting the genetic landscape of inborn errors of immunity with autoimmunity and/or autoinflammation

To leverage the collective knowledge on single-gene perturbations of autoimmunity/autoinflammation, we extracted known monogenic defects from the classification of inborn errors of immunity (IEIs)[27] as provided by the International Union of Immunological Societies (IUIS), the Online Mendelian Inheritance in Man (OMIM)[28], and OrphaNet[29] databases (Figure 1A). In total, we identified 186 gene defects underlying monogenic autoimmune/autoinflammatory conditions published over the last 50 years, with a marked acceleration of discovery facilitated by next-generation sequencing-based approaches in the last decade (Figure 1B, Supplementary Table 1). We found that 43% of IEIs present with autoimmunity, autoinflammation, or a combination thereof (Figure 1C). The Genetic and Rare Disease Information Center (GARD)[30] contains treatment information for only 38.8% of these diseases (Figure 1D), highlighting the lack of standardized, publicly accessible data on disease phenotypes, prevalence and treatment options. The mainstay of treatment for immune dysregulation in the context of IEIs is still unspecific, e.g., using general immunosuppressants such as corticosteroids, although more recently, molecular pathway-based treatment strategies have emerged, including monoclonal antibodies targeting CD20, IL1, TNF, IL6, or CD80/CD86 (Figure 1E, Supplementary Table 1).[31] Along with IEIs, genes underlying autoimmune/autoinflammatory conditions are less tolerant to loss of function (LOF) variation than other genes (Figure 1F), pointing to their central importance in homeostasis. Accordingly, we find that most of the gene defects underlying autoimmune and autoinflammatory phenotypes (87.6%) are LOF, while 8.1% can present as gain of function (GOF) (Figure 1G). Compared to IEIs, this constitutes an enrichment of GOF defects in autoimmune/autoinflammatory diseases[32], illustrating that autoimmune/autoinflammatory diseases often represent over-activation of the immune system.

Causative genetic defects for monogenic autoimmune and autoinflammatory diseases constitute a seed network

Interactomes represent all molecular interactions within the cell and can therefore be regarded as maps of cellular organization and function[33]. To chart the map of human immune dysregulation and to explore the molecular links between different monogenic autoimmune/autoinflammatory diseases, we started by compiling a comprehensive interactome from several large-scale protein-protein interaction datasets[22,34,35], resulting in a network consisting of 18,853 nodes representing proteins and 483,037 links representing physical interactions (Figure 2A and Methods). The structural and functional characteristics of our interactome are consistent with previously published ones in terms of overall distribution of interaction partners (Figure 2B) and significant connectivity of known biological pathways (Figure 2C). We first investigated the network properties of monogenic autoimmune/autoinflammatory genes. It has been shown that highly central nodes generally correspond to genes with important roles in healthy and disease

associated processes, such as essential genes or cancer driver genes[36–38]. We thus considered the centrality of autoimmune/autoinflammatory genes within the interactome as quantified by their number of interaction partners (degree), and their betweenness, closeness and eigenvector centralities. We found that compared to all nodes in the network, monogenic autoimmune/autoinflammatory genes have a 4-fold higher degree, 1.7-fold higher betweenness centrality, and 2-fold higher eigenvector centrality (Figure 2D-F, Figure S1A). We next compared these characteristics between autoimmune/autoinflammatory genes and found that gene defects that can present with both phenotypes are characterized by lower degrees compared to gene defects that present only with either autoimmunity or autoinflammation, but tend to be more central (Figure S1B-D). Genes associated with the same complex disease have been shown to form connected clusters on molecular networks[22], so-called 'disease modules'. We hypothesized that, collectively, also monogenic autoimmune/autoinflammatory diseases could behave as a joint disease module. To assess the connectivity among monogenic autoimmune/autoinflammatory genes, we measured the largest connected component (lcc) of directly connected autoimmune/autoinflammatory genes. The lcc contained 133 of 186 genes, representing a highly significant cluster compared to random expectation (Figure 2G-H, z-score=7.6, empirical p-value<0.0001, as obtained from 1,000 simulations with randomly selected genes). Within the connected cluster of autoimmune/autoinflammatory genes, we found that IUIS clinical subgroups[27] related to specific cellular mechanisms form more pronounced subclusters compared to more generic clinical phenotypes, which are more widely spread on the interactome (Figure 2I-K). Furthermore, the genes underlying autoimmune/autoinflammatory phenotypes are significantly agglomerated within the same interactome neighborhood (z-score=-27.56, p-value<0.00001, Figure S1E). A similar pattern of significant connectivity among monogenic autoimmune/autoinflammatory genes was also observed on tissue-specific co-expression networks built from large-scale transcriptomic data[39] (median z-score=5.08, p-value<0.0001; Figure S1F), suggesting transcriptional regulation among these genes that is shared across various tissues. Overall, these results imply that although autoimmune/autoinflammatory gene defects underpin a heterogeneous group of diseases, they behave like a single disease module from a systems perspective.

Identification of the AutoCore, a significantly dense subnetwork of monogenic autoimmunity and autoinflammation

We used the significantly connected lcc amongst known monogenic autoimmune/autoinflammatory genes as a starting point to identify the AutoCore, a cohesive and connected interactome submodule that represents the mechanistic basis of diseases associated with autoimmunity/autoinflammation. Our analysis above showed that a considerable proportion (28.5%) of the autoimmune/autoinflammatory gene defects are only indirectly connected to the lcc. We hypothesized that linker genes that connect the fragmented autoimmune/autoinflammatory genes are related to the cellular machinery whose perturbation underlies the different diseases. To identify these linker genes, we implemented a random walk-based propagation method widely used to uncover vital links between nodes on biological networks[40]. The method expands

the subnetwork around the monogenic autoimmune/autoinflammatory genes until all monogenic disease genes have been incorporated (Figure 3A-C). The resulting AutoCore contains 399 nodes, among them 186 autoimmune/autoinflammatory disease genes and 213 functional linker genes, connected by a total of 5,244 links (Figure 3C-D). Notably, the number of linker genes that are required to connect all fragments into a single cluster is much smaller compared to previously studied complex diseases[41]. This is likely a consequence of the higher centrality of the monogenic autoimmune/autoinflammatory genes and highlights the cohesiveness of the AutoCore. We hypothesized that the AutoCore not only connects monogenic autoimmune/autoinflammatory diseases, but represents a functionally relevant interactome submodule pinning down the core of processes leading to and involved in autoimmunity/autoinflammation. To explore this hypothesis, we first focused on the linker genes within the AutoCore. We found that linker genes are enriched in pathways involved in viral infection induced diseases such as Hepatitis B, EBV infection, Influenza A, CMV infection, NFkB signaling and cytokine signaling, as well as pathways in cancer. We found further enrichment of linker genes among frequently mutated genes in chronic lymphocytic leukemia (CLL)[42] and in highly expressed genes in lymphomas as per the Human Protein Atlas[43,44] (Figure S2A-D), illustrating their vital roles in particular for hematological malignancies. Linkers therefore represent hallmarks of autoimmune/autoinflammatory diseases, such as abrogated T-cell immunity and viral infections[45,46], and genes prevalent in hematological malignancies which occur in ~5% of IEIs[47,48]. Furthermore, we found that the linker genes are enriched in targets of drugs that are used to treat autoimmune/autoinflammatory diseases (Figure 3E), and in genes associated with polygenic, common autoimmune/autoinflammatory diseases (Figure 3F), demonstrating that the linker genes represent vital building blocks of processes involved in autoimmunity/autoinflammation. To assess the extent to which the AutoCore is specific to certain tissue or cell types, we considered the lcc of the AutoCore on tissue specific coexpression networks[39]. We found that the AutoCore is significantly enriched and connected on various tissue specific networks, with the whole blood network showing most significant enrichment (odds ratio OR = 2.33, p-value = 3.3e-16, Figure S2E), and with the connectivity most prominent on the thyroid coexpression network (z-score 7.8, p-value = 3.1e-13, Figure S2F). This illustrates that the AutoCore is linked through transcription machinery across various tissue types, but most prominently in immune-cell related tissues, as thyroid autoimmunities are among the most common endocrine autoimmune features in the clinic[49,50]. Within the AutoCore, gene defects that present with both autoimmunity/autoinflammation have the highest closeness centrality (Figure 3G). Pathway diversity – calculated by the number of pathways associated with the direct neighborhood of gene defects – showed that gene defects that can present with both autoimmune/autoinflammatory phenotypes are implicated in more diverse cellular processes compared to gene defects that only present with either autoimmunity or autoinflammation (Figure 3H-I). The high topological centrality and molecular outreach of genes that present with both phenotypes suggest that they represent high fragility nodes in the network architecture of autoimmunity/autoinflammation.

The monogenic AutoCore is at the core of common, complex autoimmunity/autoinflammation

In contrast to monogenic autoimmune/autoinflammatory diseases, their polygenic counterparts are highly prevalent in Western countries[51]. While these common diseases often present with phenotypes similar to monogenic diseases, they are generally thought to result from a combination of environmental, genetic and epigenetic factors[52]. Genome-wide association studies (GWAS) have been extensively used to investigate the genetic landscape of these diseases[53], however, much of these have remained without direct mechanistic link between the identified loci and the respective diseases. Most of the heritability of polygenic autoimmune/autoinflammatory diseases thus remains unexplained[54]. Elucidating the molecular linkers between the polygenic and monogenic autoimmune/autoinflammatory diseases could aid in better understanding the contribution of significantly associated genes from GWAS to the respective pathobiology. We therefore sought to leverage the AutoCore to investigate and quantify the relationship between complex polygenic and rare monogenic autoimmunity/autoinflammation. To this end, we considered genes associated with polygenic diseases that present with a phenotypical overlap with monogenic autoimmune/autoinflammatory diseases such as rheumatological diseases (rheumatoid arthritis, systemic lupus erythematosus (SLE), Sjögren's syndrome), inflammatory bowel disease (IBD), scleroderma, as well as other well documented polygenic autoimmune/autoinflammatory diseases where significant GWAS associations have been uncovered, including type 1 diabetes, Graves' disease and multiple sclerosis (MS). We first identified the genome-wide significant genetic associations per disease from the GWAS catalog[53]. We found that the joint lcc defined by the AutoCore and genes associated with the polygenic autoimmune/autoinflammatory diseases is significantly larger than the connected component of polygenic autoimmune/autoinflammatory genes alone (Figure 4A-B). To assess the relative position of the AutoCore and complex autoimmune/autoinflammatory diseases on the interactome, we first considered the network paths between genes associated with complex autoimmune/autoinflammatory diseases (Figure 4C). We found that genes from the AutoCore frequently (359 nodes out of 399) act as connectors between complex disease genes (Figure 4D-E). These connectors are enriched in pathways such as NFkB signaling, Hepatitis B, EBV infection, viral infections, osteoclast differentiation, TCR and TNF signaling and cytokine signaling (Figure 4F-G). Viral infections are often the first triggers of polygenic autoimmune/autoinflammatory diseases, and the immune response to these heavily relies on TNF-induced NFkB activation that initiates cytokine and T-cell dependent mechanisms[55]. Figure 4H shows that the AutoCore is strongly enriched among the connectors of all considered common autoimmunity/autoinflammation disorders. We can further quantify this by measuring the network-based distance between all common diseases and the AutoCore. The results in Figure 4I clearly show that the AutoCore takes the most central position in this disease-interaction network.

In summary, we found that the AutoCore, based on rare, monogenic disease genes, serves as a central connector between common, polygenic autoimmune/autoinflammatory diseases on the interactome. From a network perspective, the central role of the AutoCore is consistent with the stronger impact of genetic lesions in this interactome neighborhood, compared to genetic perturbations in the periphery, thus

explaining the less detrimental effects that polygenic autoimmune/autoinflammatory genes have on the same molecular machinery.

Connectors of common IBD and SLE highlight monogenic genes implicated in the polygenic phenotypes

To further characterize the molecular links between monogenic and polygenic diseases, we focused on IBD and SLE, two prototypical diseases for autoinflammation and autoimmunity. To date, 67 monogenic gene defects have been described to present with an IBD-like phenotype[56], and 34 monogenic gene defects have been linked to pediatric SLE[57]. Despite the ample similarities between the common polygenic and rare monogenic diseases on the clinical level[58,59], there is only minimal overlap between monogenic and polygenic SLE or IBD on the genetic level (17 and 6 overlapping genes, respectively, Figure 5A-B). We found that the combined monogenic-polygenic disease gene sets formed significantly connected clusters for both IBD and SLE (z-scores=6.6 and 4.8, respectively; Figure 5C-D). For both diseases, the monogenic genes were more central on the interactome than the polygenic ones, however the difference was only significant in the case of IBD (Figure S3A-H). We next assessed the network paths linking polygenic IBD. We found that AutoCore genes, specifically those corresponding to monogenic IBD genes, were frequently connectors between polygenic IBD genes (49 out of 67 monogenic IBD genes, Figure 5E-F). Figure 5G shows the interactome neighborhood of the combined monogenic and polygenic IBD genes and their interactions. We found that the monogenic IBD genes acting as connectors were enriched in C-type lectin receptor signaling, NFkB signaling, toxoplasmosis-related pathways, B-cell activation and cytokine signaling (Figure S3I-J). We also found strong enrichment in a high confidence seed IBD network from the literature[60], and the cohort-specific coexpression networks derived from intestinal gene-expression data of IBD patients[61] (Figure 5H). The lectin pathway is linked to recognition and response to the microbiota in the gut[62], intestinal inflammation, apoptotic cell clearance and intestinal repair[63,64]. The enrichment in this pathway, together with the enrichment in patient intestine-derived IBD networks point to the importance of more polygenic-like, dysbiosis-linked mechanisms to act as links between polygenic and monogenic IBD. We found similar patterns for SLE, where again, monogenic SLE genes within the AutoCore served as connectors between polygenic SLE genes (Figure 5I-J). The monogenic SLE connectors (17 out of 34 monogenic SLE genes, Figure 5J-K) were enriched in pathways involved in neurotrophin signaling, DNA replication, NK-cell-mediated cytotoxicity, cytokine and interferon signaling (Figure S3K-L). Neurological phenotypes are considerably more common in adult patients than in pediatric SLE[58,59]. We found that the SLE connector genes were also enriched in genes differentially expressed in SLE cohorts[65] (Figure 5L), pointing to a central role in adult SLE. Overall, we found that polygenic-monogenic connectors in both IBD and SLE are enriched in potential phenotype-driver genes, as well as in hallmark polygenic phenotypes.

The AutoCore separates into molecularly and phenotypically cohesive subclusters

Several strategies exist for classifying monogenic autoimmune/autoinflammatory diseases into subgroups[27,28,66], often based on non-standardized clinical terms. Here, we set out to leverage the network organization of the AutoCore, combined with Gene Ontology (GO) functional annotations to identify molecularly defined subgroups (see Methods, Figure 6A-B). Our analyses resulted in 25 subclusters, with 19 out of 25 clusters containing disease genes, from here on denoted as "disease clusters" (Figure 6B, Figure S4A-B). We found that the disease clusters are phenotypically cohesive, i.e., genes within a particular cluster are significantly more phenotypically similar based on Human Phenotype Ontology (HPO) terms[67] than expected by chance (Figure S4C). Furthermore, we found that autoimmune/autoinflammatory phenotypes are not clearly separated into distinct clusters - highlighting that autoimmunity and autoinflammation are intimately linked on the molecular level (Figure S4D-E). Less specific disease phenotypes and broader clinical subclusters are often spread across several different subclusters, whereas more specific disease phenotypes and clinical groups tend to reside in only a few, or even a single cluster (Figure S4F). The correlation between the specificity of a particular clinical term and the degree of localized clustering highlights the efficacy of our network-based approach for pinning down consensus molecular mechanisms underlying the subclusters. To extend our understanding of the fundamental mechanisms of autoimmunity/autoinflammation, we sought to identify the molecular mechanisms and pathways represented by the different clusters within the AutoCore. We found that the subclusters represent distinct cellular processes and states (Figure S5A-B). The largest AutoCore subclusters pinned down survival and death receptor signaling (cluster c15), the NFkB pathway, nucleotide and RNA remodeling and repair (c20 and c7), TNF/DNA binding related mechanisms (c8), TCR and BCR signaling (c14), the complement cascade (c18), steroid resistant inflammation (c11), actin regulation (c10), and immune dysregulation through interleukin and cytokine signaling (c16), the proteasome complex (c6), as well as vesicular traffic regulation (c13) (Figure 6C, Figure S5A-B) as major building blocks of autoimmunity/autoinflammation.

Using the AutoCore to highlight clinically targetable pathways

It has been shown previously that network proximity between drug targets and disease genes may serve as an effective predictor for a therapeutic effect[68,69]. This prompted us to investigate if we can leverage the interactome to identify potentially efficacious novel therapeutic pathways within the AutoCore. We collected available treatments for rare autoimmune/autoinflammatory diseases and their respective protein targets and measured the minimal network distance between each drug and the AutoCore that has previously been shown to be the most accurate measure of efficacy[68] (Figure 7A, Supplementary Table 1). We first established that targeted treatments already used for autoimmune/autoinflammatory diseases were closer to the AutoCore than expected by chance (Figure 7B). Next, we sought to identify if the network distance measure is able to differentiate the different therapeutic approaches used for the different clusters. We found that the targets of targeted treatments already used for diseases within a certain cluster are significantly closer to the cluster on the interactome, as compared to other drugs not used for the treatment

of diseases within a cluster (Figure 7C-D). With this proof of principle for the predictive power of network distance for treatment-cluster associations, we next aimed to identify novel pathways that are not yet exploited for the treatment for monogenic autoimmune/autoinflammatory diseases. To this end, we used a list of all drugs from DrugBank[70] and calculated the network-based minimal distance of their respective targets from the AutoCore (see Methods, Figure 7E). We found ample overlap between the pathways enriched among the targets of existing treatments of autoimmune/autoinflammatory diseases (87 out of 109, Figure 7F) and the pathways enriched among targets of novel treatment candidates as suggested by the network-based ranking. The candidate targets were enriched in several novel pathway associations, including downstream TCR signaling, platelet activation and Fce receptor signaling (Figure 7G). Unspecific modulators of downstream TCR signaling through the NFkB pathway such as glucocorticoids and NSAIDs are already utilized in the treatment of autoimmunity/autoinflammation. More recently, preclinical studies are investigating more specific manipulation of NFkB signaling, such as targeting the IKK complex or IKb[71]. In addition to the use of targeted checkpoint inhibitor for CTLA4[72–74], the efficacy of other therapies targeting TCR signaling such as the PD-1 checkpoint in murine models[75], ICOS-ICOSL interactions[76,77], blocking CD40-CD40L signaling[78] or the OX40-OX40L[79,80] costimulatory pathway or the use of multivalent therapeutics[81] are already being investigated[82]. Strategies of blocking or neutralizing Fc receptors as ways to treat inflammation are also under review[83], and certain monoclonal antibodies have shown efficacy in treating RA[84]. Taken these together, the AutoCore highlights those already utilized but so far only indirectly targeted pathways that show potential in treating autoimmune/autoinflammatory diseases.

Discussion

The premise of our work was to construct a core molecular network of human autoimmunity/autoinflammation. To our knowledge, this work represents the first effort to systematically explore the relationships among a heterogeneous group of monogenic diseases using a network-based framework. We showed that high-confidence rare disease genes can be used to identify the AutoCore, a localized subnetwork of the interactome that can aid in the interpretation of a wide range of previously disconnected biological and clinical observations. We found that autoinflammation and autoimmunity do not separate within the AutoCore, corroborating theories that propose an "immunological disease continuum"[85] of such conditions, both on an individual patient and a population level. Genes that present with both autoimmunity/autoinflammation displayed a high pathway diversity, i.e., a tendency to influence diverse molecular processes. We further found that the AutoCore lies at the center of common, polygenic autoimmune/autoinflammatory diseases. This finding is consistent with the recently proposed omnigenic model of complex diseases that hypothesizes that genes at the core and at the periphery of regulatory networks contribute differently to the heritability of complex traits[86,87]. Although the general applicability of this model to complex traits is debated[88], we find good agreement when considering monogenic and polygenic versions of the autoimmune/autoinflammatory phenotypes, using the SLE and IBD cohorts as examples. This implies that instead of separate disease entities, autoimmune/autoinflammatory conditions may represent genetically determined extreme forms of more common human autoimmune/autoinflammatory conditions. We found that the pathways that interconnect mono- and polygenic diseases are strongly enriched in host defense-related mechanisms. These linkers can therefore be regarded as triggers for autoimmunity/autoinflammation in polygenic diseases, that link them to more fine-tuned molecular mechanisms resulting in severe monogenic autoimmune/autoinflammatory, such as viral infections and the subsequent inflammatory response, dysbiosis in IBD, or highly expressed expression signatures in SLE. Previous efforts to classify autoimmune/autoinflammatory diseases were based on reclassifications of clinical terms[89], grouping diseases based on mostly qualitative criteria[90], or relying on public datasets that are usually not available for rare autoimmune and autoinflammatory conditions[17]. Here, we propose a quantitative, network-based approach for subdividing the AutoCore into 25 subclusters, 19 of which disease clusters, that were found to be phenotypically and molecularly cohesive. Furthermore, we found that the clusters were also cohesive from a therapeutic perspective, such that drugs that are used to treat diseases within a particular subcluster also target close interactome areas. We hypothesize that network-distance measures within the AutoCore can thus be used to inform novel therapeutic avenues or repurposing strategies, in line with similar recent efforts in the context of complex diseases, where network distance has been shown to be predictive of drug efficacy and synergistic effects of drug combinations[69,91]. Indeed, we find multiple novel pathways enriched in the targets of those drugs close to the AutoCore. Finally, we hope that our work may contribute to establishing a closer connection between the clinical experts of a heterogeneous, disjointed group of single-gene diseases, and systems-based methods for elucidating similarities and differences between them. Our work represents a versatile

platform for addressing a wide range of important challenges ranging from the elucidation of disease relationships to potential novel therapeutic approaches. The platform enables us to quantify and integrate isolated and often descriptive clinical observations and translate them to machine interpretable data. It can thus be used to address other diseases/disease groups in the future. Our web tool, located at https://menchelab.com/autocore allows for an interactive exploration of the AutoCore.

Materials & Methods

Identification of monogenic gene defects underlying autoimmunity and autoinflammation

We performed a literature review of IEIs based on the 2017 IUIS classification[27] to identify monogenic gene defects underlying autoimmunity and autoinflammation. We queried for autoimmune/autoinflammatory phenotypes using the IUIS classification, as well as our recent review of autoimmune manifestations in monogenic IEIs[92] and inflammatory bowel disease[56]. Only gene defects with evidence of causality were included. If the disease or gene defect was not part of the IUIS classification of IEIs, we required at least one case report with autoimmune and/or autoinflammatory phenotypes. First year of publication, inheritance pattern, and type of mutation were identified through the OMIM database [28]. In case OMIM entries were not available, case reports or review articles were used to access the information. Available therapeutic strategies were accessed from the Genetic and Rare Disease Information Center (GARD) database[30], as well as manually curated from the literature. The detailed list of gene defects, disease phenotypes and treatment options is available in Supplementary Table 1. Monogenic gene defects underlying inflammatory bowel diseases were extracted from[56], monogenic SLE gene defects were extracted from[57].

GWAS data

We used the GWAS catalog[53] to identify significant SNPs (genome-wide p-value<0.05) associated with type 1 diabetes (T1D), systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), Graves diseases, multiple sclerosis (MS), inflammatory bowel disease (IBD), scleroderma and Sjögren's syndrome. In case of duplicate SNPs, we chose the lowest p-value. We followed the official GWAS catalog guidelines to obtain SNP-to-gene associations, taking the "reported genes" as genetic associations. The detailed curation process can be found here: https://www.ebi.ac.uk/gwas/docs/methods/curation. We used GWAS SNPs coming from multiple studies, with varying cohort sizes and arrays. Although we took only significant associations from GWAS catalog, as the statistical significance of a particular SNPs could differ based on what study the SNP was identified in, we opted to take an inclusive, gene-based approach for our network analysis. This means that we did not differentiate between GWAS genes based on how many SNPs have been identified in a particular gene, or the magnitude of the p-value (as long as it reached genome-wide significance). The full list of genes and SNPs used for the genetic analysis is available in Supplementary Files 7-8.

Measure of genetic vulnerability

We used the Genome Aggregation Database (gnomAD)[31] to extract information on genetic vulnerability. Intolerance of loss of function (pLoF) scores were extracted on 20.05.2020. The Mann-Whitney U nonparametric test was used to determine the significance of difference of the set of pLOF scores from random expectation.

Pathway enrichment

Pathway enrichment and comparison was performed by the Enrichr tool[93] the reactomePA[94] package in R (R Foundation for Statistical Computing, https://www.R-project.org/) or an in-house Python (Python Software Foundation, https://www.python.org/) script based on the KEGG[95] and reactome[96] pathway datasets from the Molecular Signature Database[97]. Fisher's exact test from the SciPy[98] stats package in Python was used to define enrichment odds ratio and p-value. We have used reactomePA[94] to determine the significantly enriched reactome pathways among drug target sets.

Gene-set enrichment

The seed IBD network and IBD-cohort-based coexpression networks were extracted from[60,61]. The SLE transcriptomic signature dataset was accessed from[65]. Frequently mutated genes in cancer were extracted from[42]. The cancer expression dataset was downloaded from The Human Protein Atlas[44] on the 28.12.2020. Highly expressed genes were identified as those which had a 'High' expression annotator in at least 50% of all patients in a particular tumor type. Lists of drugs, chemicals and their targets was extracted from the DrugBank database[70] and are available in Supplementary File 14. Odds ratio and p-values were calculated using Fisher's exact test as implemented in the SciPy[98] Python package.

Network-based methods

Networks

We combined three widely used interactomes, harnessing their advantages: i) an interactome widely used for identifying links between diseases[22] ii) a large scale dataset from the HIPPIE[34] database (v2.2), iii) and a systematic high-throughput interactome that was previously shown to be able to draw meaningful connections in Mendelian diseases[35]. The joint interactome is larger and denser than the individual ones, but follows the same network characteristics as previously published interactomes, in particular concerning the close relationship between network distance and biological function. The aggregate network consists of 18,853 nodes and 483,037 edges. The full list of nodes and edges of the interactome can be found in Supplementary File 1. To represent different dimensions of the cellular machinery, in addition to the interactome, we included coexpression networks that represent genes that are linked through transcriptional and machinery[39]. We used 16 transcriptome-wide coexpression networks from[38,91]. The networks were built by combining total expression and expression of relative isoform levels into single sparse networks using data from the GTEx database[99] (v6), based on 449 human donors with genotype information and 7310 RNA-seq samples across 50 tissues. Sampled donors were 83.7% European American and 15.1% African American. The full list of nodes and edges can be found in Supplementary File 3.

Network-based characteristics

The Python package NetwrokX[100] was used to compute network-based features of genes and gene sets. *Degree* is defined as the number of connections (direct neighbors) a node has on a network. *Closeness centrality* measures the inverse of the average distance from a given node in the network to all others. The more central a node is, the smaller are the shortest path lengths to all other nodes. It is calculated by

$$C(x) = \frac{N}{\sum_y d(y,x)}$$

where *N* is the number of nodes in the graph, *d(x,y)* is the distance between nodes *x* and *y* on the network. *Localized closeness centrality* measures the relative closeness centrality of a group of nodes on a network. It is calculated by:

$$C_{loc}(x) = Mgroup \div \sum_y d(y,x)$$

where $Mgroup$ is the number of nodes in the group, and *d(x,y)* is the distance between nodes x and y on the network. *Betweenness centrality* of a node *x* measures how many of all shortest paths in the network are passing through node x. It is given by

$$g(x) = \sum\nolimits_{s \neq x \neq t} \frac{\delta_{st(x)}}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node *s* to node *t* and (*x*) is the number of shortest paths that pass through *x*. *Eigenvector centrality* of a node is a measure of its influence over the network[101]. A high eigenvector centrality means that a node is connected to many nodes with high eigenvector scores. For a graph G=(V,E) with vertices V, let A=($a_{v,t}$) be the adjacency matrix, i.e. $a_{v,t}$ =1 if vertex *v* is linked to vertex *t* and $a_{v,t}$ =0 otherwise. The relative centrality, x of a vertex *v* can then be defined as:

$$x_V = \frac{1}{\lambda} \sum_{t \epsilon M(v)}$$

$$x_t = \frac{1}{\lambda} \sum_{t \epsilon G} a_{v,t}\, x_t$$

where M(v) is a set of the e neighbors of v and lambda is a constant. The eigenvector equation stands as: Ax = x.

Connectivity measures

Largest connected component: The largest connected component of a gene set on a network is defined as the subnetwork formed by the gene set on the network that consist of only gene nodes of the gene set and

their direct interactions. The size of the largest connected component of random subsets of N nodes is expected to follow a normal distribution, given is larger than the percolation threshold. We can therefore empirically estimate the significance of a given connected component through its z-score and corresponding empirical p-value determined from 1,000 randomly selected gene sets of the same size.

Distance measures

Shortest distance: a shortest distance or path between two nodes is a path with the minimum number of edges between the two nodes.

Average distance: Distance between two sets of nodes on a network was calculated by averaging the pairwise shortest distances of each node-node pair in the two node sets. The significance of a measured average distance was assessed by comparing it to 1,000 random permutations and calculating the z-score, as well as the empirical p-value. *Minimum distance:* The minimal distance between two sets of nodes on a network was calculated by the minimum distance from the pairwise shortest distance measurements in each gene-gene pair in the two node sets. The significance of minimum distance was assessed by comparing it to 1,000 random permutations and calculating the z-score, as well as the empirical p-value.

Pathway diversity

The pathway diversity of a particular gene set quantifies the normalized number of unique KEGG and reactome pathways annotated to the set of the first neighbors of the gene set on the interactome. The pathway diversity is calculated by dividing the number of unique pathways by the number of nodes in the neighborhood. For a particular group of nodes, pathway diversity is thus given by:

$$(PathDiv)^{group} = X/N$$

Where *X* is the number of unique pathways among all neighbors of the gene set, and *N* is the number of all nodes in its neighborhood.

Identification of the *AutoCore*

We used a random walk with restart algorithm[102], using the monogenic autoimmune/autoinflammatory disease genes as seed genes and restart probability r=0.9. This resulted in a list of all genes in the interactome ranked by decreasing visiting probabilities representing their distance to the seed. From this list, genes are added to the seed graph of monogenic autoimmune/autoinflammatory genes until the graph forms a fully connected component. The resulting AutoCore subgraph contains all seed genes and a minimal set of linker genes according to the performed random walk given its specific restart value.

Finding molecular subclusters within the *AutoCore*

In a first step, a random-walk was used to identify neighbors for every gene in the AutoCore. The set of random-walk neighbors is not necessarily equal to the set of genes given by immediate connections, as it

takes into account structural network properties such as interconnections among neighbors or their degree. For all these neighborhoods gene set enrichment analysis for GO annotations (biological process, molecular function, cellular component) was performed to identify significant annotations according to Fisher's exact test. Next, every single gene in the network neighborhood of a particular gene received these additional annotations to complement the plain database knowledge with information induced by the network context. Former poorly annotated genes can thus be equipped with function due to their location in a specific functional network region (deorphanization). Annotated GO functions were used as features to construct a gene-based multi-dimensional feature space. A gene-gene matrix was calculated from the feature matrix using cosine similarity. This matrix was clustered using hierarchical clustering. The optimal number of clusters was determined by keeping the number of genes in the largest cluster small while minimizing the number of clusters consisting of only a single gene.

GO term enrichment within cluster

For the GO annotation and enrichment of nodes within clusters, we consider GO annotations from two sources. (1) plain annotations of a node as from the GO database. (2) network annotations of a node as based on its network neighborhood. For this, we calculated the enrichment of specific GO terms among the nodes within the network neighborhood using Fisher's exact test. Finally, for a given cluster and GO term X, we determine the number of genes that GO term X is linked to by plain association and by network annotation, and compute the combined enrichment level according to:

$$Enrich(X) = \frac{Freq(X)_{all} - Freq(X)_{plain}}{Freq(X)_{plain}}$$

Where $Freq(X)_{all}$ is the total frequency of term X in the cluster (plain and network annotation combined) and $Freq(X)_{plain}$ is the plain annotation frequency of a term in the cluster.

Polygenic – monogenic autoimmune/autoinflammatory connector nodes

The connector nodes between polygenic autoimmune/autoinflammatory diseases were defined as those AutoCore nodes that were among the shortest paths between polygenic disease genes. IBD and SLE specific polygenic disease connectors were defined as those monogenic IBD or SLE genes that were among the shortest paths between polygenic IBD or SLE disease genes. Enrichment (odds ratio and p-value) of connector nodes in various datasets was calculated by Fisher's exact test with the SciPy[98] stats package in Python (Python Software Foundation, https://www.python.org/).

Phenotype similarity

Phenotype similarity of two genes was based on Human Phenotype Ontology[67], release 09.10.2018. Semantic similarity was is computed for all annotated gene pairs using information content-based Resnik's similarity. In general, gene pairs annotated with terms that are deeper in the ontology structure are given higher similarity scores. Phenotypic similarity of a group of nodes within a cluster was calculated by comparing the distribution of similarity scored within gene pairs in a cluster to 1,000 random simulations. The Mann-Whitney U nonparametric test was used to determine the significance p-value.

Network visualization

The 2D-network layouts in Figure 2, Figure 4 and Figure 5 were visualized in Cytoscape[103], using the Organic layout. Visualization of the AutoCore in Figure 3 and 6 was carried out with UMAP[104] dimensionality reduction, that takes the feature matrix with all AutoCore genes as samples and GO-terms as features to find an optimal way to project similar genes based on their share features into three-dimensional space.One of the drawbacks of systems-methods are the challenges in visualizing and interactively exploring large and complex datasets. To address these challenges, we developed a web-app, and also included the AutoCore into a virtual reality (VR) application. The web-app allows for browsing through the database that contains gene and disease-based information we compiled for this publication and for a 2D interactive visualization of the AutoCore. The VR-based visualization is based on the VRNetzer platform[105] and enables an intuitive, immersive discovery of the complex AutoCore network in 3D.

Funding

The study was supported by the European Research Council (ERC, Consolidator Grant 820074 "iDysChart"), the Austrian Science Fund (FWF) project P29951-B30 (both to K.B.) and the Vienna Science and Technology Fund (WWTF) through project VRG15-005 (to JM).

Data Availability

Code analysis scripts are available at: https://github.com/jpazmandi/AutoCore. All network files are available as edge list and Cytoscape formats in Supplementary Files 16-19. Publicly available databases used for this study include OMIM (https://www.omim.org/), OrphaNet (https://www.orpha.net/consor/cgi-bin/index.php), GARD (https://rarediseases.info.nih.gov/about-gard/pages/23/about-gard), MSigDB (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp), the GWAS catalog (https://www.ebi.ac.uk/gwas/) and DrugBank (https://go.drugbank.com/). All other data are available in the main text, supplemental table, supplementary files, as well as on https://menchelab.com/autocore.

559 **List of Supplementary Data**

560

Supplementary Table 1	Detailed table on monogenic autoimmune and autoinflammatory gene defects and diseases.
Supplementary File 1	Edge list of the combined interactome.
Supplementary File 2	Network characteristics of monogenic autoimmune/autoinflammatory genes on the interactome.
Supplementary File 3	Edge lists of all coexpression networks from the GTEX resource.
Supplementary File 4	List of nodes in the AutoCore.
Supplementary File 5	AutoCore edge list.
Supplementary File 6	List of drugs used for treatment of monogenic autoimmune/autoinflammatory diseases, with their targets.
Supplementary File 7	List of significant GWAS SNPs associated with polygenic autoimmune and autoinflammatory diseases from the GWAS catalog.
Supplementary File 8	List of reported genes pinpointed by the significant GWAS SNPs associated with polygenic autoimmune and autoinflammatory diseases from the GWAS catalog.
Supplementary File 9	List of polygenic autoimmune/autoinflammatory disease connectors.
Supplementary File 10	Monogenic SLE and IBD gene test, along with polygenic IBD, and polygenic SLE connectors.
Supplementary File 11	Edge list of the intestine-derived IBD networks.
Supplementary File 12	List of the molecular subclusters within the AutoCore.
Supplementary File 13	GO annotation of AutoCore clusters.
Supplementary File 14	List of drugbank drugs used for the analysis in Figure 7.
Supplementary File 15	List of novel drugs that are significantly close to the AutoCore.
Supplementary File 16	Largest connected component of monogenic autoimmune/autoinflammatory genes, Cytoscape file.
Supplementary File 17	The AutCore, Cytoscape file.
Supplementary File 18	Polygenic-monogenic network: IBD, Cytoscape file.
Supplementary File 19	Polygenic-monogenic network: SLE, Cytoscape file.

561

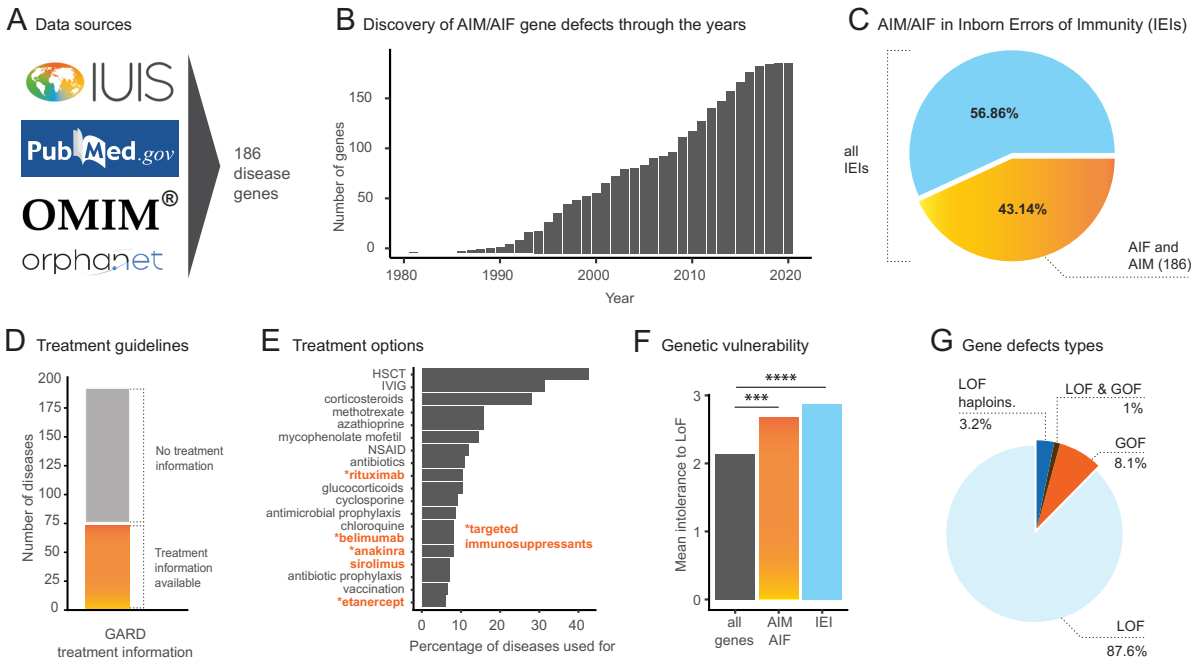


Figure 1: Charting the genetics of monogenic autoimmunity and autoinflammation. A) Resources for data collection on monogenic autoimmune/autoinflammatory (AIM/AIF) gene defects. B) Discovery of monogenic autoimmune/autoinflammatory diseases through the years. C) Percentage of gene defects with autoimmunity/autoinflammation within inborn errors of immunity (IEI). D) Treatment information availability for monogenic autoimmune/autoinflammatory diseases in GARD. E) Types of treatment available for monogenic autoimmune/autoinflammatory diseases, full list available in Supplementary Table 1. F) Tolerance to loss-of-function (LOF) variation of IEIs and autoimmune/autoinflammatory gene defects. G) Percentage of (LOF), gain-of-function (GOF) gene defects within monogenic autoimmune and autoinflammatory diseases. Abbreviations: IUIS: International Union of Immunological Societies; OMIM: Online Mendelian Inheritance in Man; AIM: autoimmunity; AIF: Autoinflammation; IEI: Inborn Errors of Immunity; avail.: available; GARD: Genetic and Rare Diseases Information Center; HSCT: Hematopoietic stem-cell transplantation; IVIG: Intravenous Immunoglobulin; NSAID: nonsteroidal anti-inflammatory drugs; Encap. bact: encapsulated bacteria; LOF: Loss of function; GOF: Gain of function; LOF haploins: LOF haploinsufficiency. AIM: Autoimmunity; AIF: Autoinflammation; * = p<0.05, ** = p<0.01, *** = p<0.001, *** = p<0.0001, **** = p<0.00001.

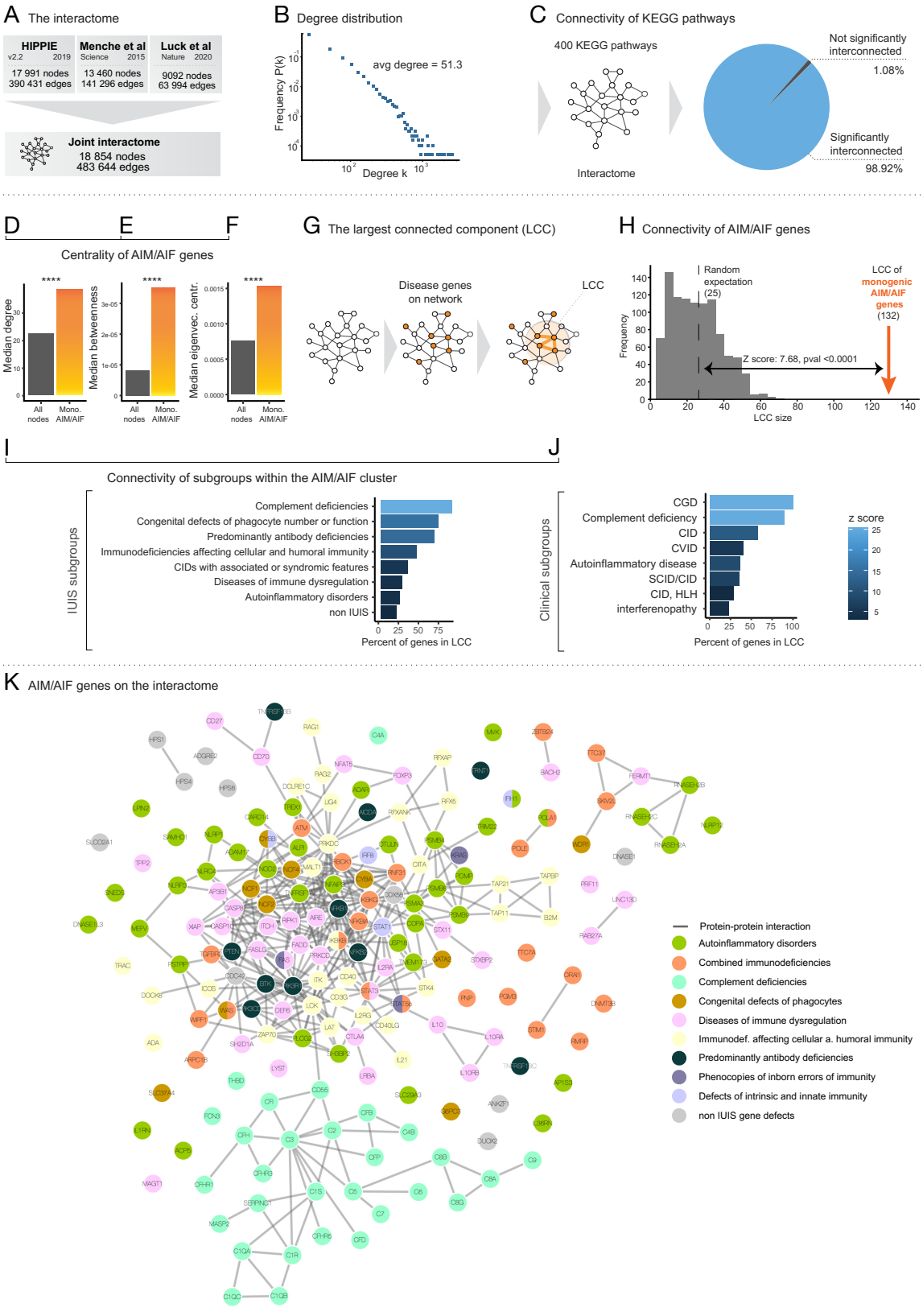

A The interactome
HIPPIE v2.2 2019
17 991 nodes
390 431 edges
Menche et al Science 2015
13 460 nodes
141 296 edges
Luck et al Nature 2020
9092 nodes
63 994 edges
Joint interactome
18 854 nodes
483 644 edges
B Degree distribution
avg degree = 51.3
Frequency P(k)
Degree k
C Connectivity of KEGG pathways
400 KEGG pathways
Interactome
Not significantly interconnected
1.08%
Significantly interconnected
98.92%
D E F
Centrality of AIM/AIF genes
Median degree
Median betweenness
Median eigenvec. centr.
All nodes
Mono. AIM/AIF
G The largest connected component (LCC)
Disease genes on network
LCC
H Connectivity of AIM/AIF genes
Random expectation (25)
LCC of monogenic AIM/AIF genes (132)
Z score: 7.68, pval <0.0001
Frequency
LCC size
I J
Connectivity of subgroups within the AIM/AIF cluster
IUIS subgroups
Complement deficiencies
Congenital defects of phagocyte number or function
Predominantly antibody deficiencies
Immunodeficiencies affecting cellular and humoral immunity
CIDs with associated or syndromic features
Diseases of immune dysregulation
Autoinflammatory disorders
non IUIS
Percent of genes in LCC
Clinical subgroups
CGD
Complement deficiency
CID
CVID
Autoinflammatory disease
SCID/CID
CID, HLH
interferenopathy
z score
K AIM/AIF genes on the interactome
Protein-protein interaction
Autoinflammatory disorders
Combined immunodeficiencies
Complement deficiencies
Congenital defects of phagocytes
Diseases of immune dysregulation
Immunodef. affecting cellular a. humoral immunity
Predominantly antibody deficiencies
Phenocopies of inborn errors of immunity
Defects of intrinsic and innate immunity
non IUIS gene defects

Figure 2: Causative genetic defects for monogenic autoimmune and autoinflammatory diseases constitute a seed network. A) Data sources to build the joint interactome. The interactome encompasses three protein-protein interaction networks. B) Degree distribution of the joint interactome. C) Connectivity of KEGG pathways on the interactome. The largest connected component (lcc) of KEGG pathways was calculated. Significance was determined by z-score based on 1,000 random permutations. D-F) Median degree-, betweenness-, and eigenvector centrality of autoimmune/autoinflammatory (AIM/AIF) gene defects. The Mann-Whitney U nonparametric test was used to determine p-value. G) Defining the largest connected component (lcc) of monogenic autoimmune/autoinflammatory gene defects. H) Connectivity of monogenic autoimmune/autoinflammatory disease genes on the interactome, measured by the largest connected component. Significance was determined by z-score based on 1000 random permutations. I) Lccs of the major IUIS subgroups of IEIs within the subnetwork of autoimmune/autoinflammatory genes on the interactome. J) Lccs of major clinical phenotypes within the subnetwork of autoimmune/autoinflammatory genes on the interactome. K) The subnetwork of monogenic autoimmune/autoinflammatory on the interactome. Nodes are colored according to the major IUIS subgroups. Abbreviations: IUIS: International Union of Immunological Societies; AIM: Autoimmunity; AIF: Autoinflammation; Lcc: Largest connected component. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, *** = $p<0.0001$, **** = $p<0.00001$.

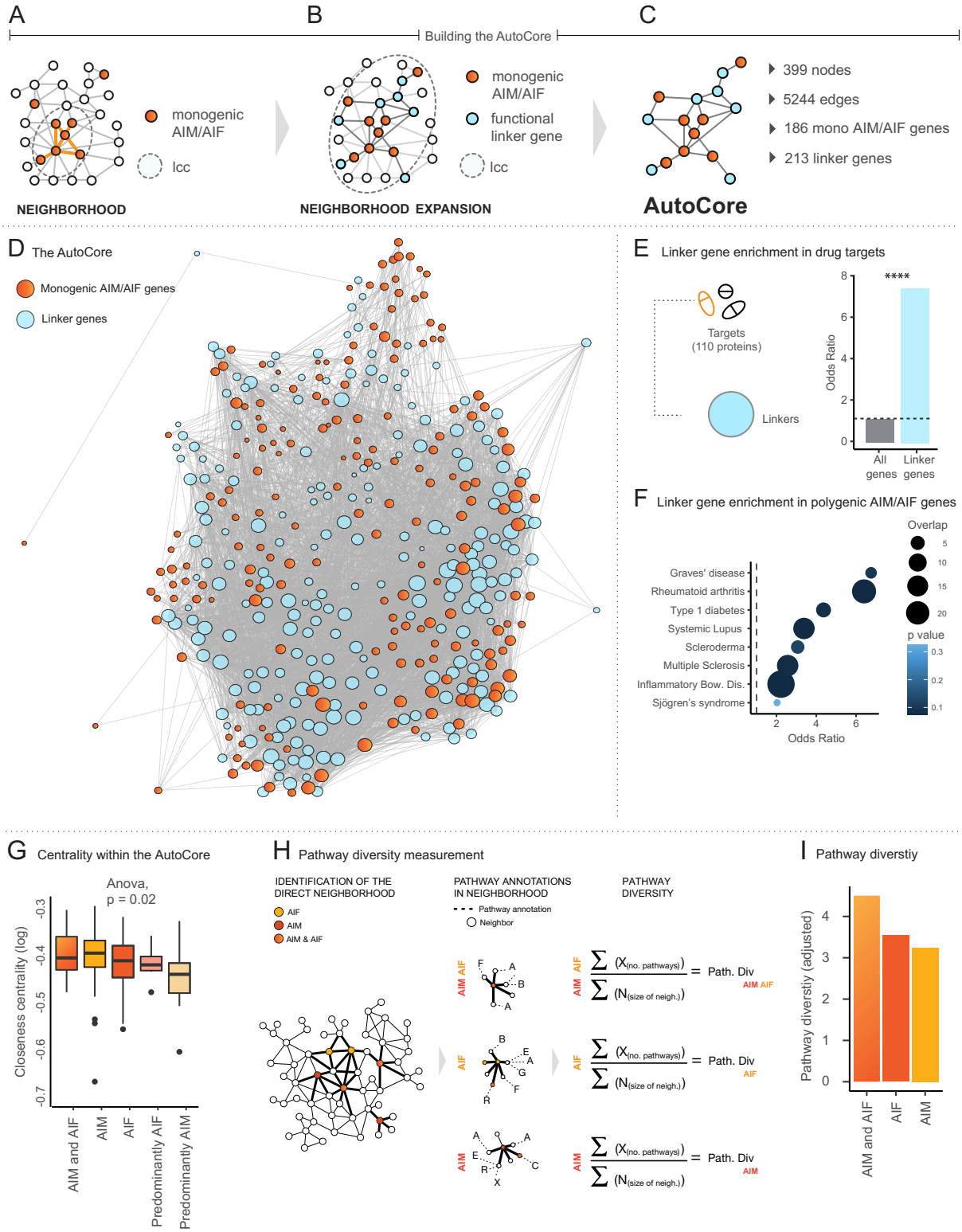


Figure 3: Identification of the functional AutoCore. A-B-C) Defining the AutoCore using monogenic autoimmune/autoinflammatory (AIM/AIF) gene defects on the interactome D) The AutoCore, a fully connected subnetwork of monogenic autoimmunity and autoinflammation on the interactome. D) The

AutoCore is enriched in targets of those drugs that are currently used to treat autoimmune and autoinflammatory conditions. E) Enrichment if linker genes in drug targets used in monogenic autoimmune/autoinflammatory diseases. F) The AutoCore is enriched in genes associated with polygenic, common autoimmune/autoinflammatory conditions. Odds ratio and p-value were determined by Fisher's exact test. G) Closeness centrality of different autoimmune/autoinflammatory phenotypes in the AutoCore. Significance of the difference between the different groups was calculated by ANOVA. H) Pathway diversity as a measurement of molecular "outreach" of gene defects. I) Pathway diversity of AIM/AIF gene defects. Abbreviations: AIM: Autoimmunity; AIF: Autoinflammation; Inflam. Bowel. Dis: Inflammatory bowel disease. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, *** = $p<0.0001$, **** = $p<0.00001$.

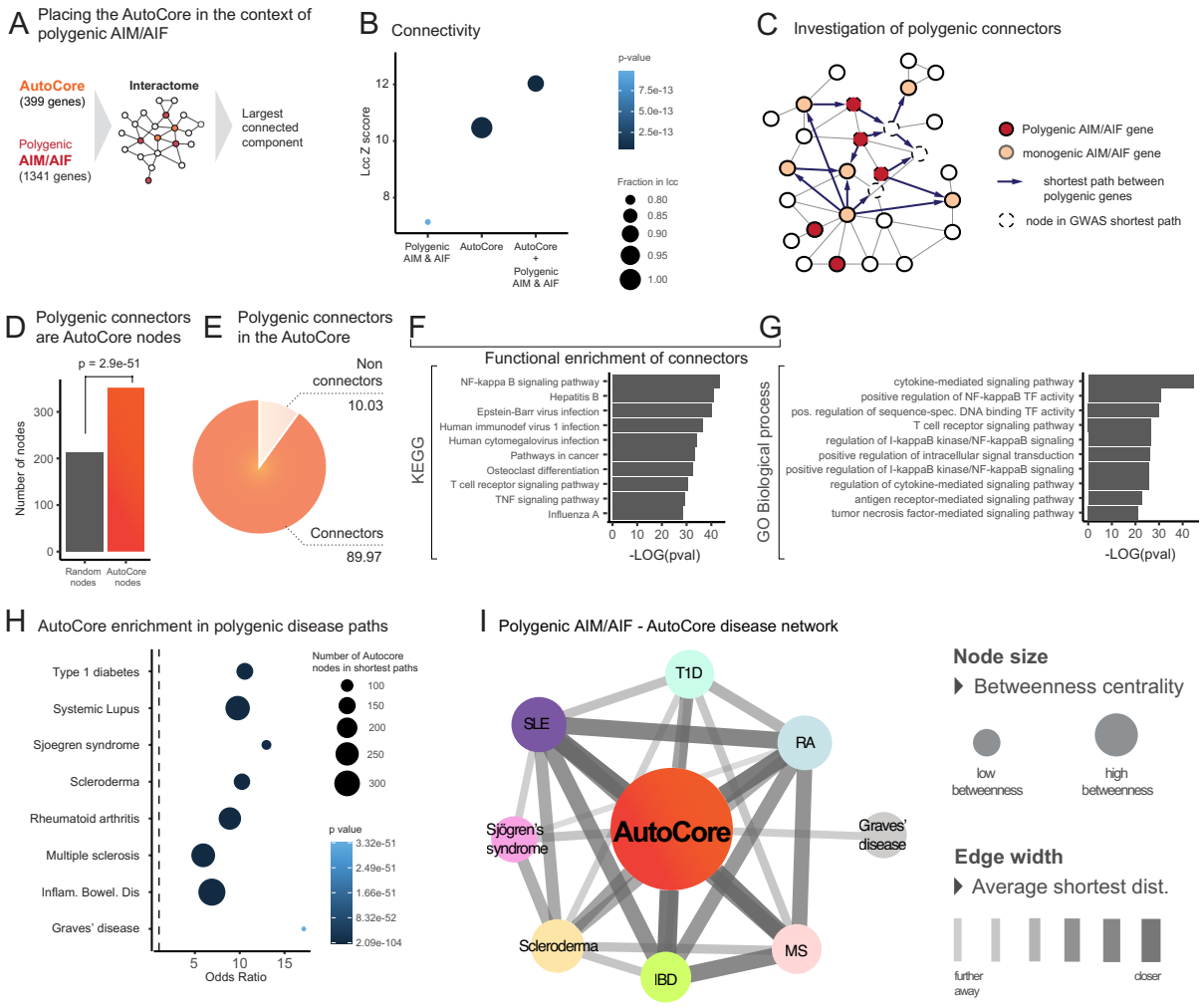


Figure 4: The monogenic AutoCore is the core of complex, polygenic autoimmune and autoinflammatory diseases. A) Finding the connected components of polygenic autoimmunity/autoinflammation (AIM/AIF) disease and the AutoCore. B) The largest connected component of polygenic autoimmune/autoinflammatory diseases and the AutoCore is more significantly larger compared to the polygenic autoimmunity/autoinflammation connected component alone. Significance was determined by z-score based on 1000 random permutations. C) Assessment of connections of polygenic autoimmune/autoinflammatory diseases based on shortest paths between polygenic genes. D) Enrichment of AutoCore nodes between polygenic disease gene shortest paths. E) Fraction of AutoCore nodes that serve as connectors between polygenic autoimmune/autoinflammatory diseases. P-value was determined by Fisher's exact test. F) Pathway and G) Gene ontology biological process enrichment of polygenic autoimmunity/autoinflammation connector genes within the AutoCore. H) Enrichment of AutoCore nodes between disease specific polygenic disease gene shortest paths on the interactome. Odds ratio and p-value was determined by Fisher's exact test. I) Polygenic autoimmune/autoinflammatory disease and AutoCore disease network. Each shortest path shorter than 1.5 is visualized. Node size represents

624 betweenness centrality in the network. Edge width represents the average shortest distance between
625 diseases. Abbreviations: AIM: Autoimmunity; AIF: Autoinflammation; Lcc: Largest connected component;
626 Inflam. Bowel. Dis: Inflammatory bowel disease. SLE: Systemic lupus erythematosus; IBD: Inflammatory
627 bowel disease; MS: Multiple sclerosis; RA: Rheumatoid arthritis; T1D: Type 1 diabetes. * = $p<0.05$, ** =
628 $p<0.01$, *** = $p<0.001$, *** = $p<0.0001$, **** = $p<0.00001$.
629

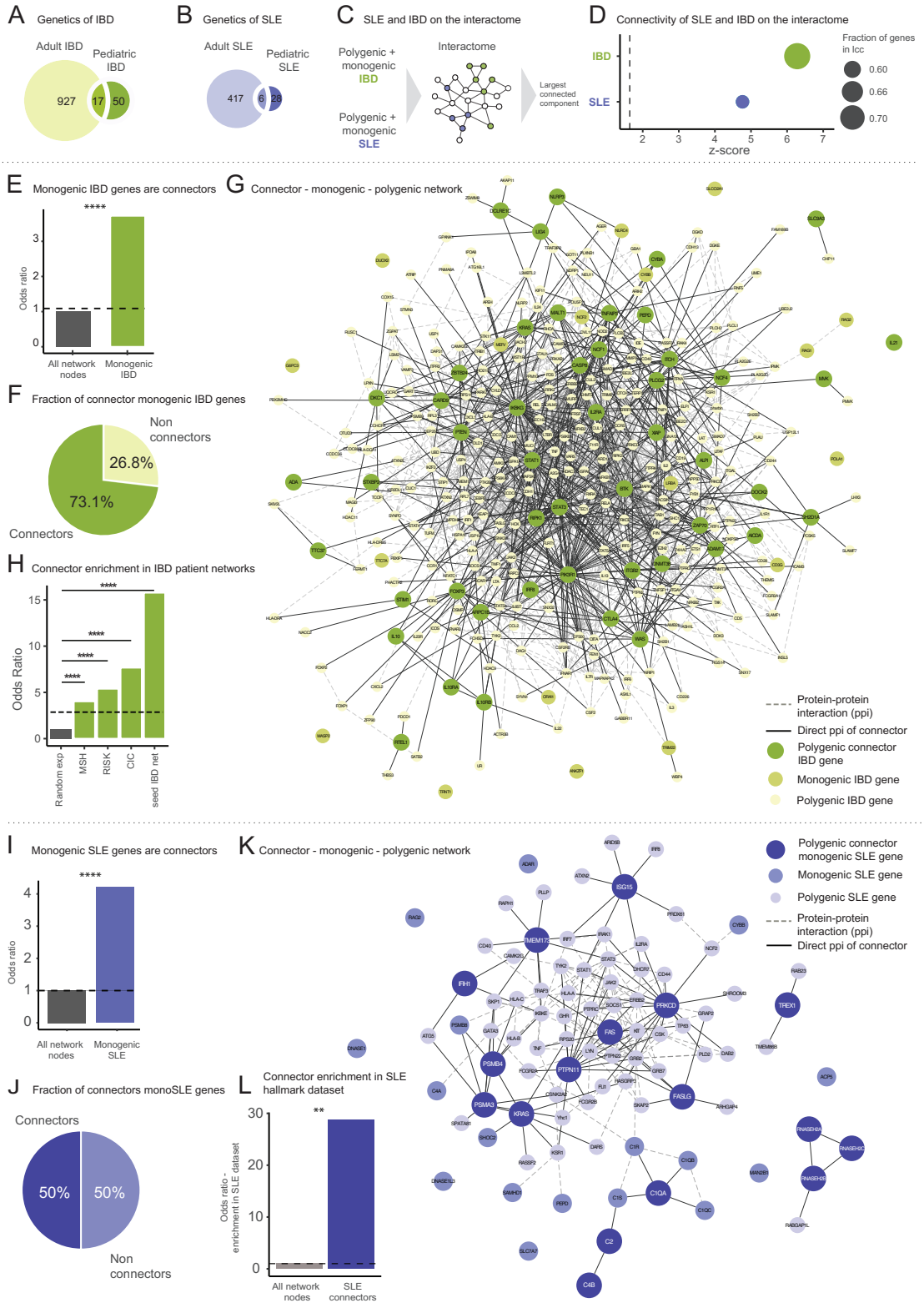


Figure 5: Disease – specific connections in SLE and IBD. A) Genetic overlap between monogenic and polygenic IBD. B) Genetic overlap between monogenic and polygenic SLE. C) Assessment of the

connectivity of monogenic and polygenic IBD and SLE. D) Connectivity of monogenic and polygenic SLE and IBD on the interactome. Significance was determined by z-score based on 1,000 random permutations. E) Enrichment of monogenic IBD genes in polygenic IBD shortest paths. Odds ratio and p-value were determined by Fisher's exact test. F) Fraction of polygenic connectors among monogenic IBD genes. G) The polygenic-monogenic IBD network on the interactome. Monogenic IBD genes, monogenic connectors and their polygenic direct neighbors are visualized. H) Enrichment of monogenic-polygenic IBD connectors in networks derived from intestinal coexpression networks from IBD patients. Odds ratio and p-value were determined by Fisher's exact test. I) Enrichment of monogenic SLE genes in polygenic SLE shortest paths. Odds ratio and p-value were determined by Fisher's exact test. J) Fraction of polygenic connectors among monogenic SLE genes. K) The polygenic-monogenic SLE network on the interactome. Monogenic SLE genes, monogenic connectors and their polygenic direct neighbors are visualized. L) Enrichment of monogenic-polygenic SLE connectors in highly expressed genes in adult SLE patients. Odds ratio and p-value were determined by Fisher's exact test. Abbreviations: IBD: Inflammatory bowel disease; SLE: Systemic lupus erythematosus; Lcc: Largest connected component; * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, *** = $p<0.0001$, **** = $p<0.00001$.

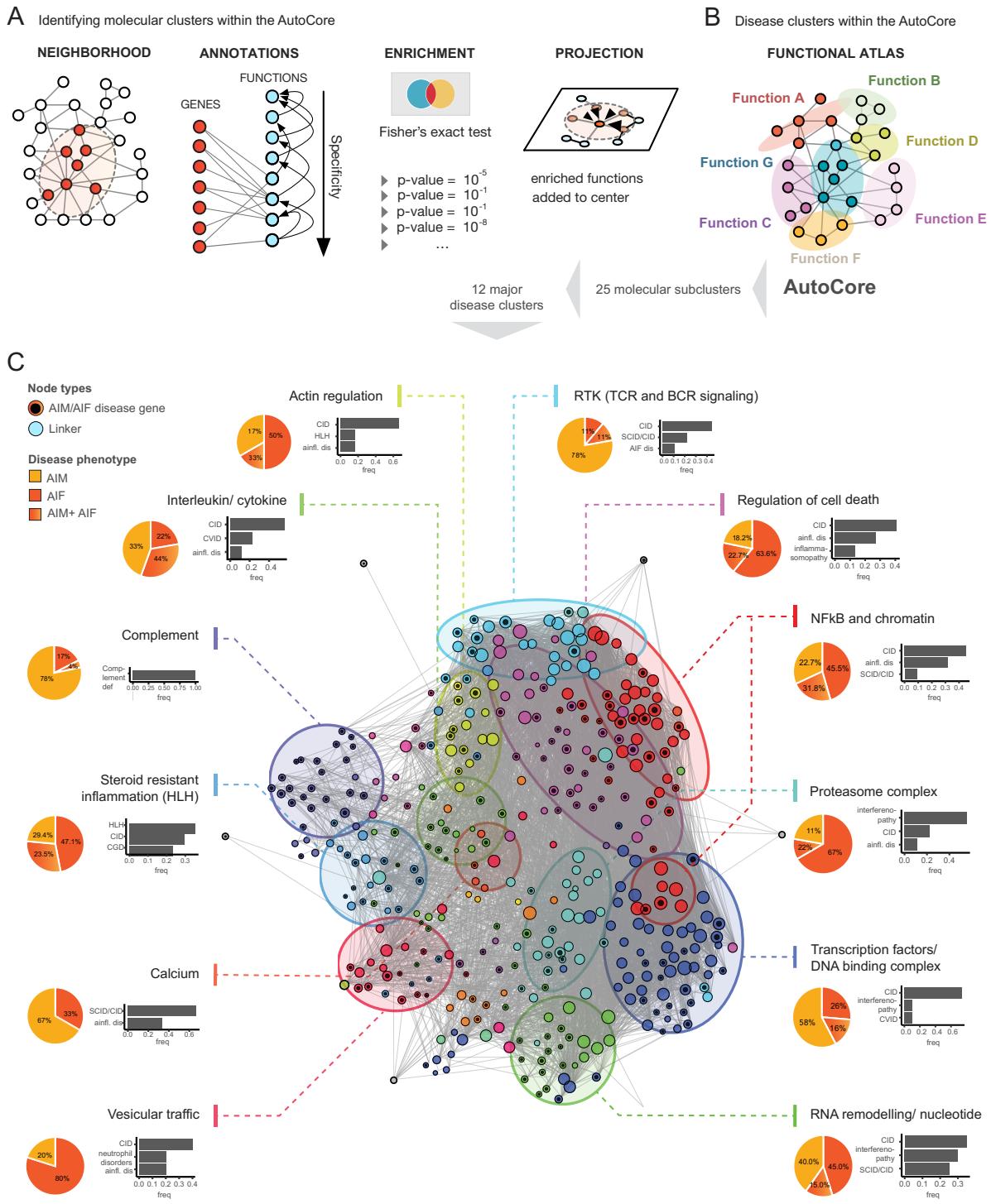


Figure 6: Finding subclusters within the AutoCore. A-B) Identification of molecular subclusters within the AutoCore. B) The AutoCore separates into 25 molecular subclusters, among them 19 disease clusters. C) Functional clusters within the AutoCore. The 12 major disease subclusters are highlighted.

654 Abbreviations: AIM: Autoimmunity; AIF: Autoinflammation; CID: Combined immunodeficiency; HLH:
655 Hemophagocytic lymphohistiocytosis; ainfl.dis: Autoinflammatory disease; SCID: Severe combined
656 immunodeficiency; Complement def.: Complement deficiency; CGD: Chronic granulomatous disease.

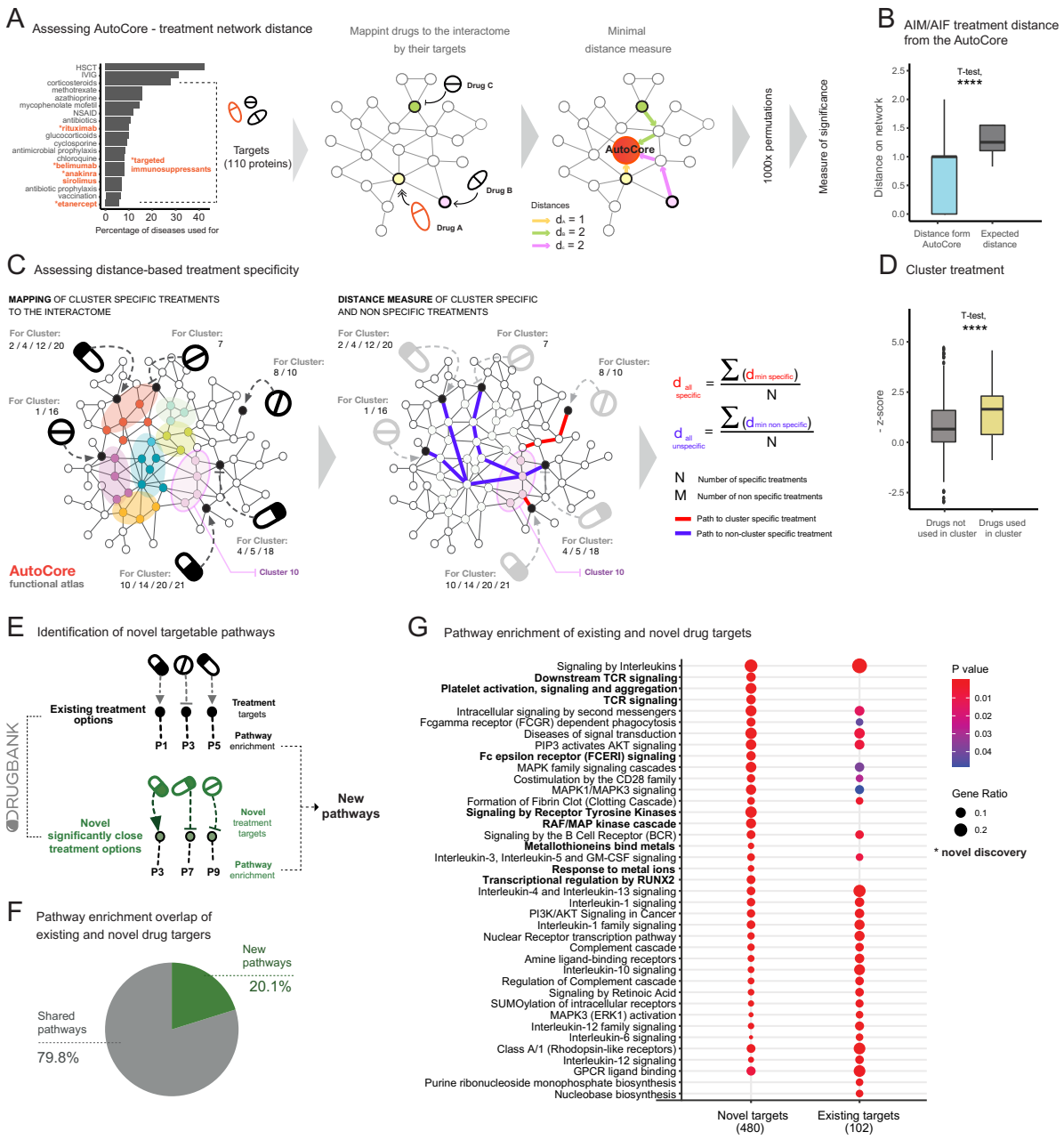


Figure 7: Exploiting the AutoCore for identification of potential treatment repurposing opportunities. A) Assessment of treatment-target distance from the AutoCore with minimal distance. B) Distance of drug targets of drugs that are used for treatment in the AutoCore. Significance of difference in distance measurements between the AutoCore and 1000 random permutations was determined by Student t-test. C) Approach to investigate cluster-specific treatment distances using the interactome and minimal distance. D) Distance of cluster-specific and cluster-non-specific drug targets on the interactome. Significance of difference in distance measurements between the cluster specific and non-specific treatments was determined by Student's t-test. E) Approach to investigate novel pathways pinpointed by significantly close rugs to the AutoCore. F) Fraction of existing and novel pathways in drug targets

667 significantly close to the AutoCore. G) Pathways enriched of novel treatment targets significantly close to
668 the AutoCore, compared to pathway enrichment of already existing treatment strategies. The top 20
669 enriched pathways for each group (novel and existing targets) are visualized. Novel pathways are
670 highlighted in bold. Odds ratio and p-value were determined by Fisher's exact test. * = $p<0.05$, ** = $p<0.01$,
671 *** = $p<0.001$, *** = $p<0.0001$, **** = $p<0.00001$.

References

1. Tobón, G. J., Youinou, P. & Saraux, A. The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. *J. Autoimmun.* **35**, 10–14 (2010).
2. McInnes, I. B. & Schett, G. Pathogenetic insights from the treatment of rheumatoid arthritis. *Lancet* **389**, 2328–2337 (2017).
3. Atkinson, M. A., Eisenbarth, G. S. & Michels, A. W. Type 1 diabetes. *Lancet* **383**, 69–82 (2014).
4. Katsarou, A. *et al.* Type 1 diabetes mellitus. *Nat Rev Dis Prim.* **3**, 17016 (2017).
5. Abraham, C. & Cho, J. H. Inflammatory bowel disease. *N. Engl. J. Med.* **361**, 2066–78 (2009).
6. Guan, Q. A Comprehensive Review and Update on the Pathogenesis of Inflammatory Bowel Disease. *J Immunol Res* **2019**, 7247238 (2019).
7. Rosenblum, M. D., Remedios, K. A. & Abbas, A. K. Mechanisms of human autoimmunity. *J. Clin. Invest.* **125**, 2228–2233 (2015).
8. Fischer, A. *et al.* Autoimmune and inflammatory manifestations occur frequently in patients with primary immunodeficiencies. *J. Allergy Clin. Immunol.* **140**, 1388-1393.e8 (2017).
9. Sobreira, N. L. M. *et al.* Matchmaker Exchange. *Curr. Protoc. Hum. Genet.* **95**, 9.31.1-9.31.15 (2017).
10. Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum. Mutat.* **36**, 928–930 (2015).
11. Haffner, M. E. Adopting Orphan Drugs — Two Dozen Years of Treating Rare Diseases. *N. Engl. J. Med.* **354**, 445–447 (2006).
12. Lucas, C. L. & Lenardo, M. J. Identifying genetic determinants of autoimmunity and immune dysregulation. *Curr. Opin. Immunol.* **37**, 28–33 (2015).
13. Loddo, I. & Romano, C. Inflammatory Bowel Disease: Genetics, Epigenetics, and Pathogenesis. *Front. Immunol.* **6**, 551 (2015).
14. Pulendran, B. & Davis, M. M. The science and medicine of human immunology. *Science* **369**, (2020).
15. López-Otín, C. & Kroemer, G. Hallmarks of Health. *Cell* **184**, 33–63 (2021).
16. Savic, S., Caseley, E. A. & McDermott, M. F. Moving towards a systems-based classification of innate immune-mediated diseases. *Nat. Rev. Rheumatol.* **16**, 222–237 (2020).
17. Arakelyan, A. *et al.* Autoimmunity and autoinflammation: A systems view on signaling pathway dysregulation profiles. *PLoS One* **12**, e0187572 (2017).
18. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).
19. Chan, S. Y. & Loscalzo, J. The emerging paradigm of network medicine in the study of human disease. *Circ. Res.* **111**, 359–374 (2012).
20. Hofree, M., Shen, J. P., Carter, H., Gross, A. & Ideker, T. Network-based stratification of tumor mutations. *Nat. Methods* **10**, 1108–15 (2013).
21. Hidalgo, C. A., Blumm, N., Barabási, A.-L. & Christakis, N. A. A Dynamic Network Approach for the Study of Human Phenotypes. *PLoS Computational Biology* **5**, e1000353 (2009).
22. Menche, J. *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science (80-.).* **347**, (2015).
23. Sharma, A. *et al.* A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum. Mol. Genet.* **24**, 3005–3020 (2015).
24. Peters, L. A. *et al.* A functional genomics predictive network model identifies regulators of inflammatory bowel disease. (2017). doi:10.1038/ng.3947
25. Li, Q. *et al.* Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. *Gastroenterology* **150**, 1196–1207 (2016).
26. Smedley, D. *et al.* Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat. Protoc.* **10**, 2004–2015 (2015).
27. Bousfiha, A. *et al.* Human Inborn Errors of Immunity: 2019 Update of the IUIS Phenotypical Classification. *J. Clin. Immunol.* **40**, 66–81 (2020).
28. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: leveraging knowledge across phenotype–gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2018).
29. Pavan, S. *et al.* Clinical Practice Guidelines for Rare Diseases: The Orphanet Database. *PLoS*

One **12**, e0170365 (2017).
30. Lewis, J., Snyder, M. & Hyatt-Knorr, H. Marking 15 years of the Genetic and Rare Diseases Information Center. *Transl Sci Rare Dis* **2**, 77–88 (2017).
31. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
32. Boisson, B., Quartier, P. & Casanova, J.-L. Immunological loss-of-function due to genetic gain-of-function in humans: autosomal dominance of the third kind. *Curr. Opin. Immunol.* **32**, 90–105 (2015).
33. Caldera, M., Buphamalai, P., Müller, F. & Menche, J. Interactome-based approaches to human disease. *Curr. Opin. Syst. Biol.* **3**, 88–94 (2017).
34. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **45**, D408–D414 (2016).
35. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).
36. Wachi, S., Yoneda, K. & Wu, R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21**, 4205–4208 (2005).
37. Ning, K., Ng, H. K., Srihari, S., Leong, H. W. & Nesvizhskii, A. I. Examination of the relationship between essential genes in {PPI} network and hub proteins in reverse nearest neighbor topology. *BMC Bioinformatics* **11**, 505 (2010).
38. Wu, C.-C., Kannan, K., Lin, S., Yen, L. & Milosavljevic, A. Identification of cancer fusion drivers using network fusion centrality. *Bioinformatics* **29**, 1174–1181 (2013).
39. Saha, A. *et al.* Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* **27**, 1843–1858 (2017).
40. Cowen, L., Ideker, T., Raphael, B. J. & Sharan, R. Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* **18**, 551–562 (2017).
41. Ghiassian, S. D., Menche, J. & Barabási, A.-L. A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome. *PLOS Comput. Biol.* **11**, e1004120 (2015).
42. Rolland, T. *et al.* A proteome-scale map of the human interactome network. *Cell* **159**, 1212–26 (2014).
43. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science (80-.).* **347**, 1260419 (2015).
44. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science (80-.).* **357**, (2017).
45. Hussein, H. M. & Rahal, E. A. The role of viral infections in the development of autoimmune diseases. *Crit. Rev. Microbiol.* **45**, 394–412 (2019).
46. Dixon-Zegeye, M. & Rutherford, A. Infections in systemic autoimmune diseases. *Handbook of Systemic Autoimmune Diseases* 143–166 (2020). doi:10.1016/b978-0-444-64217-2.00007-5
47. Mayor, P. C. *et al.* Cancer in primary immunodeficiency diseases: Cancer incidence in the United States Immune Deficiency Network Registry. *J. Allergy Clin. Immunol.* **141**, 1028–1035 (2018).
48. Hemminki, K., Huang, W., Sundquist, J., Sundquist, K. & Ji, J. Autoimmune diseases and hematological malignancies: Exploring the underlying mechanisms from epidemiological evidence. *Semin. Cancer Biol.* **64**, 114–121 (2020).
49. Ragusa, F. *et al.* Hashimotos' thyroiditis: Epidemiology, pathogenesis, clinic and therapy. *Best Practice & Research Clinical Endocrinology & Metabolism* **33**, 101367 (2019).
50. Docimo, G. *et al.* The Human Microbiota in Endocrinology: Implications for Pathophysiology, Treatment, and Prognosis in Thyroid Diseases. *Front. Endocrinol.* **11**, 586529 (2020).
51. Cooper, G. S., Bynum, M. L. K. & Somers, E. C. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J. Autoimmun.* **33**, 197–207 (2009).
52. Cho, J. H. & Gregersen, P. K. Genomics and the Multifactorial Nature of Human Autoimmune Disease. *N. Engl. J. Med.* **365**, 1612–1623 (2011).
53. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
54. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

55. Hayden, M. S., West, A. P. & Ghosh, S. NF-κB and the immune response. *Oncogene* **25**, 6758–6780 (2006).
56. Pazmandi, J., Kalinichenko, A., Ardy, R. C. & Boztug, K. Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms. *Immunol. Rev.* **287**, 162–185 (2019).
57. Alperin, J. M., Ortiz-Fernández, L. & Sawalha, A. H. Monogenic Lupus: A Developing Paradigm of Disease. *Front. Immunol.* **9**, 2496 (2018).
58. Tarr, T. *et al.* Similarities and differences between pediatric and adult patients with systemic lupus erythematosus. *Lupus* **24**, 796–803 (2015).
59. Bundhun, P. K., Kumari, A. & Huang, F. Differences in clinical features observed between childhood-onset versus adult-onset systemic lupus erythematosus: A systematic review and meta-analysis. *Med.* **96**, e8086 (2017).
60. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–24 (2012).
61. Peters, L. A. *et al.* A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat. Genet.* (2017). doi:10.1038/ng.3947
62. Li, T.-H., Liu, L., Hou, Y.-Y., Shen, S.-N. & Wang, T.-T. C-type lectin receptor-mediated immune recognition and response of the microbiota in the gut. *Gastroenterol. Rep.* **7**, 312–321 (2019).
63. Rahabi, M. *et al.* Divergent Roles for Macrophage C-type Lectin Receptors, Dectin-1 and Mannose Receptors, in the Intestinal Inflammatory Response. *Cell Reports* **30**, 4386-4398.e5 (2020).
64. Yang, Z. *et al.* C-type lectin receptor LSECtin-mediated apoptotic cell clearance by macrophages directs intestinal repair in experimental colitis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 11054–11059 (2018).
65. Haynes, W. A. *et al.* Integrated, multicohort analysis reveals unified signature of systemic lupus erythematosus. *JCI Insight* **5**, (2020).
66. Vasant, D. *et al.* ORDO: An Ontology Connecting Rare Disease, Epidemiology and Genetic Data. *Phenotype data ISMB2014* (2014).
67. Köhler, S. *et al.* The Human Phenotype Ontology in 2017. *Nucleic Acids Res.* **45**, D865–D876 (2017).
68. Guney, E., Menche, J., Vidal, M. & Barábasi, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 10331 (2016).
69. Caldera, M. *et al.* Mapping the perturbome network of cellular perturbations. *Nat. Commun.* **10**, 5140 (2019).
70. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668-72 (2006).
71. Herrington, F. D., Carmody, R. J. & Goodyear, C. S. Modulation of NF-κB Signaling as a Therapeutic Target in Autoimmunity. *J. Biomol. Screen.* **21**, 223–42 (2016).
72. Cutolo, M., Sulli, A., Paolino, S. & Pizzorni, C. CTLA-4 blockade in the treatment of rheumatoid arthritis: an update. *Expert Rev. Clin. Immunol.* **12**, 417–25 (2016).
73. Kiykim, A. *et al.* Abatacept as a Long-Term Targeted Therapy for LRBA Deficiency. *J. allergy Clin. Immunol. Pract.* **7**, 2790-2800.e15 (2019).
74. Serwas, N. K. *et al.* Human DEF6 deficiency underlies an immunodeficiency syndrome with systemic autoimmunity and aberrant CTLA-4 homeostasis. *Nat. Commun.* **10**, 3106 (2019).
75. Chamoto, K., Al-Habsi, M. & Honjo, T. Role of PD-1 in Immunity and Diseases. *Curr. Top. Microbiol. Immunol.* **410**, 75–97 (2017).
76. Yang, J. *et al.* First-in-human study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of ALPN-101, a dual CD28/ICOS antagonist, in healthy adult subjects. *Clin. Transl. Sci.* (2021). doi:10.1111/cts.12983
77. Adom, D. *et al.* ICOSL+ plasmacytoid dendritic cells as inducer of graft-versus-host disease, responsive to a dual ICOS/CD28 antagonist. *Sci. Transl. Med.* **12**, (2020).
78. Karnell, J. L., Rieder, S. A., Ettinger, R. & Kolbeck, R. Targeting the CD40-CD40L pathway in autoimmune diseases: Humoral immunity and beyond. *Adv. Drug Deliv. Rev.* **141**, 92–103 (2019).
79. Totsuka, T. *et al.* Therapeutic effect of anti-OX40L and anti-TNF-alpha MAbs in a murine model of chronic colitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **284**, G595-603 (2003).
80. Pakala, S. V, Bansal-Pakala, P., Halteman, B. S. & Croft, M. Prevention of diabetes in NOD mice at a late stage by targeting OX40/OX40 ligand interactions. *Eur. J. Immunol.* **34**, 3039–46 (2004).

81. Chittasupho, C., Siahaan, T. J., Vines, C. M. & Berkland, C. Autoimmune therapies targeting costimulation and emerging trends in multivalent therapeutics. *Ther. Deliv.* **2**, 873–89 (2011).
82. Liu, X., Gorzelanny, C. & Schneider, S. W. Platelets in Skin Autoimmune Diseases. *Front. Immunol.* **10**, 1453 (2019).
83. Ben Mkaddem, S., Benhamou, M. & Monteiro, R. C. Understanding Fc Receptor Involvement in Inflammatory Diseases: From Mechanisms to New Therapeutic Tools. *Front. Immunol.* **10**, 811 (2019).
84. Ben Mkaddem, S. *et al.* Shifting FcγRIIA-ITAM from activation to inhibitory configuration ameliorates arthritis. *J. Clin. Invest.* **124**, 3945–59 (2014).
85. McGonagle, D. & McDermott, M. F. A proposed classification of the immunological diseases. *PLoS Med.* **3**, e297 (2006).
86. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
87. Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034.e6 (2019).
88. Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* **173**, 1573–1580 (2018).
89. Grateau, G. *et al.* How should we approach classification of autoinflammatory diseases? *Nat. Rev. Rheumatol.* **9**, 624–629 (2013).
90. Pathak, S., McDermott, M. F. & Savic, S. Autoinflammatory diseases: update on classification diagnosis and management. *J. Clin. Pathol.* **70**, 1–8 (2017).
91. Guney, E., Menche, J., Vidal, M. & Barábasi, A.-L. Network-based in silico drug efficacy screening. *Nat. Commun.* **7**, 10331 (2016).
92. Köstel Bal, S., Pazmandi, J., Boztug, K. & Özen, S. Rheumatological manifestations in inborn errors of immunity. *Pediatr. Res.* **87**, 293–299 (2020).
93. Kuleshov, M. V *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90--7 (2016).
94. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–9 (2016).
95. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
96. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39**, D691-7 (2011).
97. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
98. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
99. Consortium, T. G. & The GTEx Consortium. The {GTEx} Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
100. A Hagberg, D Schult, P. S. Exploring Network Structure, Dynamics, and Function using NetworkX. *Proc. 7th Python Sci. Conf. (SciPy 2008)* **pp. 11-15**, (2008).
101. Bonacich, P. Power and Centrality: A Family of Measures. *Am. J. Sociol.* **92**, 1170–1182 (1987).
102. Köhler, S., Bauer, S., Horn, D. & Robinson, P. N. Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.* **82**, 949–58 (2008).
103. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–504 (2003).
104. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
105. Pirch Sebastian; Mueller, Felix; Iofinova, Eugenia; Pazmandi, Julia; Huetter, Christiane; Chiettini Martin; Sin, Celine; Boztug, Kaan; Podkosova, Iana; Kaufmann, Hannes; Menche, J. The VRNetzer Platform Enables Interactive Network Analysis in Virtual Reality. *Nat. Commun.* **in press**, (2021).

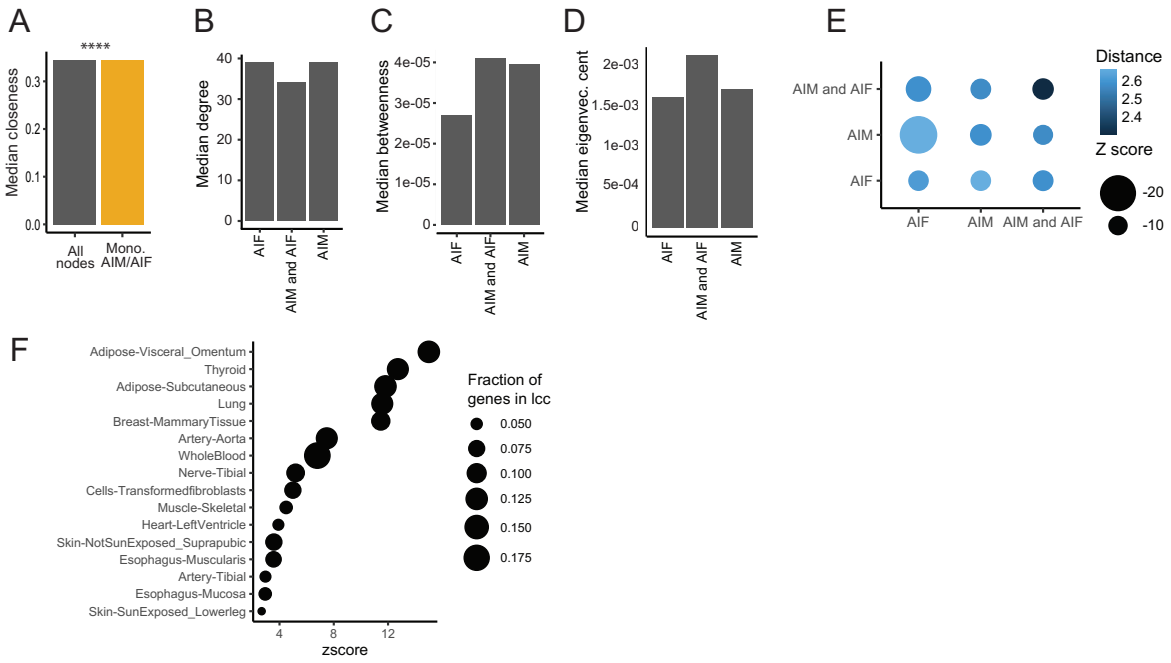


Figure S1: A) Median closeness centrality of autoimmune/ autoinflammatory gene defects. The Mann-Whitney U nonparametric test was used to determine p-value. B-D) Median degree, betweenness centrality and eigenvector centrality of monogenic autoimmune and autoinflammatory genes on the interactome. E) Average distances between autoimmune and autoinflammatory gene defects on the interactome. Significance was determined by z-score based on 1000 random permutations. F) Largest connected components of monogenic autoimmune and autoinflammatory genes on different tissue specific coexpression networks form the GTEx resource. Significance was determined by z-score based on 1000 random permutations. Abbreviations: AIM: Autoimmune; AIF: Autoinflammatory; Lcc: Largest connected component. Eigenvec. Cent: Eigenvector centrality.

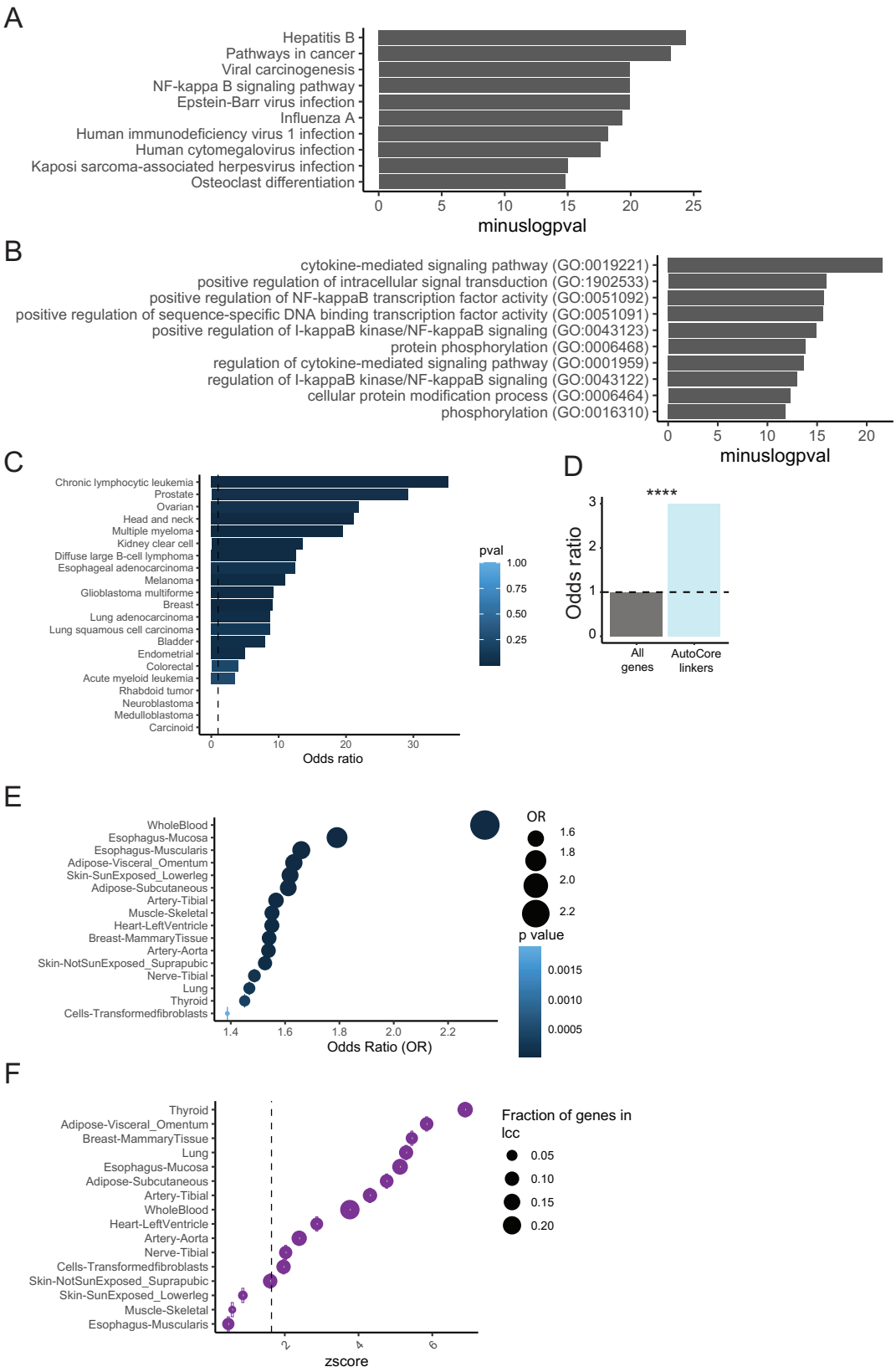


Figure S2: A) KEGG pathway enrichment of AutoCore linker nodes. Enrichment analysis was carried out using the Enrichr tool. B) Gene Ontology Biological Process enrichment of AutoCore linker nodes. Enrichment analysis was carried out using the Enrichr tool. C) Enrichment of AutoCore linker nodes in

frequently mutated genes in different cancers. Odds ratio and p-value were determined by Fisher's exact test. D) Enrichment of AutoCore linker nodes in highly expressed genes in lymphoma. Odds ratio and p-value were determined by Fisher's exact test. E) Enrichment of AutoCore nodes in tissue-specific coexpression networks from the GTEx resource. Odds ratio and p-value were determined by Fisher's exact test. F) Largest connected components of AutoCore nodes on different tissue specific coexpression networks form the GTEx resource. Significance was determined by z-score based on 1,000 random permutations.

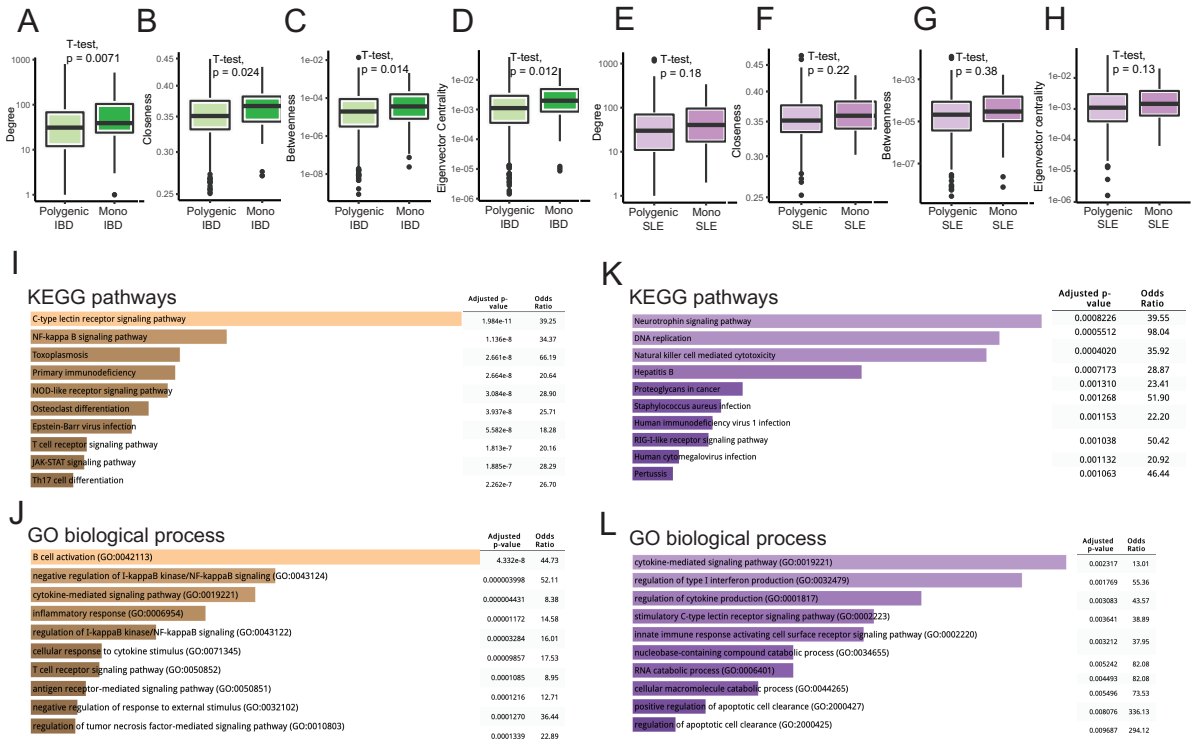


Figure S3: A-D) Degree, closeness centrality, betweenness centrality, and eigenvector centrality of monogenic and polygenic IBD genes on the interactome. Student T-test was used to determine the significance of difference between two groups. E-H) Degree, closeness centrality, betweenness centrality, and eigenvector centrality of monogenic and polygenic SLE genes on the interactome. Student T-test was used to determine the significance of difference between two groups. I-J) KEGG and Gene Ontology biological process enrichment of IBD connectors. Enrichment analysis was carried out using the Enrichr tool. I-J) KEGG and Gene Ontology biological process enrichment of SLE connectors. Enrichment analysis was carried out using the Enrichr tool.

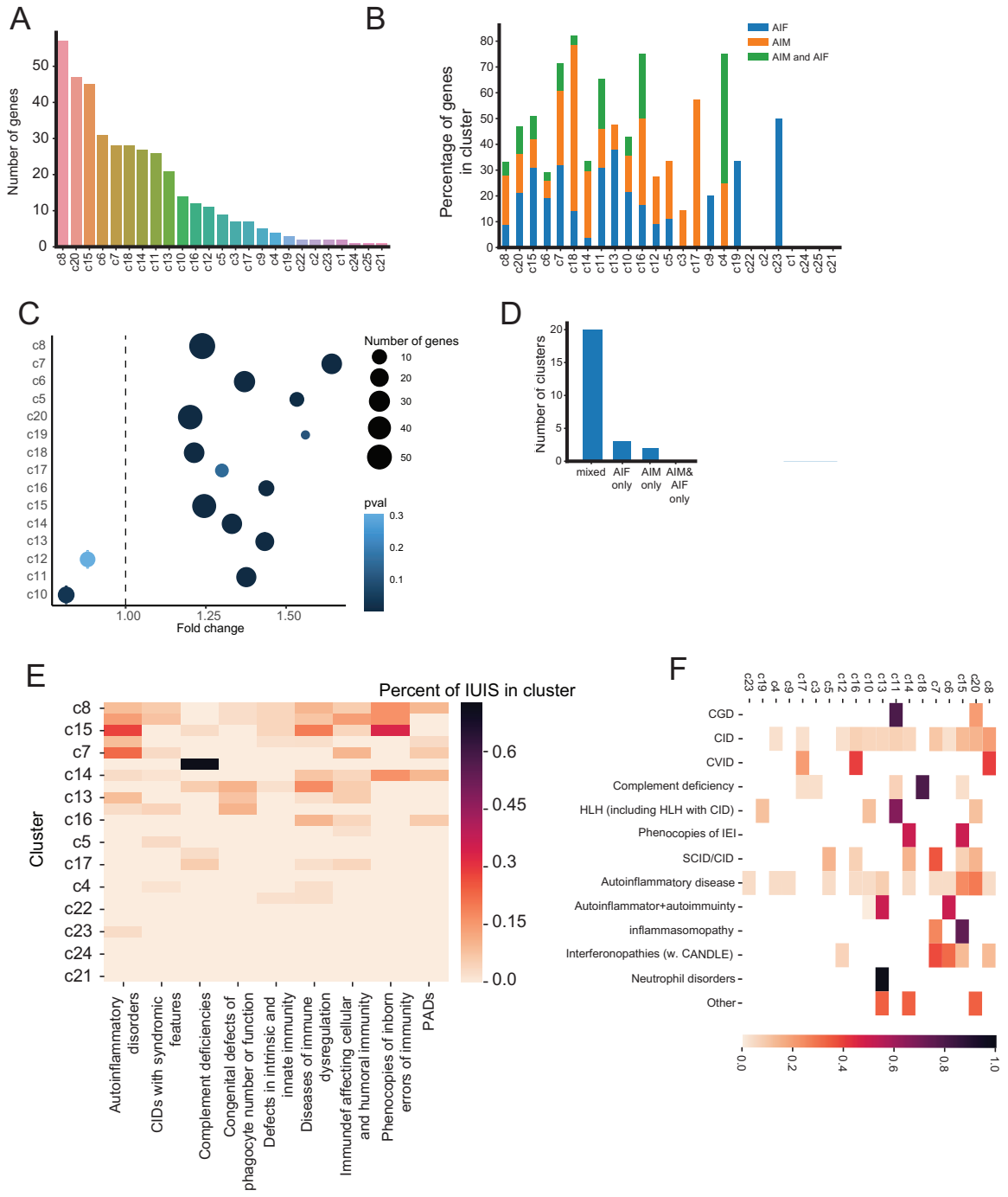
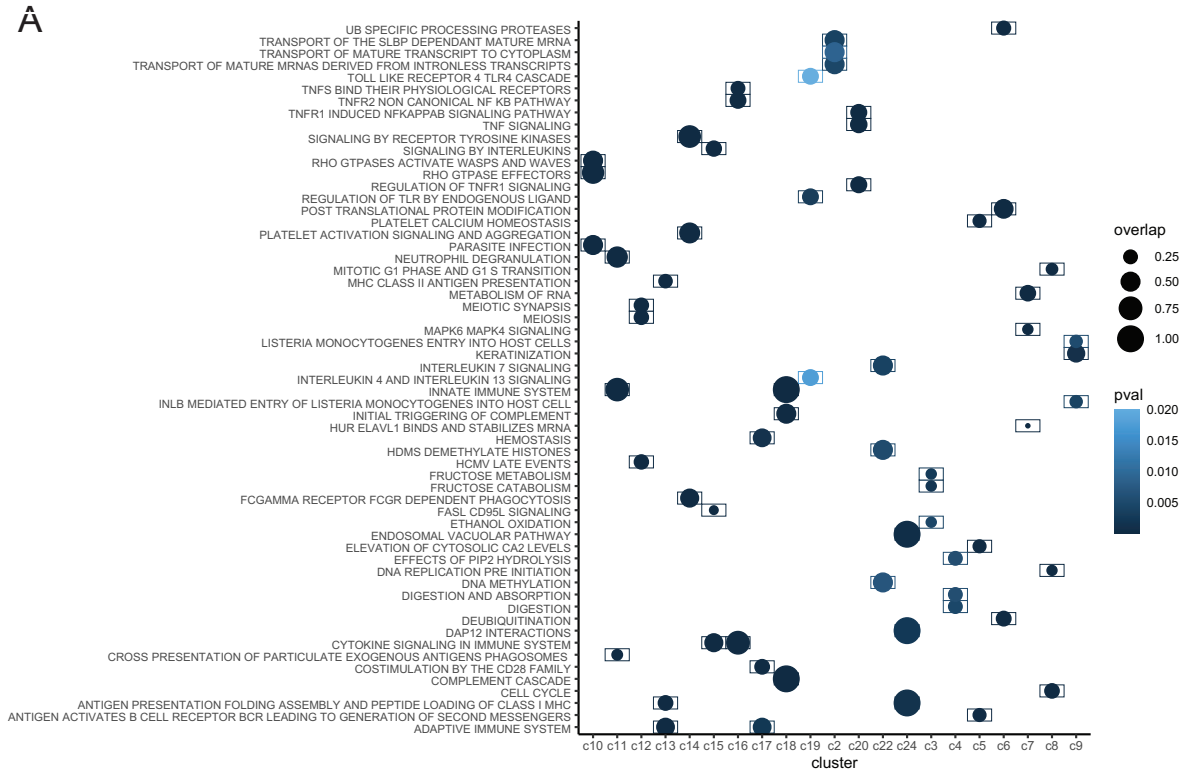


Figure S4: A) Different molecular clusters within the AutoCore. Clusters are ordered according to size. B) Fraction of AIM and AIF gene defects per molecular cluster. C) HPO phenotype similarity of clusters. Fisher's exact test was used to obtain p-value. D) Fraction of autoimmune, autoinflammatory and mixed clusters. E) Distribution of IUIS subgroups within the molecular clusters. F) Distribution of clinical phenotypes within the molecular clusters. Abbreviations: AIM: Autoimmunity; AIF: Autoinflammation; IUIS: International Union of Immunological Societies; CID: Combined immunodeficiency; CVID: Common variable immunodeficiency; SCID: Severe combined immunodeficiency; HLH: Hemophagocytic lymphohistiocytosis; ainfl.dis: Autoinflammatory disease; Complement def.: Complement deficiency; CGD: Chronic granulomatous disease.; PAD: Primary antibody deficiency.



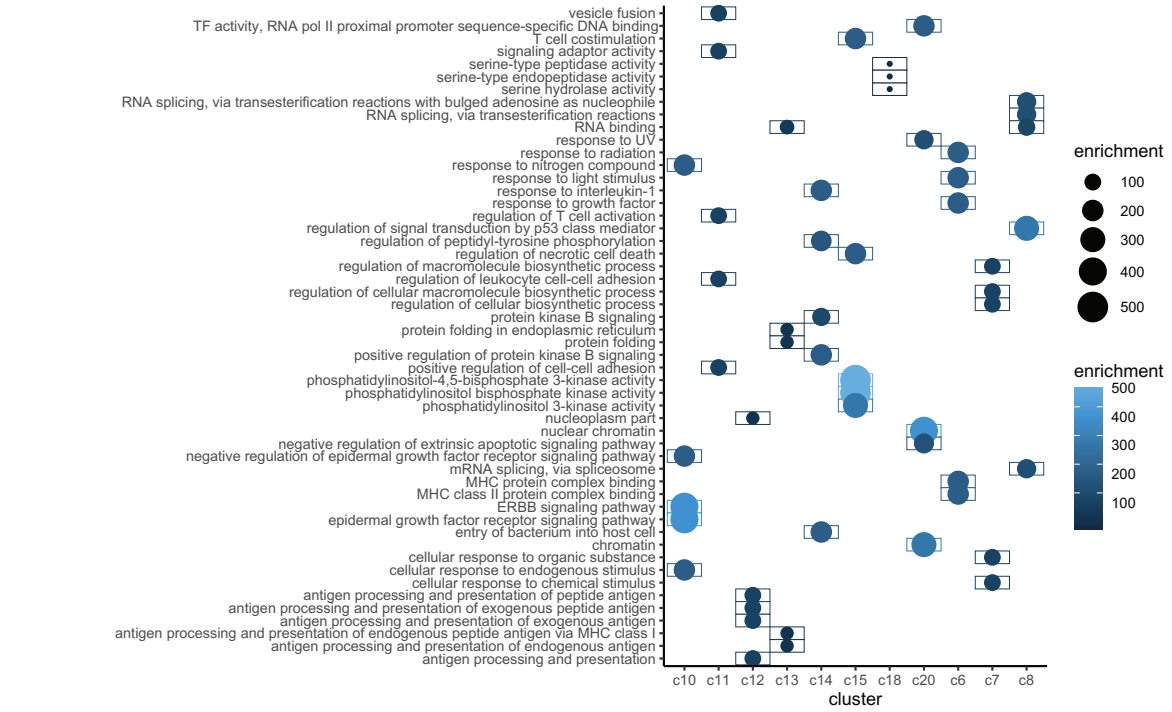


Figure S5: A) Top three enriched reactome pathway terms per molecular cluster. Odds ratio and p-value were determined by Fisher's exact test. B) Top five enriched Gene Ontology terms (biological process, molecular function and cellular component combined) per molecular cluster.

4 Discussion

Over the last decade, numerous case studies of rare disease patients have shown how severe IEI can advance our understanding of core genes and pathways of the immune system. To illustrate the diverse gene defects and molecular mechanisms affected by rare diseases of the immune system, manuscript one in the introduction showcases the diverse genetic landscape of IBD, a prototypic autoinflammatory/autoimmune disorder. IBD presents with both a rare, severe early-onset and a more common, less severe, late-onset form. Individual gene defects underlying early-onset IBD and their consequences are discussed in detail, shedding light on how genetically, phenotypically and molecularly diverse this group of monogenic diseases is.

Notwithstanding the impressive success of the single-gene mechanistic studies, we are now also increasingly recognizing the conceptual and practical limitations of the traditional one-gene, one-disease paradigm in rare disease research. On the one hand, the decentralized fashion in which most of these disease genes have traditionally been investigated has meant that data regarding these diseases is scattered throughout various databases and patient cohorts. This has resulted in a considerable diagnostic delay with most patients with rare immune mediated diseases remaining without genetic diagnosis (Chazal et al., n.d.; Gahl et al. 2016). Indeed, one of the main challenges of finding a genetic diagnosis of these diseases is often finding a matching patient cohort/diagnosis (N. Sobreira et al. 2015; N. L. M. Sobreira et al. 2017). As sharing of data and the subsequent fast and accurate diagnosis heavily relies on standardized data and vocabularies, the unavailability of accurate phenotype data for IEI significantly contributes to the diagnostic delay. Over the years, various phenotype-nomenclatures emerged that aimed to address specific applications of phenotype data, but to date none of these found strong footing for rare diseases. HPO was specifically developed as a phenotype ontology with diagnostic application in mind, exemplified by the different tools that use them for finding genetic diagnosis such as Exomiser and Lirical (Smedley et al. 2015; Robinson et al. 2020). Unfortunately, HPO is currently too incomplete to serve IEI in an adequate manner. The second manuscript of this thesis provides a proof of principle of how to use ontology-guided machine learning coupled with an expert-based review to systematically evaluate the HPO and reannotate IEI with HPO phenotype terms. We show that a directed review of distinct branches of the HPO tree and reannotation of specific IEI subgroups is able to achieve a significant increase in HPO-phenotype-based similarity matching

to genetic diagnosis in a patient cohort, as well as a more clinically accurate phenotype-based similarity between subgroups of IEI.

Beyond providing accurate phenotyping data to enable faster diagnosis and disease-matching, an objective overview of rare diseases of the immune system is currently lacking. To date, the focus on single gene defects makes it difficult to appreciate the intricate molecular network through which the individual components of the immune system are orchestrated. Indeed, despite over 400 monogenic immune diseases that have been defined to date, a unifying view of this important class of diseases is still lacking. The overarching ambition of the project showcased in research manuscript number two was to develop such a systems-level view and showcase its utility for addressing a wide range of important biomedical questions. To achieve this, we developed a network-based framework for integrating all currently known monogenic immune defects underlying autoimmunity and autoinflammation and their molecular interactions. We show that this framework, and the map of human autoimmunity and autoinflammation termed the AutoCore, are powerful tools to quantify previously only anecdotal clinical observations, to identify objective molecular subgroups of these diseases and to identify potentially druggable, novel therapeutic pathways.

In conclusion, this thesis here provides two examples of the application of systems biology to expand the knowledge base regarding rare diseases of the immune system, and to construct a systems-based framework for the investigation of a subgroup of IEI. Both of these endeavors have resulted in a significant gain of functionally relevant information regarding IEI both on a single-disease basis, but also by looking at IEI as a disease group.

4.1. Discussion and outlook for “Curation and Expansion of Human Phenotype Ontology for Defined Groups of Inborn Errors of Immunity”

IEI pose particular challenges for affected patients, as well as for medical doctors and researchers working to improve diagnostic and therapeutic approaches. Clinicians often only see a few patients with a particular rare phenotype throughout their careers, leading to considerable diagnostic delay. Data sharing across institutions and borders is crucial to achieve early diagnosis. However, phenotypic data are often described with variable quality and specificity, which hampers patient matching, genetic variant prioritization in diagnostic pipelines and global data exchange. Although well-known disease databases such as OMIM (McKusick 2007), OrphaNet (Weinreich et al. 2008) and GARD (Zhu et al. 2020) store data regarding the phenotypic traits of numerous rare diseases, the phenotypic representation of IEI in these resources is not complete. In the past decade, HPO emerged as an ontology to objectively describe human phenotypes with a diagnostic application in mind (Smedley et al. 2015; Robinson et al. 2020). Although incomplete, HPO has been the ontology most frequently adapted by the IEI community (Köhler, Kindle, and Robinson 2021). HPO-based phenotypic annotations of IEI exist, however, they are not complete enough to represent most diseases within this disease group in an adequate manner. In the first version of HPO, automated text mining of the OMIM-clinical synopsis corpus was used to identify the phenotypic spectrum and annotate diseases (Robinson et al. 2008) in order to get a general annotation of IEI. Over the last years, it was established that this initial annotation only captured a subset of relevant phenotypes of IEI, and it has become clear that a tailored, expert-lead approach is necessary to expand the current knowledgebase. Although a similar, expert-linked effort was recently documented for eye diseases (Sergouniotis et al. 2019), our publication is the first comprehensive method published so far with validation of the improvement of annotation as a result of our reannotation effort in both a patient-focused and disease-specific manner. With our initiative, our goal was to revise and expand the current HPO tree, as well as to reannotate IEI with HPO terms.

As proof of concept, we have started from four subgroups of IEI based on the IUIS clinical classification: diseases affecting cellular and humoral immunity (IUIS Table I), diseases of immune dysregulation (IUIS Table IV), autoinflammatory disorders (IUIS Table VII) and genetically undiagnosed predominantly antibody deficiencies (IUIS Table III). We have used an expert-lead, hands-on approach to review and expand the HPO tree. Specific, IEI relevant

branches of the HPO tree were discussed in detail at in-person workshops by expert groups. Based on these expert discussions, suggestions to change the tree structure and to add new HPO terms were submitted as a request to HPO, using the GitHub tracking system (Köhler et al. 2017). To date, we have requested 206 changes in the ontology structure, including 137 new term requests. Currently, 81 of the 137 requested new terms (59.12%) have been accepted to be part of the official vocabulary of HPO. Updating the HPO ontology with the requested terms is still ongoing. The HPO ontology is constantly receiving updates, based on input from initiatives like ours and also individual change requests. Therefore, the HPO ontology structure, synonyms and term descriptions are subject to change. Furthermore, the inclusion, change or exclusion of putative HPO terms is based on a discussion between and with the curators of the HPO website, using the GitHub tracker. As a result, as we are not the only decision makers in the structural change and term inclusion process, and it is possible that some of our suggestions are either added in a modified way, added in a different branch in the HPO, or after longer discussions possibly rejected.

Measuring the impact of the above mentioned individual tree structure changes is not a straightforward task. The most logical way perhaps is to perform a joint validation of all changes requested, taking all the changes relevant for the IEI disease group in consideration. To date we have reviewed the HPO tree from the perspective of the four IUIS specified subgroups of IEI we have reannotated, therefore our review of the whole HPO body is not yet complete. Once the whole HPO tree, that is relevant in the context of IEI has been revised, the assessment of the full impact of the modifications will be possible. To do this, the efficacy of the revised/modified trees should be tested on the improvement of patient-diseases and disease-disease similarity measures, without changing the underlying HPO term annotation of individual diseases.

Along with revising the structure of the HPO tree relevant for IEI, reannotation of IEI with HPO terms was our main goal. To achieve this, we have used an ontology-guided effort relying on natural language processing (NLP), using a tool trained on the HPO ontology (Arbabi et al. 2019), and used on Pubmed abstracts. The tool, termed neural concept recognizer (NCR) performed well in robust performance measurements for PubMed article abstracts. We have run the NCR tool on a collection of publications that have accurately described the phenotypic presentation of the diseases in the four subgroups, collected by experts on the different diseases. We have shown that the NCR was a powerful approach for the identification of phenotype terms from our expert-curated publication corpus, as for the majority of the new terms per disease after reannotation

now stem from the text-mining over existing HPO terms or additional suggestions by experts (Figure 3I). This text mining effort has resulted in a major quantitative gain in the number of available phenotype terms per disease, as well as a significant gain in qualitative information as the mean information content accessible per disease increased as well.

We have tested the efficacy of our reannotation approach using a cohort of IEI and performed HPO-phenotype-based patient-disease matching. For this analysis, we have used data from a real cohort of 30 IEI patients. Together, the 30 patients harbored 24 diverse genetic diagnoses. In order to validate the efficacy of the annotated disease set to improve diagnostic analysis, we aimed to use a well annotated patient cohort harboring confirmed gene defects in our reannotated diseases. With this, our goal was to illustrate the real-life applicability of an HPO-based patient-disease matching endeavor, with its difficulties, including the quality and quantity of disease annotations available, as well as the low number of patients in a specific cohort. As agreed throughout Europe according to the European Commission, a rare disease is defined as diseases affecting fewer than 1 in 2000 people (Nugent and Rhinard 2015; Eurodis). Thus, a frequent hallmark of rare diseases is the small available patient pool, especially when focusing on a diverse group of disorders. Current larger cohort studies echo this sentiment. In a recent effort, (Thaventhiran et al. 2020) sequenced 1,318 primary immunodeficiency (PID) patients and found genetic diagnosis for 135 patients spanning 42 individual diseases. In most instances, only one patient was found with a particular disease. In another effort by (Simon et al. 2020) whole exome sequencing was applied to a cohort of 106 highly consanguineous PID patients, and found a likely genetic diagnosis in 70% of cases spanning 46 gene defects. We have shown that in this real-life cohort, the reannotation has achieved a significant improvement in patient-to-diagnosis matching based on HPO phenotypic similarity, as well as in the phenotype-based ranking of the correct genetic diagnosis. These proof-of-principle results underpin the success of our effort providing a more accurate phenotypic description of IEI in HPO that is already improving diagnostic accuracy.

We have now addressed four specific subgroups of IEI. Our ongoing goals and work entail recruiting more experts in order to take our initiative further and address all other IEI disease groups. Once a fully reannotated set of IEI is available we can make a final assessment of the improvement of phenotype-disease matching and improvement in diagnosis, although we already show promising results with our focused review of four IUIS subgroups of IEI. A fully reannotated set of diseases will most likely spur on an expansion of diagnostic tools relying on HPOs of phenotyping such as, as well as popularize the tools available already such as Phenotips (Girdea

et al. 2013) Exomiser (Smedley et al. 2015) and Lirical (Robinson et al. 2020). In addition, HPO has been successfully applied to predict disease-associated lncRNAs (Le and Dao 2018) and disease-related phenotypes (Xue, Peng, and Shang 2019). Furthermore, fine-tuning and expanding the current annotation corpus with modifiers such as indicators of frequency of a phenotype in a specific disease can improve the accuracy of HPO-based phenotype matching (Köhler 2018). It is also likely that more machine-learning based technologies will be developed that use HPO for phenotyping. These tools will accelerate the rate of genetic diagnosis as non-expert centers could perform accurate patient-phenotype matching instead of having to rely on the expertise of the few rare disease centers.

A crucial step towards globally standardized phenotyping is the universal adaptation of HPO in the community. EHRs and hospitals as well as patient cohorts of individual hospitals and research centers are beginning to adapt HPOs so that there is a standardized nomenclature world-wide (Gasteiger et al. 2020). Adaptation of HPO in the IEI community can be facilitated by tools such as Doc2Hpo, a web application developed for accurate HPO curation from clinical data (C. Liu et al. 2019), or the (HPO and ORDO Ontological Module) HOOM that associates clinical entities with HPO terms (“HPO - ORDO Ontological Module - Association - Classes” n.d.). A complete annotation of IEI and IEI patients with HPO will also enable the seamless use of information and patient-exchange portals. Traditionally, IEI research is a slow-paced field in terms of the adoption of new technologies. A lot of individual research centers still rely on paper-based data storage for their patient cohorts. The implementation of electronic cohort-based information storage will naturally improve the adoption of HPO as a way of patient phenotyping for these research centers and valuable patient cohorts.

In conclusion, our manuscript demonstrates the utility of HPO and the efficacy of our effort to significantly improve the current representation of IEI in HPO which lead to significant improvement in the diagnostic accuracy of HPO for IEI. Our approach paves the way to further reannotate IEI, and drives the adoption of HPO in genetic diagnosis. Finally, as our repertoire of available accurate phenotypic and molecular data regarding diseases and individual patients expands, personalized medicine for IEI can potentially become a reality. New methods such as ours presented here will be crucial to allow us to go from a one-size-fits-all to a more personalized/precision medicine for IEI (Delhalle et al. 2018; Delmonte et al. 2019).

4.2. Discussion and outlook on “AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation.”

The paradigm that molecules do not act in isolation but elicit their effects through a connected cascade of interactions has led to the recognition that systems approaches are vital to enhance our understanding of diseases, and of health in general (López-Otín and Kroemer 2021). Since the discovery of disease modules on molecular networks and the predictive power of network-based closeness for pathobiological similarities (Menche et al. 2015), systems-methods and network medicine have been of a particular interest in research of immune mediated diseases. Indeed, the interplay of different immune cell types has been translated into a network-based view (Rieckmann et al. 2017), a systems-based classification of diseases of innate immunity has been recently proposed (Savic, Caseley, and McDermott 2020), and a holistic, pathway centered view of common, complex autoimmune and autoinflammatory phenotypes was introduced (Arakelyan et al. 2017).

Although systems methods have been applied to different complex diseases so far, they have rarely been applied to rare diseases. Therefore, despite the advances in our understanding of single-gene defects that perturb the immune system, a systematic overview of the implicated genes, molecular mechanisms and pathways is still lacking. We here aimed to provide such an overview, by leveraging network-based methods and constructing a systematic map of autoimmunity and autoinflammation starting from rare disease genes that present with autoimmunity and autoinflammation. Rare autoimmune and autoinflammatory diseases represent a high-confidence set of genetic perturbations that are historically poorly documented in open-source databases. Illustrating this, there is only phenotype and therapy-linked information available on 38.8% of these diseases in the official rare diseases data system, GARD (Lewis, Snyder, and Hyatt-Knorr 2017). We have therefore started by compiling a comprehensive dataset regarding the genetic, phenotypic, clinical presentations of rare autoimmune/autoinflammatory diseases, and the available therapeutic options. This dataset is available as a supplementary file linked to our manuscript, but also online at https://menchelab.com/autocoreapp/. This valuable data set harbors the potential to be translated into a knowledge-graph which could be used to reveal further unexpected connections between rare autoimmune/autoinflammatory diseases. Indeed, a knowledge graph was recently constructed using data from the GARD resource (Zhu et al. 2020) and has shown promise in accelerating the understanding of rare disease by the generation of disease mappings and associations through pathogenesis.

To construct a state-of-the art map of cellular interactions, we have started by building a novel interactome by combining various data sources. These included the interactome used to define disease modules of complex phenotypes (Menche et al. 2015), a database storing high-throughput protein-interaction data (Alanis-Lobato, Andrade-Navarro, and Schaefer 2016), and a recently published interactome already used to investigate Mendelian diseases (Luck et al. 2020). The combined interactome, although bigger and denser, echoed the properties of previously published interactomes in terms of the degree distribution and modularity of functions and pathways. As more and more molecular interactions are mapped out, the combined interactome can be expanded further to include newly identified connections. We hypothesize that as we have seen from the combination of the three networks mentioned above, the combined interactome would retain its properties even through multiple rounds of dynamic expansion due to the inherent scale-free and small-world feature of interactomes (Cohen and Havlin 2003; Friedel and Zimmer 2006; Rolland et al. 2014).

We show that the combined interactome can be used to quantify clinical and biologically relevant phenomena. These phenomena include the significant inter-linkedness of autoimmunity and autoinflammation on a molecular level. Our quantification of this non-separation of autoimmunity and autoinflammation on the interactome stresses the challenges of development of a purely clinical classification of rare autoimmune/autoinflammatory diseases due to the observed interconnectedness, previously termed the “immunological disease continuum” (McGonagle and McDermott 2006). Furthermore, we show the tendency of those gene defects that present with both autoimmune and autoinflammatory features to have higher centrality and a more diverse pathway repertoire in their direct network neighborhoods, as compared to gene defects that present with either one of the phenotypes. This confirmation of the increased diversity of affected processes can be leveraged for the development of methods that exploit this trait of gene defects to predict if a certain gene defect will present with autoimmune or autoinflammatory phenotypes. Indeed, network-based phenotype prediction methods are already in place, for example to identify genes linked to specific cellular processes such as cell motility (Bern et al. 2019), or to predict clinical outcomes across gene expression networks (Kang et al. 2017).

Using the combined interactome and rare disease genes, we identified the AutoCore using network propagation (Cowen et al. 2017). Within the AutoCore, 71.5% of disease genes are linked through either direct connections to one-another. Those disease genes (28.5% of all disease

genes) that do not harbor direct connections to other disease genes connected to the main cluster through the 213 linker genes which were identified through random walk with restart. The number of linker genes connecting the AutoCore is smaller than expected, and the connectivity of individual disease genes within the AutoCore is more significant than previously observed for complex diseases (Menche et al. 2015). As diseases within the AutoCore showed marked direct connectivity, we hypothesize that the 213 linker genes identified specifically for rare autoimmune and autoinflammatory disease might harbor novel rare disease-gene candidates. We have shown that the linker genes are enriched in genes identified to be associated with common complex autoimmune diseases. Just recently, PTPN2, a gene previously linked to IBD through GWAS (Rivas et al. 2011) was identified as a gene defect underlying early-onset autoimmune colitis (Parlato et al. 2020). As evidenced by tailored approaches such as those showcased for IBD (Q. Li et al. 2016) and corticospinal motor neuron disease (Novarino et al. 2014), rare diseases require specific tools for the identification of novel disease gene candidates. Therefore, a disease-specific analysis and disease-focused prioritization could reveal novel rare-disease gene candidates among the linkers genes.

In addition to their disease-gene potential, the 213 linker genes identified show ample evidence of clinical relevance by enrichment in abrogated T-cell immunity, viral infections, and hematological malignancies. All of these processes constitute clinical hallmarks of autoimmune and autoinflammatory diseases (Hussein and Rahal 2019; Dixon-Zegeye and Rutherford 2020; Mayor et al. 2018; Hemminki et al. 2020). Because of their high clinical relevance, the exploration of individual linker genes could aid in the identification of potentially novel disease mediators in autoimmunity and autoinflammation. Indeed, network-based methods have been used to identify specific key drivers in diseases such as showcased in type 2 diabetes (Sharma et al. 2018), Asthma (Sharma et al. 2015), hypertrophic cardiomyopathy (Maron et al. 2021), and late-onset IBD (Peters et al. 2017). The advantage of these network-based approaches is that the network neighborhood of candidates provides functional context of the linked cellular processes. This functional context can then be used to generate hypotheses on the molecular mechanisms affected by perturbation of the key drivers.

Polygenic autoimmune and autoinflammatory diseases, in contrast to rare disorders, are generally considered results of an interplay of various disease-predisposing factors, such as environmental exposures, genetic and epigenetic factors (Cho and Gregersen 2011). The genetic architecture of these diseases, although extensively studied, remains poorly understood, leaving most of the

heritability unexplained (Manolio et al. 2009; MacArthur et al. 2017). We showed that the AutoCore, built on rare autoimmune/autoinflammatory disease genes, is at the topological center of complex autoimmune and autoinflammatory diseases on the interactome. The omnigenic disease model (Boyle, Li, and Pritchard 2017; Wray et al. 2018) postulates that there is a relatively small number of core genes relevant to any disease, while disease modifiers on the periphery of these core genes elicit smaller effects that, taken together, are strong enough to affect cellular phenotype. We find that from a network perspective, rare autoimmune and autoinflammatory disease genes serve as a core of autoimmune/autoinflammatory processes, while genetic associations of complex polygenic autoimmune/autoinflammatory genes reside on the periphery. This finding is consistent with the observation that genetic lesions in these core genes are associated with larger perturbations in the interactome neighborhood, while variation and mutations in the complex polygenic disease genes means less detrimental effects.

We showed that the functional enrichment of specific connector nodes between monogenic and polygenic diseases reflects triggers of autoimmune and autoinflammatory phenotypes such as viral infections, response to viral infections and T cell-linked immune responses (Hayden, West, and Ghosh 2006). We found further enrichment of clinically and functionally relevant pathways and mechanisms on a disease-specific level in IBD and SLE, highlighting the AutoCore and the specific connectors nodes as potential disease-mediators and hallmark phenotype-genes for common polygenic diseases. As it has been shown that diseases that are in the close vicinity to one another on the interactome tend to present together (Menche et al. 2015), it will be intriguing to explore how the AutoCore connects to different diseases beyond autoimmune and autoinflammatory phenotypes.

Although previous efforts to objectively categorize these gene defects exist, they so far have been based on incomplete clinical terms (Grateau et al. 2013; Pathak, McDermott, and Savic 2017), or remained incomplete, as they relied on high-throughput data that is unavailable for most single-gene diseases (Arakelyan et al. 2017). Indeed, the unavailability of large-scale molecular datasets, together with the rarity of phenotypic patient profiles makes objective classification very challenging. We have used the AutoCore to identify molecular defined subclusters of rare autoimmune/autoinflammatory diseases. Although the AutoCore embodies a tightly connected subnetwork, we have found that the subclusters represent distinct molecular processes and cellular states. We found ample clustering of specific disease phenotypes and clinical groups, whereas less specific clinical phenotypes tend to be spread out across multiple clusters,

underpinning the power of this method to pinpoint clusters that are cohesive from a molecular mechanism standpoint. We further showcased that the AutoCore is therapeutically informative. We found that cluster-specific therapeutic targets tend to reside closer to the clusters they are used for, in comparison to other therapeutic targets used to treat autoimmune/autoinflammatory phenotypes. We hypothesize that this property of the AutoCore could be exploited for therapy repurposing. Network-based, disease-focused drug repurposing efforts for common, polygenic diseases such as for coronary heart disease (Cheng et al. 2018) have been published. Recently, a network-based methodology for the prediction of drug combinations that is particularly relevant for rare disease patients has been developed (Cheng, Kovács, and Barabási 2019). Most of these methods, however, rely on large-scale omics-data that is currently unavailable for rare autoimmune/autoinflammatory diseases to validate their findings. We propose that the AutoCore could first serve as a framework for network-based investigation for drug closeness, and as a framework to be used to generate larger-scale objective data sets regarding rare autoimmune/autoinflammatory diseases that can be used for validation.

Taken together, the results presented in our manuscript authenticate the utility of a unifying framework for investigating rare autoimmune and autoinflammatory diseases to quantify so far only descriptive clinical and phenotypic observations. Our results not only further the fundamental understanding of the interplay between autoimmunity, autoinflammation and the interactome, but also offer a lasting novel platform to systematically explore the molecular origins of immune homeostasis and dysregulation.

5 References

Abraham, Clara, and Judy H. Cho. 2009. “Inflammatory Bowel Disease.” *New England Journal of Medicine*.

Advances in Experimental Medicine and Biology. 1967. Plenum Press.

Aguet, François, Alvaro N. Barbeira, Rodrigo Bonazzola, Andrew Brown, Stephane E. Castel, Brian Jo, Silva Kasela, et al. n.d. “The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues.” https://doi.org/10.1101/787903.

Ahmed, R., and D. Gray. 1996. “Immunological Memory and Protective Immunity: Understanding Their Relation.” *Science* 272 (5258): 54–60.

Alanis-Lobato, Gregorio, Miguel A. Andrade-Navarro, and Martin H. Schaefer. 2016. “HIPPIE v2. 0: Enhancing Meaningfulness and Reliability of Protein--Protein Interaction Networks.” *Nucleic Acids Research*, gkw985.

Amberger, Joanna S., Carol A. Bocchini, Alan F. Scott, and Ada Hamosh. 2018. “OMIM.org: Leveraging Knowledge across Phenotype--Gene Relationships.” *Nucleic Acids Research* 47 (D1): D1038–43.

Andersen, Mads Hald, David Schrama, Per Thor Straten, and Jürgen C. Becker. 2006. “Cytotoxic T Cells.” *The Journal of Investigative Dermatology* 126 (1): 32–41.

Anderson, Ana C., and Vijay K. Kuchroo. 2007. “Autoimmune Response and Immune Tolerance.” *Immune Regulation and Immunotherapy in Autoimmune Disease*. https://doi.org/10.1007/978-0-387-36003-4_1.

Arakelyan, Arsen, Lilit Nersisyan, David Poghosyan, Lusine Khondkaryan, Anna Hakobyan, Henry Löffler-Wirth, Evie Melanitou, and Hans Binder. 2017. “Autoimmunity and Autoinflammation: A Systems View on Signaling Pathway Dysregulation Profiles.” *PloS One* 12 (11): e0187572.

Arbabi, Aryan, David R. Adams, Sanja Fidler, and Michael Brudno. 2019. “Identifying Clinical Terms in Medical Text Using Ontology-Guided Machine Learning.” *JMIR Medical Informatics* 7 (2): e12596.

Banchereau, Jacques, and Daniel Schmitt. 2012. *Dendritic Cells in Fundamental and Clinical Immunology: Volume 2*. Springer Science & Business Media.

Bansal, Mukesh, Jichen Yang, Charles Karan, Michael P. Menden, James C. Costello, Hao Tang, Guanghua Xiao, et al. 2014. “A Community Computational Challenge to Predict the Activity of Pairs of Compounds.” *Nature Biotechnology* 32 (12): 1213–22.

Barabási, Albert-László. 2007. “Network Medicine — From Obesity to the ‘Diseasome.’” *New England Journal of Medicine*. https://doi.org/10.1056/nejme078114.

Barabási, Albert-László, Natali Gulbahce, and Joseph Loscalzo. 2011a. “Network Medicine: A Network-Based Approach to Human Disease.” *Nature Reviews. Genetics* 12 (1): 56–68.

———. 2011b. “Network Medicine: A Network-Based Approach to Human Disease.” *Nature Reviews. Genetics* 12 (1): 56–68.

Barabási, Albert-László, and Zoltán N. Oltvai. 2004. “Network Biology: Understanding the Cell’s Functional Organization.” *Nature Reviews. Genetics* 5 (2): 101–13.

Barshir, Ruth, Omer Shwartz, Ilan Y. Smoly, and Esti Yeger-Lotem. 2014. “Comparative Analysis of Human Tissue Interactomes Reveals Factors Leading to Tissue-Specific Manifestation of Hereditary Diseases.” *PLoS Computational Biology* 10 (6): e1003632.

Bavelas, Alex. 1950. *Communication Patterns in Task-Oriented Groups*. na.

Beadle, G. W., and E. L. Tatum. 1941. “Genetic Control of Biochemical Reactions in Neurospora.” *Proceedings of the National Academy of Sciences of the United States of America* 27 (11): 499–506.

Beck, Tim, Tom Shorter, and Anthony J. Brookes. 2020. “GWAS Central: A Comprehensive Resource for the Discovery and Comparison of Genotype and Phenotype Data from Genome-Wide Association Studies.” *Nucleic Acids Research* 48 (D1): D933–40.

Benhar, Inbal, Anat London, and Michal Schwartz. 2012. “The Privileged Immunity of Immune Privileged Organs: The Case of the Eye.” *Frontiers in Immunology* 3 (September): 296.

Bernasconi, Nadia L., Elisabetta Traggiai, and Antonio Lanzavecchia. 2002. “Maintenance of Serological Memory by Polyclonal Activation of Human Memory B Cells.” *Science* 298 (5601): 2199–2202.

Bern, Miriam, Alexander King, Derek A. Applewhite, and Anna Ritz. 2019. “Network-Based Prediction of Polygenic Disease Genes Involved in Cell Motility.” *BMC Bioinformatics* 20 (Suppl 12): 313.

Bhattacharyya, S. B. 2016. “Introduction to SNOMED CT.” https://doi.org/10.1007/978-981-287-895-3.

Blackstone, Neil W. 2003. “Molecular Biology of the Cell. Fourth Edition.By Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and , Peter Walter. New York: Garland Science. $102.00. Xxxiv 1463 P; Ill.; Glossary (G:1–G:36); Index (I:1–I:49); Tables (T:1). ISBN: 0–8153–3218–1. [CD-ROM Included.] 2002.” *The Quarterly Review of Biology*. https://doi.org/10.1086/377840.

Blomen, Vincent A., Peter Májek, Lucas T. Jae, Johannes W. Bigenzahn, Joppe Nieuwenhuis, Jacqueline Staring, Roberto Sacco, et al. 2015. “Gene Essentiality and Synthetic Lethality in Haploid Human Cells.” *Science* 350 (6264): 1092–96.

Bodenreider, Olivier. 2004. “The Unified Medical Language System (UMLS): Integrating Biomedical Terminology.”

Nucleic Acids Research 32 (Database issue): D267–70.
Bossi, Alice, and Ben Lehner. 2009. “Tissue Specificity and the Human Protein Interaction Network.” *Molecular Systems Biology* 5 (April): 260.
Bousfiha, Ahmed Aziz, Leïla Jeddane, Fatima Ailal, Ibtihal Benhsaien, Nizar Mahlaoui, Jean-Laurent Casanova, and Laurent Abel. 2013. “Primary Immunodeficiency Diseases Worldwide: More Common than Generally Thought.” *Journal of Clinical Immunology* 33 (1): 1–7.
Bousfiha, Aziz, Leila Jeddane, Capucine Picard, Waleed Al-Herz, Fatima Ailal, Talal Chatila, Charlotte Cunningham-Rundles, et al. 2020. “Human Inborn Errors of Immunity: 2019 Update of the IUIS Phenotypical Classification.” *Journal of Clinical Immunology* 40 (1): 66–81.
Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. 2017. “An Expanded View of Complex Traits: From Polygenic to Omnigenic.” *Cell* 169 (7): 1177–86.
Braun, M. Miles, Sheiren Farag-El-Massah, Kui Xu, and Timothy R. Coté. 2010. “Emergence of Orphan Drugs in the United States: A Quantitative Assessment of the First 25 Years.” *Nature Reviews. Drug Discovery* 9 (7): 519–22.
Burnet, F. M. 2015. *The Clonal Selection Theory of Acquired Immunity - Scholar’s Choice Edition*. BiblioLife.
Caldera, Michael, Pisanu Buphamalai, Felix Müller, and Jörg Menche. 2017. “Interactome-Based Approaches to Human Disease.” *Current Opinion in Systems Biology* 3 (June): 88–94.
Caliskan, Minal, Christopher D. Brown, and Joseph C. Maranville. 2021. “A Catalog of GWAS Fine-Mapping Efforts in Autoimmune Disease.” *American Journal of Human Genetics* 108 (4): 549–63.
Cannizzo, Sara, Valentina Lorenzoni, Ilaria Palla, Salvatore Pirri, Leopoldo Trieste, Isotta Triulzi, and Giuseppe Turchetti. 2018. “Rare Diseases under Different Levels of Economic Analysis: Current Activities, Challenges and Perspectives.” *RMD Open* 4 (Suppl 1): e000794.
Carroll, Michael C. 2004. “The Complement System in Regulation of Adaptive Immunity.” *Nature Immunology* 5 (10): 981–86.
Chan, Stephen Y., and Joseph Loscalzo. 2012. “The Emerging Paradigm of Network Medicine in the Study of Human Disease.” *Circulation Research* 111 (3): 359–74.
Chazal, Pierre-Etienne, Anne-Sophie Chalandon, Ségolène Aymé, and Christian Deleuze. n.d. “Diagnostic Delay in Rare Diseases: A Documented List of (296) Rare Diseases for Which Delayed Diagnosis Would Be Especially Detrimental, Based on the French Situation.” https://doi.org/10.21203/rs.3.rs-32308/v1.
Cheng, Feixiong, Rishi J. Desai, Diane E. Handy, Ruisheng Wang, Sebastian Schneeweiss, Albert-László Barabási, and Joseph Loscalzo. 2018. “Network-Based Approach to Prediction and Population-Based Validation of in Silico Drug Repurposing.” *Nature Communications* 9 (1): 2691.
Cheng, Feixiong, István A. Kovács, and Albert-László Barabási. 2019. “Network-Based Prediction of Drug Combinations.” *Nature Communications* 10 (1): 1197.
Chmiel, Anna, Peter Klimek, and Stefan Thurner. 2014. “Spreading of Diseases through Comorbidity Networks across Life and Gender.” *New Journal of Physics* 16 (11): 115013.
Cho, Judy H., and Peter K. Gregersen. 2011. “Genomics and the Multifactorial Nature of Human Autoimmune Disease.” *The New England Journal of Medicine* 365 (17): 1612–23.
Choromański, Krzysztof, Michał Matuszak, and Jacek Miękisz. 2013. “Scale-Free Graph with Preferential Attachment and Evolving Internal Vertex Structure.” *Journal of Statistical Physics* 151 (6): 1175–83.
Cline, M. J. 1994. “Histiocytes and Histiocytosis [see Comments].” *Blood* 84 (9): 2840–53.
Cohen, Reuven, and Shlomo Havlin. 2003. “Scale-Free Networks Are Ultrasmall.” *Physical Review Letters* 90 (5): 058701.
Cooper, Glinda S., Milele L. K. Bynum, and Emily C. Somers. 2009. “Recent Insights in the Epidemiology of Autoimmune Diseases: Improved Prevalence Estimates and Understanding of Clustering of Diseases.” *Journal of Autoimmunity* 33 (3-4): 197–207.
Cooper, Max D., and Matthew N. Alder. 2006. “The Evolution of Adaptive Immune Systems.” *Cell* 124 (4): 815–22.
Cornish, Alex J., Alessia David, and Michael J. E. Sternberg. 2018. “PhenoRank: Reducing Study Bias in Gene Prioritization through Simulation.” *Bioinformatics* 34 (12): 2087–95.
Costanzo, Michael, Benjamin VanderSluis, Elizabeth N. Koch, Anastasia Baryshnikova, Carles Pons, Guihong Tan, Wen Wang, et al. 2016. “A Global Genetic Interaction Network Maps a Wiring Diagram of Cellular Function.” *Science* 353 (6306). https://doi.org/10.1126/science.aaf1420.
Cowen, Lenore, Trey Ideker, Benjamin J. Raphael, and Roded Sharan. 2017. “Network Propagation: A Universal Amplifier of Genetic Associations.” *Nature Reviews. Genetics* 18 (9): 551–62.
Csermely, Peter, Tamás Korcsmáros, Huba J. M. Kiss, Gábor London, and Ruth Nussinov. 2013. “Structure and Dynamics of Molecular Networks: A Novel Paradigm of Drug Discovery: A Comprehensive Review.” *Pharmacology & Therapeutics* 138 (3): 333–408.
Davies, Huw. 1997. “Evolution of the Adaptive Immune System.” *Introductory Immunobiology*. https://doi.org/10.1007/978-1-4899-3170-2_7.
Decker, K. 1990. “Biologically Active Products of Stimulated Liver Macrophages (Kupffer Cells).” *European Journal of Biochemistry / FEBS* 192 (2): 245–61.
Delhalle, Sylvie, Sebastian F. N. Bode, Rudi Balling, Markus Ollert, and Feng Q. He. 2018. “A Roadmap towards Personalized Immunology.” *NPJ Systems Biology and Applications* 4 (February): 9.

Delmonte, Ottavia Maria, Riccardo Castagnoli, Enrica Calzoni, and Luigi Daniele Notarangelo. 2019. “Inborn Errors of Immunity With Immune Dysregulation: From Bench to Bedside.” *Frontiers in Pediatrics* 7 (August): 353.
Delves, Peter J. 1998. “Autoimmunity.” In *Encyclopedia of Immunology (Second Edition)*, edited by Peter J. Delves, 292–96. Oxford: Elsevier.
Deng, Minghua, Shipra Mehta, Fengzhu Sun, and Ting Chen. 2002. “Inferring Domain-Domain Interactions from Protein-Protein Interactions.” In *Proceedings of the Sixth Annual International Conference on Computational Biology*, 117–26. RECOMB ’02. New York, NY, USA: Association for Computing Machinery.
Denmark, Vera Kandror, and Lloyd Mayer. 2014. “Inflammatory Bowel Diseases.” *The Autoimmune Diseases*. https://doi.org/10.1016/b978-0-12-384929-8.00060-5.
De Smet, Riet, and Kathleen Marchal. 2010. “Advantages and Limitations of Current Network Inference Methods.” *Nature Reviews. Microbiology* 8 (10): 717–29.
Dezső, Zoltán, Yuri Nikolsky, Tatiana Nikolskaya, Jeremy Miller, David Cherba, Craig Webb, and Andrej Bugrim. 2009. “Identifying Disease-Specific Genes Based on Their Topological Significance in Protein Networks.” *BMC Systems Biology* 3 (1): 36.
Di Noia, Javier M., and Michael S. Neuberger. 2007. “Molecular Mechanisms of Antibody Somatic Hypermutation.” *Annual Review of Biochemistry* 76: 1–22.
Dixon-Zegeye, Miriam, and Andrew Rutherford. 2020. “Infections in Systemic Autoimmune Diseases.” *Handbook of Systemic Autoimmune Diseases*.
Doench, John G., Nicolo Fusi, Meagan Sullender, Mudra Hegde, Emma W. Vaimberg, Katherine F. Donovan, Ian Smith, et al. 2016. “Optimized sgRNA Design to Maximize Activity and Minimize off-Target Effects of CRISPR-Cas9.” *Nature Biotechnology* 34 (2): 184–91.
Dorel, Mathurin, Bertram Klinger, Torsten Gross, Anja Sieber, Anirudh Prahallad, Evert Bosdriesz, Lodewyk F. A. Wessels, and Nils Blüthgen. 2018. “Modelling Signalling Networks from Perturbation Data.” *Bioinformatics* 34 (23): 4079–86.
Driel, Marc A. van, Jorn Bruggeman, Gert Vriend, Han G. Brunner, and Jack A. M. Leunissen. 2006. “A Text-Mining Analysis of the Human Phenome.” *European Journal of Human Genetics*. https://doi.org/10.1038/sj.ejhg.5201585.
Edris, Badreddin. 2011. “A Comparison of the Oligomap and TargetScan Algorithms for miRNA Target Analysis.” *Bmi231. Stanford. Edu*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.181.342&rep=rep1&type=pdf.
Elgueta, Raul, Victor C. de Vries, and Randolph J. Noelle. 2010. “The Immortality of Humoral Immunity.” *Immunological Reviews* 236 (July): 139–50.
Eurodis, Orphanet. n.d. “Rare Disease in Numbers: Preliminary Report from an on Going Bibliographic Study Initiated by Eurordis in Partnership with Orphanet.”
Eyre, Stephen, Gisela Orozco, and Jane Worthington. 2017. “The Genetics Revolution in Rheumatology: Large Scale Genomic Arrays and Genetic Mapping.” *Nature Reviews Rheumatology*. https://doi.org/10.1038/nrrheum.2017.80.
Fagerberg, Linn, Björn M. Hallström, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, et al. 2014. “Analysis of the Human Tissue-Specific Expression by Genome-Wide Integration of Transcriptomics and Antibody-Based Proteomics.” *Molecular & Cellular Proteomics: MCP* 13 (2): 397–406.
Fang, Han, Yiyang Wu, Hui Yang, Margaret Yoon, Laura T. Jiménez-Barrón, David Mittelman, Reid Robison, Kai Wang, and Gholson J. Lyon. 2017. “Whole Genome Sequencing of One Complex Pedigree Illustrates Challenges with Genomic Medicine.” *BMC Medical Genomics* 10 (1): 10.
Fazilleau, Nicolas, Linda Mark, Louise J. McHeyzer-Williams, and Michael G. McHeyzer-Williams. 2009. “Follicular Helper T Cells: Lineage and Location.” *Immunity* 30 (3): 324–35.
Feldman, Igor, Andrey Rzhetsky, and Dennis Vitkup. 2008. “Network Properties of Genes Harboring Inherited Disease Mutations.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (11): 4323–28.
Ferlini, A. 2017. “Solve-Rd” 1 (1): 1–1.
Ferrier, Pierre, ed. 2009. *V(D)J Recombination*. PDF. 2009th ed. Advances in Experimental Medicine and Biology. New York, NY: Springer. https://play.google.com/store/books/details?id=mGKCeEgsoWkC.
Fischer, Alain, Johan Provot, Jean-Philippe Jais, Alexandre Alcais, Nizar Mahlaoui, and members of the CEREDIH French PID study group. 2017. “Autoimmune and Inflammatory Manifestations Occur Frequently in Patients with Primary Immunodeficiencies.” *The Journal of Allergy and Clinical Immunology* 140 (5): 1388–93.e8.
Fleit, Howard B. 2012. “Janeway’s Immunobiology. Eighth Edition. By Kenneth Murphy; with Acknowledgment To, Charles A. JanewayJr., Paul Travers, and Mark Walport; with Contributions By, Allan Mowat and Casey T. Weaver. London and New York: Garland Science (Taylor & Francis Group). $125.00 (paper). Xx 868 P.; Ill.; Index. ISBN: 978-0-8153-4243-4. 2012.” *The Quarterly Review of Biology*. https://doi.org/10.1086/666778.
Fletcher, Anne L., Veronika Lukacs-Kornek, Erika D. Reynoso, Sophie E. Pinner, Angelique Bellemare-Pelletier, Mark S. Curry, Ai-Ris Collier, Richard L. Boyd, and Shannon J. Turley. 2010. “Lymph Node Fibroblastic Reticular Cells Directly Present Peripheral Tissue Antigen under Steady-State and Inflammatory Conditions.” *The Journal of Experimental Medicine* 207 (4): 689–97.
Fortunato, Santo. 2010. “Community Detection in Graphs.” *Physics Reports* 486 (3): 75–174.

Freeman, Linton C. 1977. “A Set of Measures of Centrality Based on Betweenness.” *Sociometry* 40 (1): 35–41.
Friedel, Caroline C., and Ralf Zimmer. 2006. “Toward the Complete Interactome.” *Nature Biotechnology*.
Frohman, Elliot M., Michael K. Racke, and Cedric S. Raine. 2006. “Multiple Sclerosis — The Plaque and Its Pathogenesis.” *New England Journal of Medicine*. https://doi.org/10.1056/nejmra052130.
Fujiwara, Nagatoshi, and Kazuo Kobayashi. 2005. “Macrophages in Inflammation.” *Current Drug Targets. Inflammation and Allergy* 4 (3): 281–86.
Fujiwara, Toyofumi, Yasunori Yamamoto, Jin-Dong Kim, Orion Buske, and Toshihisa Takagi. 2018. “PubCaseFinder: A Case-Report-Based, Phenotype-Driven Differential-Diagnosis System for Rare Diseases.” *American Journal of Human Genetics* 103 (3): 389–99.
Gahl, William A., Thomas C. Markello, Camilo Toro, Karin Fuentes Fajardo, Murat Sincan, Fred Gill, Hannah Carlson-Donohoe, et al. 2012. “The National Institutes of Health Undiagnosed Diseases Program: Insights into Rare Diseases.” *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 14 (1): 51–59.
Gahl, William A., John J. Mulvihill, Camilo Toro, Thomas C. Markello, Anastasia L. Wise, Rachel B. Ramoni, David R. Adams, Cynthia J. Tifft, and UDN. 2016. “The NIH Undiagnosed Diseases Program and Network: Applications to Modern Medicine.” *Molecular Genetics and Metabolism* 117 (4): 393–400.
Gainotti, Sabina, Paola Torreri, Chiuhui Mary Wang, Robert Reihs, Heimo Mueller, Emma Heslop, Marco Roos, et al. 2018. “The RD-Connect Registry & Biobank Finder: A Tool for Sharing Aggregated Data and Metadata among Rare Disease Researchers.” *European Journal of Human Genetics: EJHG* 26 (5): 631–43.
Gallet, X., B. Charloteaux, A. Thomas, and R. Brasseur. 2000. “A Fast Method to Predict Protein Interaction Sites from Sequences.” *Journal of Molecular Biology* 302 (4): 917–26.
Galli, Stephen J., Devavani Chatterjea, and Mindy Tsai. 2014. “Roles of Mast Cells and Basophils in Innate Immunity.” *The Innate Immune Response to Infection*. https://doi.org/10.1128/9781555817671.ch6.
Gasteiger, L. M., P. N. VRobinson, F. Pazmandi, and Others. 2020. “Registry Working Party of the European Society for Immunodeficiencies (ESID). Supplementation of the ESID Registry Working Definitions for the Clinical Diagnosis of Inborn Errors of Immunity with Encoded Human Phenotype Ontology (HPO) Terms.” *The Journal of Allergy and Clinical Immunology in Practice* 8: 1778.
George, Richard A., Jason Y. Liu, Lina L. Feng, Robert J. Bryson-Richardson, Diane Fatkin, and Merridee A. Wouters. 2006. “Analysis of Protein Sequence and Interaction Data for Candidate Disease Gene Prediction.” *Nucleic Acids Research* 34 (19): e130.
Ghiassian, Susan Dina, Jörg Menche, and Albert-László Barabási. 2015. “A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome.” *PLoS Computational Biology* 11 (4): e1004120.
Gillis, Jesse, Sara Ballouz, and Paul Pavlidis. 2014. “Bias Tradeoffs in the Creation and Analysis of Protein–protein Interaction Networks.” *Journal of Proteomics* 100 (April): 44–54.
Girardi, Michael, David E. Oppenheim, Carrie R. Steele, Julia M. Lewis, Earl Glusac, Renata Filler, Paul Hobby, Brian Sutton, Robert E. Tigelaar, and Adrian C. Hayday. 2001. “Regulation of Cutaneous Malignancy by γδ T Cells.” *Science* 294 (5542): 605–9.
Girdea, Marta, Sergiu Dumitriu, Marc Fiume, Sarah Bowdin, Kym M. Boycott, Sébastien Chénier, David Chitayat, et al. 2013. “PhenoTips: Patient Phenotyping Software for Clinical and Research Use.” *Human Mutation* 34 (8): 1057–65.
Gkoutos, Georgios V., Paul N. Schofield, and Robert Hoehndorf. 2018. “The Anatomy of Phenotype Ontologies: Principles, Properties and Applications.” *Briefings in Bioinformatics* 19 (5): 1008–21.
Glauser, Michel P. 1996. “The Inflammatory Cytokines.” *Drugs* 52 (2): 9–17.
Goh, Kwang-Il, Michael E. Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. 2007. “The Human Disease Network.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (21): 8685–90.
Grateau, Gilles, Véronique Hentgen, Katia Stankovic Stojanovic, Isabelle Jéru, Serge Amselem, and Olivier Steichen. 2013. “How Should We Approach Classification of Autoinflammatory Diseases?” *Nature Reviews. Rheumatology* 9 (10): 624–29.
Grau, Bernardo Cuenca, Ian Horrocks, Boris Motik, Bijan Parsia, Peter Patel-Schneider, and Ulrike Sattler. 2008. “OWL 2: The next Step for OWL.” *Journal of Web Semantics* 6 (4): 309–22.
Grimbacher, B., and ESID Registry Working Party. 2014. “The European Society for Immunodeficiencies (ESID) Registry 2014.” *Clinical and Experimental Immunology* 178 Suppl 1 (December): 18–20.
Guermonprez, Pierre, Jenny Valladeau, Laurence Zitvogel, Clotilde Théry, and Sebastian Amigorena. 2002. “Antigen Presentation and T Cell Stimulation by Dendritic Cells.” *Annual Review of Immunology* 20: 621–67.
Guney, Emre, Jörg Menche, Marc Vidal, and Albert-László Barábasi. 2016. “Network-Based in Silico Drug Efficacy Screening.” *Nature Communications* 7 (February): 10331.
Guney, Emre, and Baldo Oliva. 2012. “Exploiting Protein-Protein Interaction Networks for Genome-Wide Disease-Gene Prioritization.” *PloS One* 7 (9): e43557.
Hartwell, L. H., J. J. Hopfield, S. Leibler, and A. W. Murray. 1999. “From Molecular to Modular Cell Biology.” *Nature* 402 (6761 Suppl): C47–52.
Hayden, M. S., A. P. West, and S. Ghosh. 2006. “NF-κB and the Immune Response.” *Oncogene*.

https://doi.org/10.1038/sj.onc.1209943.
Hecker, Michael, Sandro Lambeck, Susanne Toepfer, Eugene van Someren, and Reinhard Guthke. 2009. “Gene Regulatory Network Inference: Data Integration in Dynamic models—A Review.” *Bio Systems* 96 (1): 86–103.
Hemminki, Kari, Wuqing Huang, Jan Sundquist, Kristina Sundquist, and Jianguang Ji. 2020. “Autoimmune Diseases and Hematological Malignancies: Exploring the Underlying Mechanisms from Epidemiological Evidence.” *Seminars in Cancer Biology* 64 (August): 114–21.
Hidalgo, César A., Nicholas Blumm, Albert-László Barabási, and Nicholas A. Christakis. 2009. “A Dynamic Network Approach for the Study of Human Phenotypes.” *PLoS Computational Biology*.
Hoffman, William, Fadi G. Lakkis, and Geetha Chalasani. 2016. “B Cells, Antibodies, and More.” *Clinical Journal of the American Society of Nephrology: CJASN* 11 (1): 137–54.
Hogan, Simon P., Helene F. Rosenberg, Redwan Moqbel, Simon Phipps, Paul S. Foster, Paige Lacy, A. Barry Kay, and Marc E. Rothenberg. 2008. “Eosinophils: Biological Properties and Role in Health and Disease.” *Clinical and Experimental Allergy: Journal of the British Society for Allergy and Clinical Immunology* 38 (5): 709–50.
Hopkins, Andrew L. 2008. “Network Pharmacology: The next Paradigm in Drug Discovery.” *Nature Chemical Biology* 4 (11): 682–90.
“HPO - ORDO Ontological Module - Association - Classes.” n.d. Accessed May 4, 2021. https://bioportal.bioontology.org/ontologies/HOOM?p=classes&conceptid=root.
Hu, Jessica Xin, Cecilia Engel Thomas, and Søren Brunak. 2016. “Network Biology Concepts in Complex Disease Comorbidities.” *Nature Reviews. Genetics* 17 (10): 615–29.
Hussein, Hadi M., and Elias A. Rahal. 2019. “The Role of Viral Infections in the Development of Autoimmune Diseases.” *Critical Reviews in Microbiology* 45 (4): 394–412.
Huttlin, Edward L., Raphael J. Bruckner, Joao A. Paulo, Joe R. Cannon, Lily Ting, Kurt Baltier, Greg Colby, et al. 2017. “Architecture of the Human Interactome Defines Protein Communities and Disease Networks.” *Nature* 545 (7655): 505–9.
Huttlin, Edward L., Lily Ting, Raphael J. Bruckner, Fana Gebreab, Melanie P. Gygi, John Szpyt, Stanley Tam, et al. 2015. “The BioPlex Network: A Systematic Exploration of the Human Interactome.” *Cell* 162 (2): 425–40.
“IAMRARE® Registry Program.” 2017. February 16, 2017. https://rarediseases.org/iamrare-registry-program/.
Isik, Zerrin, Christoph Baldow, Carlo Vittorio Cannistraci, and Michael Schroeder. 2015. “Drug Target Prioritization by Perturbed Gene Expression and Network Information.” *Scientific Reports* 5 (November): 17417.
Janeway, C. A., P. Travers, M. Walport, and M. J. Shlomchik. 2005. “An Introduction to Immunobiology and Innate Immunity.” *Immunobiology. 6th Ed. New York, NY: Ed Garland Science Publishing. P*, 1–35.
Janeway, Charles A. n.d. “J., Travers, P., Walport, M., & Shlomchik, MJ (2001). The Complement System and Innate Immunity.” *Immunobiology: The Immune System in Health and Disease*.
Jeong, H., S. P. Mason, A. L. Barabási, and Z. N. Oltvai. 2001. “Lethality and Centrality in Protein Networks.” *Nature* 411 (6833): 41–42.
Kanehisa, M., and S. Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28 (1): 27–30.
Kanehisa, Minoru, Michihiro Araki, Susumu Goto, Masahiro Hattori, Mika Hirakawa, Masumi Itoh, Toshiaki Katayama, et al. 2008. “KEGG for Linking Genomes to Life and the Environment.” *Nucleic Acids Research* 36 (Database issue): D480–84.
Kang, Tianyu, Wei Ding, Luoyan Zhang, Daniel Ziemek, and Kourosh Zarringhalam. 2017. “A Biological Network-Based Regularized Artificial Neural Network Model for Robust Phenotype Prediction from Gene Expression Data.” *BMC Bioinformatics* 18 (1): 565.
Kastner, Daniel L., Ivona Aksentijevich, and Raphaela Goldbach-Mansky. 2010. “Autoinflammatory Disease Reloaded: A Clinical Perspective.” *Cell*. https://doi.org/10.1016/j.cell.2010.03.002.
Katz, Leo. 1953. “A New Status Index Derived from Sociometric Analysis.” *Psychometrika* 18 (1): 39–43.
Keiser, Michael J., Vincent Setola, John J. Irwin, Christian Laggner, Atheir I. Abbas, Sandra J. Hufeisen, Niels H. Jensen, et al. 2009. “Predicting New Molecular Targets for Known Drugs.” *Nature* 462 (7270): 175–81.
Kerrien, Samuel, Bruno Aranda, Lionel Breuza, Alan Bridge, Fiona Broackes-Carter, Carol Chen, Margaret Duesbury, et al. 2012. “The IntAct Molecular Interaction Database in 2012.” *Nucleic Acids Research* 40 (Database issue): D841–46.
Kim, Daniel, and John Rossi. 2008. “RNAi Mechanisms and Applications.” *BioTechniques* 44 (5): 613–16.
Kitsak, Maksim, Amitabh Sharma, Jörg Menche, Emre Guney, Susan Dina Ghiassian, Joseph Loscalzo, and Albert-László Barabási. 2016. “Tissue Specificity of Human Disease Module.” *Scientific Reports* 6 (October): 35241.
Klenerman, Paul. 2017. *The Immune System: A Very Short Introduction*. Oxford University Press.
Klimek, Peter, Silke Aichberger, and Stefan Thurner. 2016. “Disentangling Genetic and Environmental Risk Factors for Individual Diseases from Multiplex Comorbidity Networks.” *Scientific Reports* 6 (December): 39658.
Köhler, Sebastian. 2018. “Improved Ontology-Based Similarity Calculations Using a Study-Wise Annotation Model.” *Database: The Journal of Biological Databases and Curation* 2018 (January). https://doi.org/10.1093/database/bay026.
Köhler, Sebastian, Gerhard Kindle, and Peter N. Robinson. 2021. *The Human Phenotype Ontology in 2021*. Universität.

Köhler, Sebastian, Nicole A. Vasilevsky, Mark Engelstad, Erin Foster, Julie McMurry, Ségolène Aymé, Gareth Baynam, et al. 2017. “The Human Phenotype Ontology in 2017.” *Nucleic Acids Research* 45 (D1): D865–76.
Kölmel, Hans Wolfgang. 1977. “Plasma Cells.” *Atlas of Cerebrospinal Fluid Cells*. https://doi.org/10.1007/978-3-642-66623-0_5.
Korn, Thomas, Estelle Bettelli, Mohamed Oukka, and Vijay K. Kuchroo. 2009. “IL-17 and Th17 Cells.” *Annual Review of Immunology* 27 (1): 485–517.
Köstel Bal, Sevgi, Julia Pazmandi, Kaan Boztug, and Seza Özen. 2020. “Rheumatological Manifestations in Inborn Errors of Immunity.” *Pediatric Research* 87 (2): 293–99.
Koyama, Shohei, Ken J. Ishii, Cevayir Coban, and Shizuo Akira. 2008. “Innate Immune Response to Viral Infection.” *Cytokine* 43 (3): 336–41.
Krauthammer, Michael, Charles A. Kaufmann, T. Conrad Gilliam, and Andrey Rzhetsky. 2004. “Molecular Triangulation: Bridging Linkage and Molecular-Network Information for Identifying Candidate Genes in Alzheimer’s Disease.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (42): 15148–53.
Kreutzer, Jeffrey S., John DeLuca, and Bruce Caplan, eds. 2011. “International Statistical Classification of Diseases and Related Health Problems.” In *Encyclopedia of Clinical Neuropsychology*, 1347–1347. New York, NY: Springer New York.
Kubicek, Stefan, Joshua C. Gilbert, Dina Fomina-Yadlin, Alexander D. Gitlin, Yuan Yuan, Florence F. Wagner, Edward B. Holson, et al. 2012. “Chromatin-Targeting Small Molecules Cause Class-Specific Transcriptional Changes in Pancreatic Endocrine Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (14): 5364–69.
La Cava, Antonio. 2009. “Natural Tregs and Autoimmunity.” *Frontiers in Bioscience* 14 (January): 333–43.
Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921.
Le, Duc-Hau, and Lan T. M. Dao. 2018. “Annotating Diseases Using Human Phenotype Ontology Improves Prediction of Disease-Associated Long Non-Coding RNAs.” *Journal of Molecular Biology* 430 (15): 2219–30.
Lee, D. -. S., J. Park, K. A. Kay, N. A. Christakis, Z. N. Oltvai, and A. -. L. Barabasi. 2008. “The Implications of Human Metabolic Network Topology for Disease Comorbidity.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0802208105.
Lewis, Janine, Michelle Snyder, and Henrietta Hyatt-Knorr. 2017. “Marking 15 Years of the Genetic and Rare Diseases Information Center.” *Transl Sci Rare Dis* 2 (1-2): 77–88.
Li, Jiao, Si Zheng, Bin Chen, Atul J. Butte, S. Joshua Swamidass, and Zhiyong Lu. 2016. “A Survey of Current Trends in Computational Drug Repositioning.” *Briefings in Bioinformatics* 17 (1): 2–12.
Li, Min, Jiayi Zhang, Qing Liu, Jianxin Wang, and Fang-Xiang Wu. 2014. “Prediction of Disease-Related Genes Based on Weighted Tissue-Specific Networks by Using DNA Methylation.” *BMC Medical Genomics* 7 Suppl 2 (October): S4.
Lin, Dekang, and Others. 1998. “An Information-Theoretic Definition of Similarity.” In *Icml*, 98:296–304.
Lindsay Whitton, J. 2013. *Antigen Presentation*. Springer Science & Business Media.
Li, Qi, Cheng Hiang Lee, Lauren A. Peters, Lucas A. Mastropaolo, Cornelia Thoeni, Abdul Elkadri, Tobias Schwerd, et al. 2016. “Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease.” *Gastroenterology* 150 (5): 1196–1207.
Liu, Cong, Fabricio Sampaio Peres Kury, Ziran Li, Casey Ta, Kai Wang, and Chunhua Weng. 2019. “Doc2Hpo: A Web Application for Efficient and Accurate HPO Concept Curation.” *Nucleic Acids Research* 47 (W1): W566–70.
Liu, Wei, Jianying Wang, Tengjiao Wang, and Hongwei Xie. 2014. “Construction and Analyses of Human Large-Scale Tissue Specific Networks.” *PloS One* 9 (12): e115074.
Liu, Zhichao, Liyuan Zhu, Ruth Roberts, and Weida Tong. 2019. “Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We?” *Trends in Genetics: TIG* 35 (11): 852–67.
Li, Zhan-Chao, Meng-Hua Huang, Wen-Qian Zhong, Zhi-Qing Liu, Yun Xie, Zong Dai, and Xiao-Yong Zou. 2016. “Identification of Drug-Target Interaction from Interactome Network with ‘Guilt-by-Association’ Principle and Topology Features.” *Bioinformatics* 32 (7): 1057–64.
López-Otín, Carlos, and Guido Kroemer. 2021. “Hallmarks of Health.” *Cell* 184 (1): 33–63.
Luck, Katja, Dae-Kyum Kim, Luke Lambourne, Kerstin Spirohn, Bridget E. Begg, Wenting Bian, Ruth Brignall, et al. 2020. “A Reference Map of the Human Binary Protein Interactome.” *Nature* 580 (7803): 402–8.
MacArthur, Jacqueline, Emily Bowler, Maria Cerezo, Laurent Gil, Peggy Hall, Emma Hastings, Heather Junkins, et al. 2017. “The New NHGRI-EBI Catalog of Published Genome-Wide Association Studies (GWAS Catalog).” *Nucleic Acids Research* 45 (D1): D896–901.
Magger, Oded, Yedael Y. Waldman, Eytan Ruppin, and Roded Sharan. 2012. “Enhancing the Prioritization of Disease-Causing Genes through Tissue Specific Protein Interaction Networks.” *PLoS Computational Biology* 8 (9): e1002690.
Ma, Hongwu, Anatoly Sorokin, Alexander Mazein, Alex Selkov, Evgeni Selkov, Oleg Demin, and Igor Goryanin. 2007. “The Edinburgh Human Metabolic Network Reconstruction and Its Functional Analysis.” *Molecular Systems*

Biology 3 (September): 135.
Malhotra, Deepali, Jonathan L. Linehan, Thamotharampillai Dileepan, You Jeong Lee, Whitney E. Purtha, Jennifer V. Lu, Ryan W. Nelson, et al. 2016. “Tolerance Is Established in Polyclonal CD4(+) T Cells by Distinct Mechanisms, according to Self-Peptide Expression Patterns.” *Nature Immunology* 17 (2): 187–95.
Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, et al. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461 (7265): 747–53.
Marcotte, Cynthia J. Verjovsky, and Edward M. Marcotte. 2002. “Predicting Functional Linkages from Gene Fusions with Confidence.” *Applied Bioinformatics* 1 (2): 93–100.
Markt, Patrick, Gerhard Dürnberger, Jacques Colinge, and Stefan Kubicek. 2012. “CLOUD – CeMM Library of Unique Drugs.” *Journal of Cheminformatics*. https://doi.org/10.1186/1758-2946-4-s1-p23.
Maron, Bradley A., Rui-Sheng Wang, Sergei Shevtsov, Stavros G. Drakos, Elena Arons, Omar Wever-Pinzon, Gordon S. Huggins, et al. 2021. “Individualized Interactomes for Network-Based Precision Medicine in Hypertrophic Cardiomyopathy with Implications for Other Clinical Pathophenotypes.” *Nature Communications*. https://doi.org/10.1038/s41467-021-21146-y.
Marone, Gianni, Massimo Triggiani, Arturo Genovese, and Amato De Paulis. 2005. “Role of Human Mast Cells and Basophils in Bronchial Asthma.” *Advances in Immunology* 88: 97–160.
Martín, Alejandro, Raquel Hervás, Gonzalo Méndez, and Susana Bautista. 2018. “PICTAR: Una Herramienta de Elaboración de Contenido Para Personas Con TEA Basada En La Traducción de Texto a Pictogramas.” In *Proceedings of the 19th International Conference on Human--Computer Interaction (Interacción 2018)*. http://nil.fdi.ucm.es/sites/default/files/Pictar-Interaccion2018.pdf.
Matys, V., E. Fricke, R. Geffers, E. Gößling, M. Haubrock, R. Hehl, K. Hornischer, et al. 2003. “TRANSFAC ® : Transcriptional Regulation, from Patterns to Profiles.” *Nucleic Acids Research* 31 (1): 374–78.
Mayor, Paul C., Kevin H. Eng, Kelly L. Singel, Scott I. Abrams, Kunle Odunsi, Kirsten B. Moysich, Ramsay Fuleihan, et al. 2018. “Cancer in Primary Immunodeficiency Diseases: Cancer Incidence in the United States Immune Deficiency Network Registry.” *The Journal of Allergy and Clinical Immunology* 141 (3): 1028–35.
McCarthy, Mark I., Gonçalo R. Abecasis, Lon R. Cardon, David B. Goldstein, Julian Little, John P. A. Ioannidis, and Joel N. Hirschhorn. 2008. “Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2344.
McGonagle, Dennis, and Michael F. McDermott. 2006. “A Proposed Classification of the Immunological Diseases.” *PLoS Medicine* 3 (8): e297.
McKusick, Victor A. 2007. “Mendelian Inheritance in Man and Its Online Version, OMIM.” *American Journal of Human Genetics* 80 (4): 588–604.
Medawar, P. B. 1961. “Immunological Tolerance.” *Science* 133 (3449): 303–6.
Meffre, Eric, and Hedda Wardemann. 2008. “B-Cell Tolerance Checkpoints in Health and Autoimmunity.” *Current Opinion in Immunology* 20 (6): 632–38.
Melé, Marta, Pedro G. Ferreira, Ferran Reverter, David S. DeLuca, Jean Monlong, Michael Sammeth, Taylor R. Young, et al. 2015. “The Human Transcriptome across Tissues and Individuals.” *Science* 348 (6235): 660–65.
Menche, Jörg, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, and Albert-László Barabási. 2015a. “Disease Networks. Uncovering Disease-Disease Relationships through the Incomplete Interactome.” *Science* 347 (6224): 1257601.
———. 2015b. “Disease Networks. Uncovering Disease-Disease Relationships through the Incomplete Interactome.” *Science* 347 (6224): 1257601.
MENDEL, and G. 1865. “Versuche Uber Pflanzen-Hybriden.” *Vorgelegt in Den Sitzungen*. https://ci.nii.ac.jp/naid/10018628927/.
Meresse, B., J. Ripoche, M. Heyman, and N. Cerf-Bensussan. 2009. “Celiac Disease: From Oral Tolerance to Intestinal Inflammation, Autoimmunity and Lymphomagenesis.” *Mucosal Immunology*. https://doi.org/10.1038/mi.2008.75.
Metcalfe, D. D., D. Baram, and Y. A. Mekori. 1997. “Mast Cells.” *Physiological Reviews* 77 (4): 1033–79.
Moder, Martin, Georgia Velimezi, Michel Owusu, Abdelghani Mazouzi, Marc Wiedner, Joana Ferreira da Silva, Lydia Robinson-Garcia, et al. 2017. “Parallel Genome-Wide Screens Identify Synthetic Viable Interactions between the BLM Helicase Complex and Fanconi Anemia.” *Nature Communications* 8 (1): 1238.
Molinelli, Evan J., Anil Korkut, Weiqing Wang, Martin L. Miller, Nicholas P. Gauthier, Xiaohong Jing, Poorvi Kaushik, et al. 2013. “Perturbation Biology: Inferring Signaling Networks in Cellular Systems.” *PLoS Computational Biology* 9 (12): e1003290.
Monie, Tom P. 2017. “The Innate Immune System in Health and Disease.” *The Innate Immune System*. https://doi.org/10.1016/b978-0-12-804464-3.00006-5.
Moreira, Alvaro, Friedrich-Alexander-University Erlangen-Nürnberg, Germany, Barbara Torres, Juliano Peruzzo, Alberto Mota, Kilian Eyerich, et al. 2017. “Skin Symptoms as Diagnostic Clue for Autoinflammatory Diseases.” *Anais Brasileiros de Dermatologia*. https://doi.org/10.1590/abd1806-4841.20175208.
Mosmann, T. R., and R. L. Coffman. 1989. “TH1 and TH2 Cells: Different Patterns of Lymphokine Secretion Lead to Different Functional Properties.” *Annual Review of Immunology* 7 (1): 145–73.
Mueller, Daniel L. 2010. “Mechanisms Maintaining Peripheral Tolerance.” *Nature Immunology* 11 (1): 21–27.

Nauseef, William M., and Niels Borregaard. 2014. “Neutrophils at Work.” *Nature Immunology* 15 (7): 602–11.
Neefjes, Jacques, Marlieke L. M. Jongsma, Petra Paul, and Oddmund Bakke. 2011. “Towards a Systems Understanding of MHC Class I and MHC Class II Antigen Presentation.” *Nature Reviews. Immunology* 11 (12): 823–36.
Newman, Mark E. J. 2008. “The Mathematics of Networks.” *The New Palgrave Encyclopedia of Economics* 2 (2008): 1–12.
Noh, Heeju, Jason E. Shoemaker, and Rudiyanto Gunawan. 2018. “Network Perturbation Analysis of Gene Transcriptional Profiles Reveals Protein Targets and Mechanism of Action of Drugs and Influenza A Viral Infection.” *Nucleic Acids Research* 46 (6): e34.
Novarino, Gaia, Ali G. Fenstermaker, Maha S. Zaki, Matan Hofree, Jennifer L. Silhavy, Andrew D. Heiberg, Mostafa Abdellateef, et al. 2014. “Exome Sequencing Links Corticospinal Motor Neuron Disease to Common Neurodegenerative Disorders.” *Science* 343 (6170): 506–11.
Nugent, Neill, and Mark Rhinard. 2015. *The European Commission*. Macmillan International Higher Education.
Ofran, Yanay, and Burkhard Rost. 2003. “Predicted Protein-Protein Interaction Sites from Local Sequence Information.” *FEBS Letters* 544 (1-3): 236–39.
Onnela, J-P, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A-L Barabási. 2007. “Structure and Tie Strengths in Mobile Communication Networks.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (18): 7332–36.
Oti, M., B. Snel, M. A. Huynen, and H. G. Brunner. 2006. “Predicting Disease Genes Using Protein-Protein Interactions.” *Journal of Medical Genetics* 43 (8): 691–98.
Parikshak, Neelroop N., Vivek Swarup, T. Grant Belgard, Manuel Irimia, Gokul Ramaswami, Michael J. Gandal, Christopher Hartl, et al. 2016. “Genome-Wide Changes in lncRNA, Splicing, and Regional Gene Expression Patterns in Autism.” *Nature* 540 (7633): 423–27.
Parlato, Marianna, Qing Nian, Fabienne Charbit-Henrion, Frank M. Ruemmele, Fernando Rodrigues-Lima, Nadine Cerf-Bensussan, and Immunobiota Study Group. 2020. “Loss-of-Function Mutation in PTPN2 Causes Aberrant Activation of JAK Signaling Via STAT and Very Early Onset Intestinal Inflammation.” *Gastroenterology* 159 (5): 1968–71.e4.
Pathak, Shelly, Michael F. McDermott, and Sinisa Savic. 2017. “Autoinflammatory Diseases: Update on Classification Diagnosis and Management.” *Journal of Clinical Pathology* 70 (1): 1–8.
Paulsen, Friedrich, Fabian Garreis, and Lars Bräuer. 2013. “Role of the Innate Immune System.” *Innate Immunity and the Eye*. https://doi.org/10.5005/jp/books/11844_1.
Paul, W. E., and R. A. Seder. 1994. “Lymphocyte Responses and Cytokines.” *Cell* 76 (2): 241–51.
Pavan, Sonia, Kathrin Rommel, María Elena Mateo Marquina, Sophie Höhn, Valérie Lanneau, and Ana Rath. 2017. “Clinical Practice Guidelines for Rare Diseases: The Orphanet Database.” *PloS One* 12 (1): e0170365.
Pellegrini, M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. 1999. “Assigning Protein Functions by Comparative Genome Analysis: Protein Phylogenetic Profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (8): 4285–88.
Perniola, Roberto. 2018. “Twenty Years of AIRE.” *Frontiers in Immunology* 9 (February): 98.
Peters, Lauren A., Jacqueline Perrigoue, Arthur Mortha, Alina Iuga, Won-Min Song, Eric M. Neiman, Sean R. Llewellyn, et al. 2017. “A Functional Genomics Predictive Network Model Identifies Regulators of Inflammatory Bowel Disease.” *Nature Genetics* 49 (10): 1437.
Piñero, Janet, Ariel Berenstein, Abel Gonzalez-Perez, Ariel Chernomoretz, and Laura I. Furlong. 2016. “Uncovering Disease Mechanisms through Network Biology in the Era of Next Generation Sequencing.” *Scientific Reports* 6 (April): 24570.
Posey, Jennifer E., Tamar Harel, Pengfei Liu, Jill A. Rosenfeld, Regis A. James, Zeynep H. Coban Akdemir, Magdalena Walkiewicz, et al. 2017. “Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation.” *The New England Journal of Medicine* 376 (1): 21–31.
Posey, Jennifer E., Jill A. Rosenfeld, Regis A. James, Matthew Bainbridge, Zhiyv Niu, Xia Wang, Shweta Dhar, et al. 2016. “Molecular Diagnostic Experience of Whole-Exome Sequencing in Adult Patients.” *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 18 (7): 678–85.
Prulj, Nataa. 2019. *Analyzing Network Data in Biology and Medicine: An Interdisciplinary Textbook for Biological, Medical and Computational Scientists*. Cambridge University Press.
Ramoni, Rachel B., John J. Mulvihill, David R. Adams, Patrick Allard, Euan A. Ashley, Jonathan A. Bernstein, William A. Gahl, et al. 2017. “The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease.” *American Journal of Human Genetics* 100 (2): 185–92.
“Rare Diseases Registry Program (RaDaR).” 2017. October 6, 2017. https://ncats.nih.gov/radar.
Raulet, David H. 2004. “Interplay of Natural Killer Cells and Their Receptors with the Adaptive Immune Response.” *Nature Immunology* 5 (10): 996–1002.
Reaves, Natalie D. 2003. “A Model of Effective Health Policy: The 1983 Orphan Drug Act.” *Journal of Health & Social Policy* 17 (4): 61–71.
Resnik, Philip. 1995. “Using Information Content to Evaluate Semantic Similarity in a Taxonomy.” *arXiv [cmp-Lg]*. arXiv. http://arxiv.org/abs/cmp-lg/9511007.

Retterer, Kyle, Jane Juusola, Megan T. Cho, Patrik Vitazka, Francisca Millan, Federica Gibellini, Annette Vertino-Bell, et al. 2016. “Clinical Application of Whole-Exome Sequencing across Clinical Indications.” *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 18 (7): 696–704.
Ricklin, Daniel, Edimara S. Reis, and John D. Lambris. 2016. “Complement in Disease: A Defence System Turning Offensive.” *Nature Reviews. Nephrology* 12 (7): 383–401.
Rieckmann, Jan C., Roger Geiger, Daniel Hornburg, Tobias Wolf, Ksenya Kveler, David Jarrossay, Federica Sallusto, et al. 2017. “Social Network Architecture of Human Immune Cells Unveiled by Quantitative Proteomics.” *Nature Immunology* 18 (5): 583–93.
Rivas, Manuel A., Mélissa Beaudoin, Agnes Gardet, Christine Stevens, Yashoda Sharma, Clarence K. Zhang, Gabrielle Boucher, et al. 2011. “Deep Resequencing of GWAS Loci Identifies Independent Rare Variants Associated with Inflammatory Bowel Disease.” *Nature Genetics* 43 (11): 1066–73.
Robinson, Peter N., Sebastian Köhler, Sebastian Bauer, Dominik Seelow, Denise Horn, and Stefan Mundlos. 2008. “The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease.” *American Journal of Human Genetics* 83 (5): 610–15.
Robinson, Peter N., Vida Ravanmehr, Julius O. B. Jacobsen, Daniel Danis, Xingmin Aaron Zhang, Leigh C. Carmody, Michael A. Gargano, et al. 2020. “Interpretable Clinical Genomics with a Likelihood Ratio Paradigm.” *American Journal of Human Genetics* 107 (3): 403–17.
Rolland, Thomas, Murat Taşan, Benoit Charloteaux, Samuel J. Pevzner, Quan Zhong, Nidhi Sahni, Song Yi, et al. 2014. “A Proteome-Scale Map of the Human Interactome Network.” *Cell* 159 (5): 1212–26.
Romagnani, Sergio. 2006. “Immunological Tolerance and Autoimmunity.” *Internal and Emergency Medicine* 1 (3): 187–96.
Sabidussi, Gert. 1966. “The Centrality Index of a Graph.” *Psychometrika* 31 (4): 581–603.
Sakaguchi, Shimon, Tomoyuki Yamaguchi, Takashi Nomura, and Masahiro Ono. 2008. “Regulatory T Cells and Immune Tolerance.” *Cell*. https://doi.org/10.1016/j.cell.2008.05.009.
Sallusto, F., D. Lenig, R. Förster, M. Lipp, and A. Lanzavecchia. 1999. “Two Subsets of Memory T Lymphocytes with Distinct Homing Potentials and Effector Functions.” *Nature* 401 (6754): 708–12.
Sandelin, Albin, Wynand Alkema, Pär Engström, Wyeth W. Wasserman, and Boris Lenhard. 2004. “JASPAR: An Open-access Database for Eukaryotic Transcription Factor Binding Profiles.” *Nucleic Acids Research* 32 (suppl_1): D91–94.
Sarma, J. Vidya, and Peter A. Ward. 2011. “The Complement System.” *Cell and Tissue Research* 343 (1): 227–35.
Savic, Sinisa, Emily A. Caseley, and Michael F. McDermott. 2020. “Moving towards a Systems-Based Classification of Innate Immune-Mediated Diseases.” *Nature Reviews. Rheumatology* 16 (4): 222–37.
Schriml, Lynn Marie, Cesar Arze, Suvarna Nadendla, Yu-Wei Wayne Chang, Mark Mazaitis, Victor Felix, Gang Feng, and Warren Alden Kibbe. 2012. “Disease Ontology: A Backbone for Disease Semantic Integration.” *Nucleic Acids Research* 40 (Database issue): D940–46.
Schubert, Michael, Bertram Klinger, Martina Klünemann, Anja Sieber, Florian Uhlitz, Sascha Sauer, Mathew J. Garnett, Nils Blüthgen, and Julio Saez-Rodriguez. 2018. “Perturbation-Response Genes Reveal Signaling Footprints in Cancer Gene Expression.” *Nature Communications* 9 (1): 20.
Segal, Michael M., Alanna K. Rahm, Nathan C. Hulse, Grant Wood, Janet L. Williams, Lynn Feldman, Gregory J. Moore, et al. 2017. “Experience with Integrating Diagnostic Decision Support Software with Electronic Health Records: Benefits versus Risks of Information Sharing.” *EGEMS (Washington, DC)* 5 (1): 23.
Sergouniotis, Panagiotis I., Emmanuel Maxime, Dorothée Leroux, Annie Olry, Rachel Thompson, Ana Rath, Peter N. Robinson, Hélène Dollfus, and ERN-EYE Ontology Study Group. 2019. “An Ontological Foundation for Ocular Phenotypes and Rare Eye Diseases.” *Orphanet Journal of Rare Diseases* 14 (1): 8.
Sharma, Amitabh, Arda Halu, Julius L. Decano, Megha Padi, Yang-Yu Liu, Rashmi B. Prasad, Joao Fadista, et al. 2018. “Controllability in an Islet Specific Regulatory Network Identifies the Transcriptional Factor NFATC4, Which Regulates Type 2 Diabetes Associated Genes.” *NPJ Systems Biology and Applications* 4 (July): 25.
Sharma, Amitabh, Jörg Menche, C. Chris Huang, Tatiana Ort, Xiaobo Zhou, Maksim Kitsak, Nidhi Sahni, et al. 2015. “A Disease Module in the Interactome Explains Disease Heterogeneity, Drug Response and Captures Novel Pathways and Genes in Asthma.” *Human Molecular Genetics* 24 (11): 3005–20.
Simon, Amos J., Adi Cohen Golan, Atar Lev, Tali Stauber, Ortal Barel, Ido Somekh, Christoph Klein, et al. 2020. “Whole Exome Sequencing (WES) Approach for Diagnosing Primary Immunodeficiencies (PIDs) in a Highly Consanguineous Community.” *Clinical Immunology* 214 (May): 108376.
Smedley, Damian, Julius O. B. Jacobsen, Marten Jäger, Sebastian Köhler, Manuel Holtgrewe, Max Schubach, Enrico Siragusa, et al. 2015. “Next-Generation Diagnostics and Disease-Gene Discovery with the Exomiser.” *Nature Protocols* 10 (12): 2004–15.
Smedley, Damian, Sebastian Köhler, Johanna Christina Czeschik, Joanna Amberger, Carol Bocchini, Ada Hamosh, Julian Veldboer, Tomasz Zemojtel, and Peter N. Robinson. 2014. “Walking the Interactome for Candidate Prioritization in Exome Sequencing Studies of Mendelian Diseases.” *Bioinformatics* 30 (22): 3215–22.
Smith, Cynthia L., Carroll-Ann W. Goldsmith, and Janan T. Eppig. 2005. “The Mammalian Phenotype Ontology as a Tool for Annotating, Analyzing and Comparing Phenotypic Information.” *Genome Biology* 6 (1): R7.
Snel, B., G. Lehmann, P. Bork, and M. A. Huynen. 2000. “STRING: A Web-Server to Retrieve and Display the

Repeatedly Occurring Neighbourhood of a Gene." *Nucleic Acids Research* 28 (18): 3442–44.
Sobreira, Nara L. M., Harindra Arachchi, Orion J. Buske, Jessica X. Chong, Ben Hutton, Julia Foreman, François Schiettecatte, et al. 2017. "Matchmaker Exchange." *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]* 95 (October): 9.31.1–9.31.15.
Sobreira, Nara, François Schiettecatte, David Valle, and Ada Hamosh. 2015. "GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene." *Human Mutation* 36 (10): 928–30.
Spirin, Victor, and Leonid A. Mirny. 2003. "Protein Complexes and Functional Modules in Molecular Networks." *Proceedings of the National Academy of Sciences of the United States of America* 100 (21): 12123–28.
Srivas, Rohith, John Paul Shen, Chih Cheng Yang, Su Ming Sun, Jianfeng Li, Andrew M. Gross, James Jensen, et al. 2016. "A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy." *Molecular Cell* 63 (3): 514–25.
Steinman, Ralph M., Daniel Hawiger, and Michel C. Nussenzweig. 2003. "Tolerogenic Dendritic Cells." *Annual Review of Immunology* 21: 685–711.
Sticht, Carsten, Carolina De La Torre, Alisha Parveen, and Norbert Gretz. 2018. "miRWalk: An Online Resource for Prediction of microRNA Binding Sites." *PloS One* 13 (10): e0206239.
Stokman, Marijn F., Bert van der Zwaag, Nicole C. A. J. van de Kar, Mieke M. van Haelst, Albertien M. van Eerde, Joost W. van der Heijden, Hester Y. Kroes, et al. 2018. "Clinical and Genetic Analyses of a Dutch Cohort of 40 Patients with a Nephronophthisis-Related Ciliopathy." *Pediatric Nephrology* 33 (10): 1701–12.
Stvrtinová, V., J. Jakubovsk\`y, and I. Hulín. 1995. "Inflammation and Fever from Pathophysiology: Principles of Disease. Computing Centre, Slovak Academy of Sciences: Academic Electronic Press."
Sveinbjornsson, Gardar, Anders Albrechtsen, Florian Zink, Sigurjón A. Gudjonsson, Asmundur Oddson, Gísli Másson, Hilma Holm, et al. 2016. "Weighting Sequence Variants Based on Their Annotation Increases Power of Whole-Genome Association Studies." *Nature Genetics* 48 (3): 314–17.
Swainston, Neil, Kieran Smallbone, Hooman Hefzi, Paul D. Dobson, Judy Brewer, Michael Hanscho, Daniel C. Zielinski, et al. 2016. "Recon 2.2: From Reconstruction to Model of Human Metabolism." *Metabolomics: Official Journal of the Metabolomic Society* 12 (June): 109.
Takeuchi, Osamu, and Shizuo Akira. 2010. "Pattern Recognition Receptors and Inflammation." *Cell* 140 (6): 805–20.
Taruscio, D., G. Baynam, H. Cederroth, S. C. Groft, E. W. Klee, K. Kosaki, P. Lasko, et al. 2020. "The Undiagnosed Diseases Network International: Five Years and More!" *Molecular Genetics and Metabolism* 129 (4): 243–54.
Taylor, Rachel L., Neil R. A. Parry, Stephanie J. Barton, Christopher Campbell, Claire M. Delaney, Jamie M. Ellingford, Georgina Hall, et al. 2017. "Panel-Based Clinical Genetic Testing in 85 Children with Inherited Retinal Disease." *Ophthalmology* 124 (7): 985–91.
Thaventhiran, James E. D., Hana Lango Allen, Oliver S. Burren, William Rae, Daniel Greene, Emily Staples, Zinan Zhang, et al. 2020. "Whole-Genome Sequencing of a Sporadic Primary Immunodeficiency Cohort." *Nature* 583 (7814): 90–95.
Theofilopoulos, Argyrios N., Dwight H. Kono, and Roberto Baccala. 2017. "The Multiple Pathways to Autoimmunity." *Nature Immunology* 18 (7): 716–24.
Thiffault, Isabelle, Emily Farrow, Lee Zellmer, Courtney Berrios, Neil Miller, Margaret Gibson, Raymond Caylor, et al. 2019. "Clinical Genome Sequencing in an Unbiased Pediatric Cohort." *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 21 (2): 303–10.
Tong, Amy Hin Yan, Guillaume Lesage, Gary D. Bader, Huiming Ding, Hong Xu, Xiaofeng Xin, James Young, et al. 2004. "Global Mapping of the Yeast Genetic Interaction Network." *Science* 303 (5659): 808–13.
Trujillano, Daniel, Aida M. Bertoli-Avella, Krishna Kumar Kandaswamy, Maximilian Er Weiss, Julia Köster, Anett Marais, Omid Paknia, et al. 2017. "Clinical Exome Sequencing: Results from 2819 Samples Reflecting 1000 Families." *European Journal of Human Genetics: EJHG* 25 (2): 176–82.
Tsokos, George C. 2011. "Systemic Lupus Erythematosus." *The New England Journal of Medicine* 365 (22): 2110–21.
Turesson, Carl, and Eric L. Matteson. 2009. "Clinical Features of Rheumatoid Arthritis: Extra-Articular Manifestations." *Rheumatoid Arthritis*. https://doi.org/10.1016/b978-032305475-1.50014-8.
Vandin, Fabio, Eli Upfal, and Benjamin J. Raphael. 2011. "Algorithms for Detecting Significantly Mutated Pathways in Cancer." *Journal of Computational Biology*. https://doi.org/10.1089/cmb.2010.0265.
Vanunu, Oron. 2009. *Associating Genes and Protein Complexes with Disease Via Network Propagation*. Tel Aviv University.
Vasant, Drashtti, Laetitia Chanas, James Malone, Marc Hanauer, Annie Olry, Simon Jupp, Peter N. Robinson, Helen Parkinson, and Ana Rath. 2014. "Ordo: An Ontology Connecting Rare Disease, Epidemiology and Genetic Data." In *Proceedings of ISMB*. Vol. 30. https://www.researchgate.net/profile/Drashtti_Vasant/publication/281824026_ORDO_An_Ontology_Connecting_Rare_Disease_Epidemiology_and_Genetic_Data/links/55f99bc408aeafc8ac266edf/ORDO-An-Ontology-Connecting-Rare-Disease-Epidemiology-and-Genetic-Data.pdf.
Vasilevsky, Nicole A., Erin D. Foster, Mark E. Engelstad, Leigh Carmody, Matt Might, Chip Chambers, Hugh J. S. Dawkins, et al. 2018. "Plain-Language Medical Vocabulary for Precision Diagnosis." *Nature Genetics* 50 (4): 474–76.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et al. 2001. “The Sequence of the Human Genome.” *Science* 291 (5507): 1304–51.
Vergoulis, Thanasis, Ioannis S. Vlachos, Panagiotis Alexiou, George Georgakilas, Manolis Maragkakis, Martin Reczko, Stefanos Gerangelos, Nectarios Koziris, Theodore Dalamagas, and Artemis G. Hatzigeorgiou. 2012. “TarBase 6.0: Capturing the Exponential Growth of miRNA Targets with Experimental Support.” *Nucleic Acids Research* 40 (Database issue): D222–29.
Vinkšel, Mateja, Karin Writzl, Aleš Maver, and Borut Peterlin. 2021. “Improving Diagnostics of Rare Genetic Diseases with NGS Approaches.” *Journal of Community Genetics*, January. https://doi.org/10.1007/s12687-020-00500-5.
Wakap, Stéphanie Nguengang, Deborah M. Lambert, Annie Olry, Charlotte Rodwell, Charlotte Gueydan, Valérie Lanneau, Daniel Murphy, Yann Le Cam, and Ana Rath. 2019. “Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database.” *European Journal of Human Genetics: EJHG* 28 (2): 165–73.
Wang, Lifeng, Fu-Sheng Wang, and M. Eric Gershwin. 2015. “Human Autoimmune Diseases: A Comprehensive Update.” *Journal of Internal Medicine*. https://doi.org/10.1111/joim.12395.
Wang, Li-Xin, Yang Li, and Guan-Zhi Chen. 2018. “Network-Based Co-Expression Analysis for Exploring the Potential Diagnostic Biomarkers of Metastatic Melanoma.” *PloS One* 13 (1): e0190447.
Wang, Xiao-Dong, Jia-Liang Huang, Lun Yang, Dong-Qing Wei, Ying-Xin Qi, and Zong-Lai Jiang. 2014. “Identification of Human Disease Genes from Interactome Network Using Graphlet Interaction.” *PloS One* 9 (1): e86142.
Wang, Xiujuan, Natali Gulbahce, and Haiyuan Yu. 2011. “Network-Based Methods for Human Disease Gene Prediction.” *Briefings in Functional Genomics* 10 (5): 280–93.
Weinberg, Andrew. 2006. “Faculty Opinions Recommendation of Reciprocal Developmental Pathways for the Generation of Pathogenic Effector TH17 and Regulatory T Cells.” *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. https://doi.org/10.3410/f.1032342.374969.
Weinreich, Steffanie S., R. Mangon, J. J. Sikkens, M. E. Teeuw, and M. C. Cornel. 2008. “Orphanet: A European Database for Rare Diseases.” *Nederlands Tijdschrift Voor Geneeskunde* 152 (9): 518–19.
Weirauch, Matthew T. 2011. “Gene Coexpression Networks for the Analysis of DNA Microarray Data.” In *Applied Statistics for Network Biology*, 215–50. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.
Wray, Naomi R., Cisca Wijmenga, Patrick F. Sullivan, Jian Yang, and Peter M. Visscher. 2018. “Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model.” *Cell* 173 (7): 1573–80.
Xiao, Feifei, Zhixiang Zuo, Guoshuai Cai, Shuli Kang, Xiaolian Gao, and Tongbin Li. 2008. “miRecords: An Integrated Resource for microRNA–target Interactions.” *Nucleic Acids Research* 37 (suppl_1): D105–10.
Xue, Hansheng, Jiajie Peng, and Xuequn Shang. 2019. “Predicting Disease-Related Phenotypes Using an Integrated Phenotype Similarity Measurement Based on HPO.” *BMC Systems Biology* 13 (Suppl 2): 34.
Yan, Changhui, Drena Dobbs, and Vasant Honavar. 2004. “A Two-Stage Classifier for Identification of Protein–protein Interface Residues.” *Bioinformatics* 20 (suppl_1): i371–78.
Yeger-Lotem, Esti, and Roded Sharan. 2015. “Human Protein Interaction Networks across Tissues and Diseases.” *Frontiers in Genetics* 6 (August): 257.
Yildirim, Muhammed A., Kwang-Il Goh, Michael E. Cusick, Albert-László Barabási, and Marc Vidal. 2007. “Drug-Target Network.” *Nature Biotechnology* 25 (10): 1119–26.
Zelezniak, Aleksej, Tune H. Pers, Simão Soares, Mary Elizabeth Patti, and Kiran Raosaheb Patil. 2010. “Metabolic Network Topology Reveals Transcriptional Regulatory Signatures of Type 2 Diabetes.” *PLoS Computational Biology* 6 (4): e1000729.
Zhang, Bin, Chris Gaiteri, Liviu-Gabriel Bodea, Zhi Wang, Joshua McElwee, Alexei A. Podtelezhnikov, Chunsheng Zhang, et al. 2013. “Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer’s Disease.” *Cell* 153 (3): 707–20.
Zhang, Bin, and Steve Horvath. 2005. “A General Framework for Weighted Gene Co-Expression Network Analysis.” *Statistical Applications in Genetics and Molecular Biology* 4 (August): Article17.
Zhang, Fan, Bo Gao, Liangde Xu, Chunquan Li, Dapeng Hao, Shaojun Zhang, Meng Zhou, et al. 2013. “Allele-Specific Behavior of Molecular Networks: Understanding Small-Molecule Drug Response in Yeast.” *PloS One* 8 (1): e53581.
Zhang, Jie, Yang Xiang, Liya Ding, Kristin Keen-Circle, Tara B. Borlawsky, Hatice Gulcin Ozer, Ruoming Jin, Philip Payne, and Kun Huang. 2010. “Using Gene Co-Expression Network Analysis to Predict Biomarkers for Chronic Lymphocytic Leukemia.” *BMC Bioinformatics* 11 Suppl 9 (October): S5.
Zhang, Shu, Weixia Liu, Xiaoyan Liu, Jiaxin Qi, and Chunmei Deng. 2017. “Biomarkers Identification for Acute Myocardial Infarction Detection via Weighted Gene Co-Expression Network Analysis.” *Medicine* 96 (47): e8375.
Zhou, Xuezhong, Jörg Menche, Albert-László Barabási, and Amitabh Sharma. 2014. “Human Symptoms–disease Network.” *Nature Communications* 5 (1): 1–10.
Zhu, Qian, Hongfang Liu, Christopher G. Chute, and Matthew Ferber. 2015. “EHR Based Genetic Testing Knowledge Base (iGTKB) Development.” *BMC Medical Informatics and Decision Making* 15 Suppl 4 (November): S3.
Zhu, Qian, Dac-Trung Nguyen, Ivan Grishagin, Noel Southall, Eric Sid, and Anne Pariser. 2020. “An Integrative Knowledge Graph for Rare Diseases, Derived from the Genetic and Rare Diseases Information Center (GARD).” *Journal of Biomedical Semantics*.

Curriculum vitae

Julia Pazmandi pazmandi.julia@gmail.com
D.O.B.: 7th January 1990 ORCID ID: 0000-0003-3411-9598

Current position

2015 - present Predoctoral Fellow
Ludwig Boltzmann Institute of Rare and Undiagnosed diseases and CeMM Research Center for Molecular Medicine
Group of Kaan Boztug, Group of Jörg Menche

Education

2013 - 2015 Masters in Molecular Medicine, Uppsala University, Sweden
2014 - 2015 Exchange semester (Erasmus), Imperial College London, United Kingdom
2009 - 2013 Bachelor in Molecular Bionics, Pazmany Peter Cathlic University, Hungary

2003 - 2009 High School, Szerb Antal High School, Hungary

Skills and competences

Computational

Programming Python, R, Matlab
Data science Statistical analysis, network theory

Molecular biology

Techniques PCR, qPCR, Wester blotting, molecular cloning, culture of adherent and non-adherent cells, mRNA and protein extraction

Languages

Hungarian Native speaker
English Full professional fluency
German Professional fluency

Honours, Awards and Scholarships

2019 Falling Walls Austria Winner: Breaking the Wall of Science and Fiction
2015 Phd fellowship (CeMM)
2014 Erasmus Scholarship (Imperial College London)

Work Experience

2015 - present Systems biology methods towards rare autoimmune and autoinflammatory diseases
Revision and expansion of the Human Phenotype Ontology

	Development of VR-based methods for network analysis CeMM and LBI-RUD, Austria Group of Kaan Boztug, Group of Jörg Menche
2014 - 2015	Epigenetics of the ketogenic diet as an adjuvant therapy in glioblastoma Multiforme Imperial College London Laboratory of Nelofer Syed
2011 - 2013	Genetics of late-onset congenital adrenal hypoplasia Semmelweis Hospital Budapest Laboratory of Attila Patocs

Leadership Experience

2018 - present	Human Phenotype Ontology (HPO) for Immune Mediated Diseases Consortium, Co-organizer
2020	3nd Workshop on HPO for Immune Mediated Diseases Consortium
2019	2nd Workshop on HPO for Immune Mediated Diseases Consortium
2018	1st Workshop on HPO for Immune Mediated Diseases Consortium
2015 - present	Supervision of master and rotation students

Publications

1. **Julia Pazmandi**, Sevgi Köstel Bal, Felix Müller, Celine Sin, Christiane V. R. Hütter, Jörg Menche*#, Kaan Boztug*#. 2021. AutoCore: network-based identification of a core module defining human autoimmunity and autoinflammation. * these authors contributed equally, # to whom correspondence should be addressed.
2. Sebastian Pirch, Felix Müller, Eugenia Iofinova, **Julia Pazmandi**, Christiane V. R. Hütter, Martin Chiettini, Celine Sin, Kaan Boztug, Iana Podkosova, Hannes Kaufmann & Jörg Menche. The VRNetzer platform enables interactive network analysis in Virtual Reality. *Nat. Commun.* **12**, 2432 (2021).
3. Matthias Haimel PhD*, **Julia Pazmandi MSc***, Raúl Jiménez Heredia MSc, Jasmin Dmytrus PhD, Sevgi Köstel Bal M.D.,PhD, Samaneh Zoghi PhD, Paul van Daele M.D., Tracy A. Briggs PhD, Carine Wouters M.D., Brigitte Bader-Meunier M.D., Florence A. Aeschlimann M.D., Roberta Caorsi M.D., Despina Eleftheriou M.D., Esther Hoppenreijs M.D., Elisabeth Salzer M.D.,PhD, Shahrzad Bakhtiar M.D., Beata Derfalvi M.D., Francesco Saettini M.D., Maaike A. A. Kusters M.D.,PhD , Reem Elfeky M.D., Johannes Trück M.D., Jacques G. Rivière M.D., Mirjam van der Burg PhD , Marco Gattorno M.D., Markus G. Seidel M.D., Siobhan Burns M.D., Klaus Warnatz M.D., Fabian Hauck M.D.,PhD , Paul Brogan M.D., Kimberly C. Gilmour PhD, Catharina Schuetz M.D., Anna Simon M.D.,PhD, Christoph Bock PhD , Sophie Hambleton PhD , Esther de Vries PhD , Peter Robinson M.D., Marielle van Gijn PhD †#, Kaan Boztug M.D.†#. 2021. “Curation and Expansion of Human Phenotype Ontology for Inborn Errors of Immunity”. *JACI* * and †, these authors contributed equally # to whom correspondence should be addressed. 2021.
4. Artem Kalinichenko , Giovanna Perinetti Casoni , Loïc Dupré , Luca Trotta , Jakob Huemer , Donatella Galgano , Yolla German , Ben Haladik , **Julia Pazmandi** , Marini Thian , Özlem Yüce Petronczki , Samuel C. Chiang , Mervi Taskinen , Anne Hekkala , Saila Kauppila , Outi Lindgren , Terhi Tapiainen , Michael J. Kraakman , Kim Vettenranta , Alexis J. Lomakin , Janna Saarela , Mikko R. J. Seppänen , Yenan T. Bryceson , Kaan Boztug. RhoG deficiency abrogates cytotoxicity of human lymphocytes and causes hemophagocytic lymphohistiocytosis. *Blood* **137**, 2033–2045 (2021).

5. Sebastian Köhler, Michael Gargano, Nicolas Matentzoglu, Leigh C Carmody, David Lewis-Smith, Nicole A Vasilevsky, Daniel Danis, Ganna Balagura, Gareth Baynam, Amy M Brower, Tiffany J Callahan, Christopher G Chute, Johanna L Est, Peter D Galer, Shiva Ganesan, Matthias Griese, Matthias Haimel, **Julia Pazmandi**, Marc Hanauer, Nomi L Harris, Michael J Hartnett, Maximilian Hastreiter, Fabian Hauck, Yongqun He, Tim Jeske, Hugh Kearney, Gerhard Kindle, Christoph Klein, Katrin Knoflach, Roland Krause, David Lagorce, Julie A McMurry, Jillian A Miller, Monica C Munoz-Torres, Rebecca L Peters, Christina K Rapp, Ana M Rath, Shahmir A Rind, Avi Z Rosenberg, Michael M Segal, Markus G Seidel, Damian Smedley, Tomer Talmy, Yarlalu Thomas, Samuel A Wiafe, Julie Xian, Zafer Yüksel, Ingo Helbig, Christopher J Mungall, Melissa A Haendel, Peter N Robinson. The Human Phenotype Ontology in 2021 *Nucleic Acids Research*, Volume 49, Issue D1, 8 January 2021, Pages D1207–D1217, https://doi.org/10.1093/nar/gkaa1043
6. Lukas M Gasteiger, Peter N Robinson, **Julia Pazmandi**, Kaan Boztug, Mikko R J Seppänen, Markus G Seidel, Registry Working Party of the European Society for Immunodeficiencies (ESID). Supplementation of the ESID registry working definitions for the clinical diagnosis of inborn errors of immunity with encoded human phenotype ontology (HPO) terms. J Allergy Clin Immunol Pract 2020 May;8(5):1778. doi: 10.1016/j.jaip.2020.02.019.
7. Köstel Bal, S., **Pazmandi, J.**, Boztug, K. & Özen, S. Rheumatological manifestations in inborn errors of immunity. Pediatr. Res. 87, 293–299 (2020).
8. Calzoni E, Platt CD, Keles S, Kuehn HS, Beaussant-Cohen S, Zhang Y, **Pazmandi J**, Lanzi G, Pala F, Tahiat A, Artac H, Heredia RJ, Dmytrus J, Reisli I, Uygun V, Uygun D, Bingol A, Basaran E, Djenouhat K, Benhalla N, Bendahmane C, Emiroglu M, Kirchhausen T, Pasham M, Jones J, Wallace JG, Zheng L, Boisson B, Porta F, Rosenzweig SD, Su H, Giliani S, Lenardo M, Geha RS, Boztug K, Chou J, Notarangelo LD. F-BAR domain only protein 1 (FCHO1) deficiency is a novel cause of combined immune deficiency in human subjects. J Allergy Clin Immunol. 2019 Jun;143(6):2317-2321.e12. doi: 10.1016/j.jaci.2019.02.014. Epub 2019 Feb 26.
9. **Pazmandi J**, Kalinichenko A, Ardy RC, Boztug K. Early-onset inflammatory bowel disease as a model disease to identify key regulators of immune homeostasis mechanisms. Immunol Rev. 2019 Jan;287(1):162-185. doi: 10.1111/imr.12726. Review.
10. Mödinger Y, Rapp A, **Pazmandi J**, Vikman A, Holzmann K, Haffner-Luntzer M, Huber-Lang M, Ignatius A. C5aR1 interacts with TLR2 in osteoblasts and stimulates the osteoclast-inducing chemokine CXCL10. J Cell Mol Med. 2018 Dec;22(12):6002-6014. doi: 10.1111/jcmm.13873. Epub 2018 Sep 24.
11. Doleschall M, Szabó JA, **Pázmándi J**, Szilágyi Á, Koncz K, Farkas H, Tóth M, Igaz P, Gláz E, Prohászka Z, Korbonits M, Rácz K, Füst G, Patócs A. Common genetic variants of the human steroid 21-hydroxylase gene (CYP21A2) are related to differences in circulating hormone levels. PLoS One. 2014 Sep 11;9(9):e107244. doi: 10.1371/journal.pone.0107244. eCollection 2014

*these authors contributed equally

Communications

Conference talks

1. **Pazmandi J*,** Haimel M* et al. **"**Human Phenotype Ontology for Immune Mediated Disorders**".** 2020 Meeting of the European Society of Immunodeficiencies, online conference.
2. **Pazmandi J,** et al. "IBDomics: network analysis of inflammatory bowel disease reveals novel monogenic causes of intestinal inflammation". NetSci conference, 2020, online conference.

3. **Pazmandi J*,** Haimel M* et al. **"**Human Phenotype Ontology Immune Mediated Disorders Consortium**"** 2019 Meeting of the European Society of Immunodeficiencies, Brussels, Belgium.
4. **Pazmandi J**, Menche J , Boztug K. "Systems-analysis of DNA repair defects in primary immunodeficiencies and cancer". From Lab to Life – Childhood Cancer Research Initiatives 2018, Vienna, Austria.
5. **Pazmandi J**, Menche J , Boztug K. "A systems view of autoimmunity and autoinflammation". 2017 Meeting of the European Society of Immunodeficiencies, Edinburgh, United Kingdom.
6. **Pazmandi J** et al. "Finding the needle in the haystack: identifying causal variants with systems biology". 2017 Meeting of the European Society of Immunodeficiencies, Edinburgh, United Kingdom.

Conference posters

1. **Pazmandi J**, Menche J , Boztug K. "Systems based analysis of inflammatory bowel disease". 2019 ESPGHAN, Glasgow, United Kingdom.
2. **Pazmandi J**, Menche J , Boztug K. "Systems based analysis of monogenic and polygenic autoimmune and autoinflammatory diseases ". Genomics of Rare Diseases, 2019, Cambridge, United Kingdom.
3. **Pazmandi J*,** Haimel M* et al. "Human Phenotype Ontology Immune Mediated Disorders Consortium". Immunogenomics of Disease: Accelerating to Patient Benefit, 2019, Cambridge, United Kingdom.
4. **Pazmandi J**, Menche J , Boztug K. "Identifying novel causal genes in inflammatory bowel disease". 13th Congress of ECCO: European Crohn's and Colitis Organisation 2018, Vienna, Austria.
5. **Pazmandi J**, Menche J , Boztug K. "An interdisciplinary approach to identify novel causal genes in rare diseases" 2016 LBG Meeting for Health Sciences, Vienna, Austria.
6. **Pazmandi J** et al. "An interdisciplinary approach identifying novel causal genes in rare diseases". MLPM Symposium Barcelona 2016, Barcelona, Spain.
7. **Pazmandi J** et al. "An interdisciplinary approach identifying novel causal genes in rare diseases". 10th Annual Young Scientist Association Symposium of the Medical University of Vienna 2016, Vienna, Austria.