

## Title

*Translational bioinformatics and machine learning framework for biomarker discovery, disease prediction, and patient profiling for precision medicine*

## Running Title

*Multi-omics driven ML for predictive medicine*

## Authors

Zeeshan Ahmed<sup>1,2 \*</sup>, Prithvi Govindareddy<sup>2</sup>, William DeGroat<sup>2</sup>, Rishabh Narayanan<sup>2</sup>, Elizabeth Peker<sup>2</sup>, Saman Zeeshan<sup>3</sup>

## Affiliations

1. Department of Medicine, Robert Wood Johnson Medical School, Rutgers Health, 125 Paterson St, New Brunswick, NJ, 08901, USA.
2. Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers Health, 112 Paterson St, New Brunswick, 08901, NJ, USA.
3. Department of Biomedical and Health Informatics, UMKC School of Medicine, 2411 Holmes Street, Kansas City, 64108, MO, USA.

**\*Corresponding author:** Zeeshan Ahmed, Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers University, 112 Paterson Street, New Brunswick, 08901, NJ, USA. (zahmed@ifh.rutgers.edu).

## Abstract

Precision medicine aims to advance our ability from a "one-size-fits-all" approach to personalized and predictive healthcare across diverse populations. It promotes integration of multi-omics and phenotypic data to understand disease mechanisms and discover novel biomarkers and risk factors, which could be used to predict and prevent critical diseases in individual patients across diverse populations. The potential implications of precision medicine approach can accelerate our ability to classify patients at higher risk of developing critical diseases, improve diagnostic capabilities, develop deeper understanding of individual risk, investigate racial differences and demographic characteristics, and find relationships between genetic variants, expressions, and diseases. This study focuses on implementing an innovative and data driven framework of translational bioinformatics and Machine Learning (ML) techniques to analyze multi-omics, including RNA-seq and Whole-Genome Sequencing (WGS) data, generated using blood samples of randomly consented patients. First, we utilized bioinformatics pipelines to identify differentially expressed genes and their pathogenic and likely pathogenic variants for the downstream data analysis, annotation, and visualization. Then, applied a nexus of ML models for multi-omics biomarker discovery, disease prediction, density-based clustering, single-patient profiling, and pathogenicity classification. WGS data analysis supported the exploration of genetic variation and diversity among patients to identify known and novel biomarkers, whereas RNA-seq data analysis improved our understanding of functional and biological pathways that underlying disease states. We classified and clustered pathogenic variants and expressions across various genes and discovered numerous diseases leading risk factors. Our results include gene-disease associations and captured common pathways across the broader population, demonstrating a level of sensitivity and accuracy that has broad clinical implications. We validated our results through clinical records, and state of the science literature. This study delves into the strengths of multi-omics data integration and capabilities of ML application in genetically diverse and complex patient cohorts. Our approach has the potential to elucidate complex gene-disease interactions for genetically diverse populations, which can support earlier diagnoses for patients in many disease realms.

## Keywords:

Multi-omics, Genomics, Bioinformatics, Machine Learning, Biomarker, Disease Prediction, Patient Profile, Precision medicine

## Introduction

The prevalence of diagnostic errors in medicine result in substantial but preventable harm across various medical fields [1, 2, 3, 4]. This is largely because most modern medicine depends on the treatment of generic symptoms and empirical evidence from clinical trials rather than focusing on the personalized care tailored to unique biological composition of individual patients [5, 6]. Critically, reducing the rate of misdiagnosis will ultimately improve patient outcomes, increase available treatment options, and address growing costs. The field of precision medicine has gained increasing attention for its potential to transform patient care through individualized diagnosis and treatment strategies. To determine key biological indicators that distinguish healthy and diseased patients, it is imperative to comprehend the particularities and unique aspects of each patient [7]. In aspiring to improve diagnostic accuracy by understanding the intricacies of human biology, scientists frequently explore various types of omics, which include e.g., genomics, transcriptomics, epigenomics, proteomics, microbiomics, metabolomics, and multimodal integrations of those [8, 9, 10, 11]. In clinical settings, the timely identification of significant patterns among millions of these multi-omics and clinical features is paramount for discovering underlying biological pathways and modifiable risk factors to support early detection and prevention of complex and rare disorders e.g., neural, respiratory, cardiovascular, immunological, infectious, cancer and skeletal [12]. However, individual omics modalities alone are insufficient to capture the full complexity of biological systems [13]. The effective integration of multi-omics with phenotypic data has the potential to generate holistic multimodal patient profiles, necessary to underscore key characteristics and improve prognosis, support early diagnosis, and provide more economical personalized treatment [7].

Traditional bioinformatics and statistical modelling have long been the default choice for analyzing and interpreting big data [14]. Most importantly, Genome-wide Association Studies (GWAS) and Gene Expression Analysis (GEA) have remarkably assisted in understanding the genetic basis of human disease by uncovering millions of loci associated with various complex phenotypes [15, 16, 17]. Whole-Genome Sequencing (WGS) data analysis supported the exploration of genetic variation and diversity among patients to identify known and novel biomarkers [18], whereas RNA-seq data analysis improved our understanding of functional and biological pathways that underlie disease states [19]. Although these approaches provide important insights into multifactorial diseases, they have limitations when used individually. RNA-seq based gene expression analysis on its own fails to capture the effects of non-coding genes, whereas linking WGS variant data to specific diseases remains challenging [19]. Most genetic studies focus on a targeted set of disease-specific genes, rather than an unbiased investigation of the entire

human genome and transcriptome, limiting the applicability and generalizability of the results in clinical settings. Furthermore, these are unable to predict disease and detect all the heritability explained by Single Nucleotide Polymorphisms (SNPs) and can only target specific variants of complex diseases [18]. The unbiased integration of comprehensive multi-omics data can provide a deeper view into causal relationships between genetic variants, expressions, and complex diseases. However, it remains persistently difficult due to the inherent heterogeneity and large volumes of multi-omics data [20]. Progress in the development of translational bioinformatic approaches has significantly enabled robust data processing and investigation of high-throughput omics data [20]. However, due to their inability to integrate heterogeneous data types, these tools have historically underperformed in identifying multi-faceted interactions in high-dimensional multi-omics data [21].

The cutting-edge Artificial Intelligence (AI) and Machine Learning (ML) approaches have proven effective at uncovering elucidative knowledge on disease-causing biomarkers and the biological underpinnings of a plethora of human diseases [7]. The analysis of integrated multi-omics and phenotypic data with a nexus of translational bioinformatics and ML techniques has the potential to accurately predict disease risk in patients by identifying patterns that may otherwise be eluded by the traditional bioinformatics approaches [22]. Furthermore, the current state-of-science ML techniques typically involve employing single-model supervised ML on pre-labelled case/control cohorts, have proven less effective at disease prediction, especially when using multi-omics data [21, 23]. In response, we developed multiple ML pipelines and applications that support flexible and ensemble-based analytics frameworks, capable of predicting diseases with higher accuracy in a case/control cohort [23, 24, 25, 26, 27]. However, to achieve unbiased and effective personalized treatment, it is imperative to equally explore and develop unsupervised learning techniques on patients whose disease states are unknown a priori. The current unsupervised ML methods remain substantially less developed than supervised approaches, often requiring extensive amounts of prelabelled data [28, 29]. Since unsupervised techniques involve finding associations among data points to describe higher-level patterns, the importance of unbiased representation within a population is paramount [30].

In this study, we implemented a FAIR (Findable, Accessible, Interoperable, Reusable) and data driven nexus of translational bioinformatics and ML techniques to analyze unlabeled multi-omics data from a cohort of randomly consented patients. Random participant selection rather than disease-specific targeted recruitment was planned and helped preserve genetic diversity that may better guide unsupervised clustering and discriminatory ML models. Genomic and transcriptomic data generated from the collected

blood sample of each patient were independently analyzed and reintegrated to identify predictive biomarkers and associated disease states. Nevertheless, this integration remains challenging due to the heterogeneity and scale of multi-omics data.

## Results

The reported results are achieved by analyzing multi-omics data (RNA-seq and WGS, generated from blood samples) of randomly consented patients (n=96) with the implementation of designed translational bioinformatics and data driven ML framework, explained in the methods section (Figure 1). These results include classified, clustered, and integrated gene expressions and variants, annotated gene-disease associations, and common pathways to understand relationships between disease causing multi-omics biomarkers and clinical outcomes. Furthermore, we produced single-patient profiles based on the discovered biomarkers and associated diseases, and presented overall disease distribution across patient cohort, including the most commonly identified biomarkers and associated diseases.

### ***Gene expression analysis-based results***

Gene expression analysis revealed 11,235 protein-coding genes from 56,861 transcripts with a mean TPM greater than 10 thresholds. Expression was visualized by generating heatmap across the cohort (Figure 2A, 2B, and 2C). Using a nested cross-validation scheme, our calibrated disease prediction pipeline achieved a mean F1 score ( $pos\ label = 0$ ) of 0.672 (95% CI = 0.580, 0.752) and an MCC of 0.649 (95% CI = 0.550, 0.738). The model attained a balanced accuracy of 0.810 (95% CI = 0.755, 0.862) and an overall accuracy of 0.912 (95% CI = 0.892, 0.933). We explored how classification thresholds affect the model's performance and varied the decision threshold from 0.1 to 0.9. F1 scores ranged from 0.190 to 0.700 and balanced accuracy from 0.551 to 0.844, illustrating the trade-off between precision and recall at different probabilities. A threshold of 0.70 produced the highest F1 (0.700) without sacrificing balanced accuracy (0.844). This was used to make "confident" positive predictions. We utilized the mRMR algorithm with an MIQ criterion to select genes (n=15) representative of the Cardiovascular Disease (CVD) classification: *GPSM3*, *RPS15*, *GSTO1*, *GSTK1*, *GPX4*, *RPS11*, *RPLP2*, *RPS14*, *RPL37A*, *GMFG*, *RPS18*, *RPS19BP1*, *GABARAP*, *RPS10*, and *RPS21*. Before applying the model, outlier samples (SID = 666, 675, 693, 694, 698, 711, 731, 732, and 743) were removed based on PCA criterion of two standard deviations from the center of the distribution. With that, we predicted occurrence of CVD in 30 patients.

UMAP, followed by HDBSCAN clustering, identified ten clusters in the pseudo-labeled dataset, along with a set of outlier samples that did not fit any cluster's criteria. Clusters 1, 2, and 3 overlapped with the patients predicted to have CVD (Figure 2D, 2E), suggesting a clear separation of disease-associated

transcriptional profiles. DE and subsequent enrichment analysis indicated that three of the clusters (1, 6, and 7) were particularly enriched for disease-associated biological pathways. Using our individualized KNN-weight reference approach, each patient was compared against a smooth, population-level baseline and ranked according to the twenty genes with the largest absolute LFC values. Patient-specific enrichment analysis of these twenty genes against GWAS catalog, ClinVar, DisGeNET, and Online Mendelian Inheritance in Man (OMIM) Disease highlighted recurrent disease-relevant terms and pathways across the cohort. *CD177* appeared frequently among the top genes (Figure 2F). In GWAS catalog, the most identified terms were labelled as “Abdominal Aortic Calcification Levels”, “*CD177* Antigen Levels”, and “Response to SSRI (Symptom Remission)” (Figure 2G). OMIM disease annotations repeatedly included “blood”, “ichthyosis”, and “lymphoma” (Figure 2H). DisGeNET terms focused on “Myeloid Leukemia, Chronic”, “Thalassemia”, and “Inflammation” (Figure 2I). Finally, ClinVar analysis frequently identified “Diamond-Blackfan anemia”, “severe congenital neutropenia”, and “susceptibility to malaria” (Figure 2J). Collectively, these results underscore the capacity of this pipeline to reveal individualized disease associations while also capturing common pathways across the broader population.

### ***Gene variant analysis-based results***

Mutation profiling revealed 43,451,462 from a total of 396,788,923 variants across all 96 patients of our cohort for the top 5000 DEG derived from RNA-seq expression analysis. Annotations were sourced from dbSNP and ClinVar to provide crucial metadata for each variant, including known RS numbers, clinical significance, and associated clinical disease names. Further preprocessing was performed to identify 972,823 variants with known RS numbers, constituting approximately 0.25% of our original dataset and serving as the basis for subsequent pathogenicity classification, clustering, and visualization. The identification of pathogenic, likely pathogenic, and risk-factor variants was supported through ClinVar disease annotations. A large majority of the filtered variants were marked benign or of unknown pathogenicity and consequently pruned. Only 1,747 total variants (0.18% of the filtered set) across 84 genes were identified as pathogenic or known risk factors for diseases, with between 10 and 26 variants recalled for each patient in our cohort. Normalized counts for these 84 genes are displayed on a logarithmic scale in Figure 3A. Gene disease associations were identified by cross-referencing gene information from dbSNP with disease metadata from ClinVar. We discovered 241 different diseases associated with these 1,747 pathogenic variants. Some of the most common diseases associated with patients in our cohort included myocardial infarction, tuberculosis, malaria, bacteremia, asthma, and diabetes. Figure 3B visualizes the distribution of all 241 identified diseases, while Figure 3C depicts the 50

most frequent diseases among all participants. Approximately 37% (88 diseases) of all 241 identified diseases were predicted for only one patient, while only 20% (49 diseases) were predicted in at least 10 patients. This relative sparsity of shared variants and diseases is attributable to the high genetic diversity of our random cohort. Agglomerative hierarchical clustering was performed using a matrix of variant prevalence to identify 10 distinct clusters of individuals. Visualization of the high-dimensional data necessitates lower-dimensional embedding, which was enabled through a 2D UMAP projection, as presented in Figure 3D. Separate comparative gene-disease networks for the top 50 diseases identified through WGS analysis as well as ClinVar-based diseases identified from RNA-seq analysis are presented in Figure 3E and 3F, respectively.

### ***Patient profiling and disease distribution across cohort***

Notably, the clusters remained isolated with minimal overlap between groups. Adjacent data points are generally indicative of greater similarity between the corresponding individuals, offering greater comparability among otherwise unrelated individuals. Upon identifying gene expressions, pathogenic variants, and predicted clusters, we constructed individual profiles for each patient, including relevant biomarkers and associated diseases (Table 1 and Figure 4). Key findings include most common, rare, and complex diseases across the patient cohort (n=96, patients), divided into seventeen out of eighteen targeted disease categories described in the methods section (Figure 4A and 4B). Our analysis revealed hematologic and immune-related disorders were the most common findings, including blood and blood-forming organ diseases (81/96, patients) and immune system diseases (52/96, patients). Additionally, most influencing diseases are related to nervous system (63/96, patients); genitourinary system (51/96, patients); endocrine, nutritional and metabolic (47/96); and congenital and genetic (44/96, patients) disorders. Diseases observed with moderate occurrence in the overall cohort were related to the skin and subcutaneous tissue (36/96, patients); musculoskeletal (32/96, patients); respiratory system (32/96, patients); viral/bacterial/infectious (27/96, patients); circulatory system and CVD (20/96, patients); mental/behavioral/neurodevelopmental (18/96, patients); eye and adnexa (16/96, patients); digestive system (14/96, patients); and ear, nose and throat (13/96, patients) diseases. We discovered cancer (10/96, patients) among the complex and rare disorders, and observed some abnormal clinical findings (34/96, patients) as well (Figure 4C). The overall cohort demonstrated presence of multisystem comorbidities, excluding injuries and wounds.

## Discussion

This study delves into the application of translational bioinformatics and ML application with multi-omics integration, using a genetically diverse and complex patient cohort. Through constructed individual profiles based on gene expressions, pathogenic variants, and associated diseases, we identified eight mutually predictive genes: *ACKR1*, *C4B*, *IFI44L*, *IFITM3*, *LGALS2*, *RETN*, *RHD*, and *TNFRSF13B*. Next, we applied VAREANT [20] (integrating public annotation databases dbSNP and ClinVar) to identify gene-disease relationships among 148 diseases across RNA-based and WGS cohorts. After manual review and removal of redundancies, 134 distinct diseases were retained. During gene-disease annotation, we realized that there could be a possibility of missing data among annotation databases. Therefore, we conducted an extensive literature review using state-of-the-science studies available from PubMed Central and were able to discover 389 additional diseases associated with these eight genes, which were missing in the annotation databases. Combined, using annotation databases (134 diseases) and scientific literature review (389 diseases), we identified 523 distinct gene-disease associations (Table 2). To better understand the complexities and relationships, we constructed gene-disease networks of all diseases identified through annotation databases (Figure 5A), scientific literature review (Figure 5B), combined annotations from both (Figure 5C), and disease associations with more than one gene only (Figure 5D). Furthermore, we identified that 74 diseases were associated with at least two and at up to seven of the eight mutually predictive genes. These figures reflect original annotation outputs prior to the redundancy removal, therefore include 537 total diseases associated with eight genes.

*ACKR1* (atypical chemokine receptor 1) was found to be one of the eight common genes among patient cohort. It is part of the Duffy blood group and encodes glycosylated membrane receptor protein for malarial parasites *Plasmodium vivax* and *Plasmodium knowlesi* [31]. A total of sixty-four diseases were found to be associated with *ACKR1* (twenty-five from annotation databases; and thirty-nine from scientific literature review). Major disease categories associated with this gene include cancer (breast [32] and prostate [33]); CVD and circulatory system disorders (atherosclerosis [34], coronary artery disease [35], and ascending thoracic aortic aneurysm [36]); blood and blood-forming organs diseases (sickle cell anemia [37]); and viral, bacterial, and infectious diseases (malarial disease [38], resistance to *Plasmodium vivax* infection [39], hepatitis C virus [40], and Human Immunodeficiency Virus-1 (HIV-1) [41]). Furthermore, we discovered its binding with other disorders, including musculoskeletal/connective tissue, endocrine/metabolic, neurodevelopmental/nervous system, skin, digestive, respiratory, and rare diseases such as non-infectious posterior uveitis [42] and preeclampsia [43]. Second most common gene among

the patient cohort is *C4B* (complement C4B). It encodes complement factor 4, a key effector for adaptive and innate humoral immunity [44]. A total of ninety-nine diseases were found to be associated with *C4B* (sixteen from annotation databases and eighty-three from scientific literature reviews), which mainly include neurodevelopmental/mental/behavioral and nervous system diseases (schizophrenia [45], attention deficit hyperactivity disorder [46], multiple sclerosis [47], autism [48], Parkinson's disease [49], and Alzheimer's [50]); CVD/circulatory systems (myocardial infarction [51]); musculoskeletal system/connective tissue diseases (Systemic Lupus Erythematosus (SLE) [52], Felty's syndrome [53], juvenile dermatomyositis [54], and rheumatoid arthritis [55]); and endocrine/metabolic disorders (type 1 diabetes mellitus [56], Graves' disease [57], and congenital adrenal hyperplasia [58]). In addition, we discovered its relationship with diseases e.g., infectious, notably pertussis (whooping cough) [59], cancer (lung [60] and colorectal [61]), six blood and blood-forming organ-related, digestive system and teeth, eye and adnexa, genitourinary system, respiratory system, and miscellaneous diseases such as leprosy [62], Kawasaki disease [63], and amyotrophic lateral sclerosis [64].

*IFI44L* (interferon induced protein 44) is the third most common gene found in the overall patient cohort. It is a protein coding gene that has vast implications in immune response [65] and found to be related to seventy-one diseases (seven from annotation databases, and sixty-four from literature review), which include musculoskeletal/connective tissue disorders (Sjogren's syndrome [66], juvenile dermatomyositis [67], rheumatoid arthritis [68], and SLE [69]); infectious diseases (HIV [70], COVID-19 [71], and tuberculosis [72]); cancer disorders (hepatocellular carcinoma where *IFI44L* plays a role in cancer stemness, metastasis, and drug resistance [73]); and endocrine/metabolic disorders (Aicardi-Goutieres Syndrome [74]). Additional categories included neurological, skin, blood, CVD/circulatory system, pregnancy/puerperium related, respiratory and rare antiphospholipid syndrome (an autoimmune disorder found in the blood disease category, prominently associated with increased *IFI44L* levels [75]). Fourth most common gene is *IFITM3* (interferon induced transmembrane protein 3), part of a protein family involved in host defense against pathogens and cellular transformation [76]. A total of forty-nine diseases were found to be associated with *IFITM3* (sixteen from annotation databases and thirty-three from literature review), which mainly include infectious diseases (influenza [77], coronavirus [78], zika virus [79], foot and mouth disease [80], and foamy viruses represented the rarest association [81]); neurological disorders (schizophrenia [82], Alzheimer's disease [83] and Huntington's disease [84]); and cancer (hepatocellular carcinoma [85], cutaneous melanoma [86], and *JAK2V617F+* myelofibrosis [87]). However, other associated disease categories include musculoskeletal/connective tissue, respiratory, digestive, and miscellaneous diseases with SLE showing multiple supporting studies [88].

*LGALS2* (galectin 2) found to be the fifth mutual gene in the patient cohort. It is an oxidative stress-response gene largely implicated in cardiovascular and cancer pathways [89]. A total of twenty-four diseases were found to be linked to this gene (three from annotation databases and twenty-one from literature review), which mainly include CVD/circulatory system predominated diseases (myocardial infarction [90] and coronary heart disease [91]); cancer (gastric [92] and colorectal [93]). Furthermore, it showed relationships with musculoskeletal/connective tissue, digestive, endocrine/metabolic disorders, and miscellaneous diseases with Behcet's disease emerging as one of the only significantly rare diseases [94]. Sixth common gene is *RETN* (resistin), secreted by adipocytes and implicated in metabolic, inflammatory, infectious, and cancer pathways [95]. A total of 110 diseases were identified for *RETN*, (nineteen diseases from annotation databases, and ninety-one from literature review), which mainly include relationships with musculoskeletal/connective tissue diseases (psoriatic arthritis [96], osteoarthritis [97], SLE [98], systemic sclerosis [99], and rheumatoid arthritis [100]); multiple CVD/circulatory system diseases (coronary artery disease [101], atherosclerosis [102], heart failure [103], pulmonary arterial hypertension [105] and the rare Kawasaki disease [101]); cancer (breast [106], colorectal [107], and the rare type of gallbladder [108]); endocrine/metabolic disorders (obesity [109], metabolic syndrome [110], and type 2 diabetes mellitus [111]); infectious diseases (urogenital Schistosomiasis [112]). Additional associations include digestive, neurological, genitourinary, skin, respiratory, blood, and miscellaneous diseases (chronic kidney disease [113], acne vulgaris [114], and Alzheimer's disease [115]).

*RHD* (Rh blood group D antigen) is seventh most common gene, which is part of the Rh blood group system, which is the most important protein-based blood group system [116]. A total of thirty diseases were found for *RHD* (twenty from annotation databases and ten from literature review). This gene is primarily implicated in anti-D alloimmunization during pregnancy, leading to hemolytic disease of the fetus and newborn anemia [117, 118]. Its additional associations include CVD/circulatory system (coronary artery disease [119]), blood, cancers, infectious and miscellaneous disorders (sickle cell disease [120], chronic myeloid leukemia [121], and schizophrenia [122]), and perinatal period. Interestingly, the Rh-null blood type, also known as "golden blood", is the rarest blood type in the world, and pertains to diseases such as Rhnull disease, characterized by stomatocytosis and chronic hemolytic anemia [123]. Last common gene is the *TNFRSF13B* (TNF receptor superfamily member 13B). It encodes transmembrane activator and calcium modulator and cyclophilin ligand interactor (TACI) and plays an instrumental role in humoral immunity [124]. A total of seventy six diseases were identified for *TNFRSF13B* (twenty-eight diseases from annotation databases and 48 diseases from literature review), which mainly include cancer (multiple

myeloma [125], chronic lymphocytic leukemia [126], and breast cancer [127]); congenital malformations/chromosomal abnormalities; blood diseases (immunodeficiency [128]); musculoskeletal/connective tissue disorders (SLE [129], rheumatoid arthritis [130] and Sjogren's syndrome [131]); and infectious diseases (SARS-COV-2 [132]). Other categories include diseases associated with respiratory system, CVD/circulatory system, neurological, ear and mastoid process, and miscellaneous diseases (rare Meniere's disease [133], and IgA nephropathy [134]).

Extensive literature review and association analysis of overlapping diseases among the eight mutually predictive genes revealed rheumatoid arthritis as the most common disease, associated with seven genes (*ACKR1*, *C4B*, *IFI44L*, *IFITM3*, *LGALS2*, *RETN*, and *TNFRSF13B*). SLE followed, associated with 6 genes (*ACKR1*, *RETN*, *IFITM3*, *TNFRSF13B*, *C4B*, and *IFI44L*) with strong literature support across all associations. Behcet's disease and hepatitis C virus were found to be connected to five genes (*ACKR1*, *C4B*, *IFI44L*, *IFITM3*, *RETN*). These overlaps highlight shared disease pathways and suggest potential gene-gene interactions demanding further investigation. To clinically validate and better understand the implications of our results, we matched them with the diagnoses extracted from the electronic health records (EPIC Health System). In total, we discovered and reported 412 disorders (after redundancy removal) and from which 116 diseases were matched with diagnosis reported in the clinical records, which include 18 CVD/circulatory system, 13 endocrine/nutritional and metabolic, 12 cancer, 11 musculoskeletal/connective tissue, 10 digestive system, 9 genitourinary system, 8 infectious, 6 nervous system, 6 respiratory system, 5 skin and subcutaneous tissue diseases, 5 neurodevelopmental, 5 blood, and 3 eye and adnexa diseases, and 3 uncategorized/multiple category diseases/isolated and 2 abnormal clinical disorders. Out of these 116 matched diseases, 87 were completely and 29 partially matched based on reflected umbrella terminology, minor disease progression differences, or ICD9–ICD10 coding discrepancies. The remaining 296 unmatched diseases spanned similar categories, including 34 infectious diseases, 33 cancer, 21 CVD/circulatory system, 21 musculoskeletal/connective tissue, 20 blood and blood-forming organ, 15 respiratory system, 12 endocrine/metabolic, 11 congenital malformations/chromosomal abnormalities, 10 nervous system, 6 skin and subcutaneous tissue, 5 eye/adnexa/ear, 3 digestive system diseases, 3 neurodevelopmental, 3 conditions originating in the perinatal period, and 99 uncategorized/combination of categories/isolated in category. In addition, clinical records confirmed strong overlapping for rheumatoid arthritis, systemic lupus erythematosus, and hepatitis C virus. However, Behcet's disease did not appear in clinical records for the observed patients suggesting a potentially novel inflammatory pathway identified through multi-omics/genomics analysis.

We hypothesize that these unmatched and missing diseases in clinical diagnoses may represent potential novel or preclinical pathways, therefore, require further investigation and validation in the future.

## Methods

The overall implemented methodology of this study includes human blood sample collection for multi-omics data generation and preprocessing, and individual and integrated analysis of RNA-seq and WGS data using translational bioinformatics and data driven ML techniques.

### ***Multi-omics data generation & preprocessing***

RNA-seq and WGS were performed on DNA and RNA extracted from blood samples of randomly consented patients ( $n = 96$ ). The sequencing libraries were prepared by random fragmentation, followed by 5' and 3' adapter ligation. Adapter-ligated fragments were then PCR amplified and gel purified. Illumina NovaSeq6000 S4 (2x150bp) was utilized for sequencing with 30X coverage using Illumina compatible libraries (e.g., Nextera DNA Flex and TruSeq Stranded mRNA). All human samples were used in accordance with relevant guidelines and regulations, and all experimental protocols were approved by the Institutional Review Board (IRB). We obtained informed consent from all patients. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institution and the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Preprocessing of the raw RNA-seq data was carried out using the GVVIZ pipeline, which mainly applies Hierarchical Indexing for Spliced Alignment of Transcripts (HISAT) with Bowtie2 to align the sequences against reference human genome (hg38). Next, RNA by Expectation Maximization (RSEM) was utilized for the quantification and identification of Differentially Expressed Genes (DEG) by aligning reads to reference de novo transcriptome assemblies [135]. Furthermore, we performed gene-disease data annotation with GVVIZ to identify, classify, and connect significantly expressed genes to human phenotypes and diseases. The expected RNA-seq counts were formatted into a matrix, where rows corresponded to Subject Identifiers (SIDs) and columns represented features. We then retained only those features that were based on protein-coding genes with a mean Transcripts Per Million (TPM) above 10 across the cohort. To preprocess raw WGS data and extract variants into a usable dataset [136], we applied our pipeline i.e., JWES (Java-based Whole Genome/Exome) [137]. JWES combines various tools, including Burrows-Wheeler Aligner (BWA) for reference alignment [138], Genome Analysis Toolkit (GATK) for variant calling [139], and SnpEff for variant annotation [140]. The final outcome of JWES is a Variant Call Formatted (VCF) file, which automatically parsed and stored into relational databases. We utilized VCF extracted information for further downstream analysis, including variant-level quality filtering, classification of

coding variations, identification of functionally relevant alleles, and annotation. Preprocessed data from JWES served as the entry point for our subsequent WGS-based variant investigation using ML techniques.

### ***RNA-seq data analysis***

The RNA-seq data analysis methodology includes expression data preprocessing, variance filtering, disease prediction, clustering, and enrichment analysis (Figure 1). In one of our previous studies, we generated RNA-seq dataset ( $n = 61$ ) from a diverse cohort of patients [141]. We utilized these samples to train an accurate, generalizable model for disease prediction. Our custom scikit-learn pipeline initiated with a K-Nearest Neighbors (KNN) imputer [142], using the five nearest neighbors with distance-weighted contributions to fill in missing values. After imputation analysis, a custom quantile transformer mapped the features to a uniform distribution ( $n$  quantiles = 40). This normalization strategy enabled our model's applicability across RNA-seq datasets and minimized the batch effects [143]. Next, we performed feature selection using the minimum Redundancy Maximum Relevance (mRMR) algorithm with a Mutual Information Quotient (MIQ) criterion to retain the most informative features ( $n=15$ ) [144, 145]. Our pipeline's primary learner is a Random Forest (RF) classifier, instantiated with trees ( $n=100$ ) and balanced class weights [146, 27]. We implemented a nested cross-validation scheme to provide a performance estimate and tune our model's hyperparameters. Our outer loop employed repeated stratified 5-fold cross-validation, with each fold repeated five times, resulting in unique dataset splits ( $n=25$ ). For each outer split, training fold was subjected to an inner cross-validation procedure, where a randomized search over a space of hyperparameters was conducted. We sampled 10 parameter configurations in each inner search, and the pipeline was refitted for each configuration.

After each inner search, the best parameter settings were identified based on the highest F1 score of the minority class. The scikit-learn pipeline was then retrained on the entire outer training set. Probability calibration was performed on this retrained classifier using CalibratedClassifierCV (*method* = 'sigmoid'), employing a "prefit" strategy such that the calibration module reused the same outer training data to learn a post-processing function that aligned the predicted probabilities with the empirical outcomes [142]. Our calibrated model was evaluated on the unseen outer testing fold, producing per-fold estimates of the F1 score of the healthy control class, Matthews Correlation Coefficient (MCC), balanced accuracy, and standard accuracy. During each of the outer folds ( $n=25$ ), we recorded both predicted labels and calibrated predicted probabilities on the external holdout data, ultimately generating sets of performance metrics ( $n=25$ ) that were aggregated and reported with 95% Confidence Intervals (CIs) derived from 1000-sample bootstrapping. Following the nested cross-validation, entire dataset was used for a final hyperparameter

search over the same parameter space, relying on a standard 5-fold cross-validation strategy to identify the most effective pipeline settings. The resulting pipeline was trained on the full data and subjected to sigmoid calibration on all samples. Our prediction pipeline was used on our randomly consented cohort after outliers identified by Principal Component Analysis (PCA) were removed. The features were subsetted to the genes previously selected by mRMR. Calibrated probability estimates were generated for each of the remaining samples, after which a 0.70 probability threshold was used to define “confident” positive classifications.

We merged disease predictions with gene expression features to construct a pseudo-labeled version of our randomly consented cohort’s dataset. Using DESeq2, we normalized RNA-seq raw count values [147, 148] and filtered for highly variable genes. Next, we calculated the Coefficient of Variation (CV) for each feature and retained the top 5000 genes. The dataset was re-normalized from raw counts using the same methodology as this set of features, as DESeq2 normalization is dependent on the composition of the entire dataset [148]. Uniform Manifold Approximation and Projection (UMAP) was used to perform dimensionality reduction [149]. We configured the output to two dimensions and incorporated pseudo-labeled information as a soft constraint (*target metric* = 'categorical', *target weight* = 0.05). Clustering was carried out on these UMAP embeddings using HDBSCAN [150, 151]. Minimum cluster size and sample parameters were both set to “3”. Next, we used DESeq2 to identify DEG in each cluster relative to all other samples. We deemed genes with an adjusted *p-value* less than “0.05” and a minimum absolute Log2 Fold Change (LFC) greater than “1.5” as differentially expressed. Genes satisfying these cutoffs for each cluster were subjected to enrichment analysis using GSEAPy’s Enrichr module [152, 153].

### **WGS data analysis**

The analysis of WGS data valuably provided insights into complex diseases and their biological bases [154, 155]. To identify relevant and pathogenic mutations associated with disease, we divided our WGS-based methodology into three phases (Figure 1): variant extraction and annotation; pathogenicity classification; and clustering. We processed WGS data in parallel with RNA-seq driven expression data to independently identify and validate significant biomarkers and their relationship with diseases (if any exists) to construct a comprehensive single-patient profile derived from multi-omics data. Next, we employed a two-stage preprocessing pipeline including variant filtering and annotation to extract functionally significant and pathogenic variants, and to identify disease-related biomarkers. For the consistent and comparable results between both methodologies, we performed WGS-based variant analysis using the same preliminary gene set generated by our RNA-seq data methodology. Specifically, all significantly expressed protein-coding

genes identified during RNA-seq data analysis formed the primary filtering criterion for extracting targeted variants. Targeting these specific genes ensured greater consistency and comparability with our RNA-seq methodology, as mutually identified biomarkers better reveal underlying pathways and strengthen confidence in retrieved results. It also aids in minimizing the effects of confounding data by narrowing the scope of analysis to only deleterious and impactful variants. Preprocessed variants were extracted using our pipeline i.e., *V*ariant *R*eduction and *A*Nnotation (*VAREANT*) [20], with the rest pruned from downstream analysis (Figure 1, W1). As a means of determining disease associations and relevant gene information, each variant was annotated with VAREANT pipeline using two public annotation databases: the Single Nucleotide Polymorphism Database (dbSNP) [156] and ClinVar [157] (Figure 1 W2). Unknown variants (i.e., those without RS numbers) were discarded from the subsequent analysis.

We focused solely on pathogenic variants and those with known disease correlations to identify significant biomarkers to create unique biological profiles per individual. Specifically, we identified all those variants marked as pathogenic, likely pathogenic, or a risk factor for disease (Figure 1, W3). Those otherwise marked benign or with unknown pathogenicity were excluded from the dataset to reduce the influence of confounding and irrelevant data. Before further scoring, it was essential to first normalize gene-variant counts. Since longer genes are more likely to have reads containing pathogenic variants, we first normalized the variant counts based on gene length to avoid underrepresenting shorter genes (Figure 1, W4). Protein-coding gene lengths were sourced from Ensembl BioMart and used to inversely weigh the prevalence of variants when scoring their importance [158]. Pathogenic variants and corresponding genes were then scored using their normalized prevalence among individuals in our cohort (Figure 1, W5).

We divided patients into clusters, based on the distribution of pathogenic variants to identify disease trends and better understand similarities between individuals among the cohort. Patients with similar pathogenic variants are likely to share biological similarities and disease phenotypes. An AI/ML-ready dataset was constructed to aggregate patient data with gene-specific variant data. Particularly, a matrix indicating the prevalence of pathogenic variants in specific genes and individuals was used to cluster patients with an agglomerative (i.e., bottom-up) hierarchical clustering approach (Figure 1, W6). Individuals with the most similar variant profiles are recursively merged to form higher-order clusters. To visualize and interpret the distribution of these clusters, we employed UMAP in 2D (Figure 1, W6). UMAP is a nonlinear projection technique optimized for preserving the local structure of high-dimensional data [149]. Cluster-specific feature annotations, including relevant genes and diseases, were aggregated into individual patient profiles (Figure 1, W7). These patient profiles offer greater comparability for single-

patient data to drive precision medicine in addition to having clinical importance in early diagnosis of disease. For each patient, we identified a set of significant genes alongside correlated diseases determined from ClinVar annotations during pathogenicity analysis. Additionally, we utilized ICD-10 codes to further categorize, label, and interpret diseases.

### ***Gene-disease association and patient profiling***

We normalized raw RNA-seq count data from our cohort of randomly consented patients using the same DESeq2 normalization approach utilized in the previous subsection. Next, we calculated CV across all samples to ensure that outlier expression levels would not affect downstream analysis [160]. Raw counts were not re-normalized during any step in this analysis. We built a pipeline to compare a subject of interest to a weighted-average representative of the cohort. First, the dataset was partitioned into two sets: a vector corresponding to initial and a matrix containing all other samples in the cohort. Next, a KNN approach was employed to estimate the distance for every sample in the cohort. Distances were converted into Gaussian weights by centering a normal distribution on the mean of the observed distances and scaling by their standard deviation. This weighting scheme enabled us to construct the profile by taking a weighted average of the reference expression counts, ultimately producing a single synthetic sample that reflects a smooth, population-level baseline for the target patient. Gene-level differences were quantified by computing the LFC for each gene and adding a small constant to prevent division by zero. LFC values were sorted by their absolute magnitude, and the top 20 genes were selected for further investigation. GSEAPy's Enrichr was again utilized, producing patient-specific annotations based on those 20 genes [152, 153]. For each annotation, enrichment statistics, gene sets, and significance measures were determined for individual patients of the cohort.

We utilized Open Targets Platform API (Application Programming Interface) for gene-disease associations with a confidence score  $\geq 0.5$  [160] to filter out weaker links. Each disease returned by Open Targets was assigned to one of the eighteen disease categories (Blood and blood-forming organ diseases; Nervous system diseases; Immune system diseases; Genitourinary system diseases; Endocrine, nutritional and metabolic diseases; Congenital and genetic diseases; Skin and subcutaneous tissue diseases; Abnormal clinical findings; Musculoskeletal diseases; Respiratory system diseases; Viral/bacterial/infectious diseases; Circulatory system diseases/CVD; Mental/behavioral/neurodevelopmental diseases; Eye and adnexa diseases; Digestive system diseases; Ear, nose and throat diseases; Cancers; and Injuries and wounds) using a three-layer classification system. The first layer used the therapeutic area tags provided directly by Open Targets (e.g., 'cardiovascular disease', 'cancer or benign tumor') and mapped them into

standardized clinical categories. The second layer applied a medical priority rule for diseases tagged under multiple therapeutic areas, ensuring the most clinically specific category was selected over broader parent categories (e.g., diseases tagged as both 'nervous system disease' and 'disorder of visual system' were correctly assigned to Eye and Adnexa rather than the broader Nervous System). The third layer applied a curated set of name-based overrides for cases where Open Targets' tagging was clinically inaccurate, for instance, Alzheimer's disease was tagged as a psychiatric disorder in Open Targets but was correctly reassigned to Nervous System diseases. In total, over 100 such overrides were applied and validated. Additionally, 25 entries returned by Open Targets were identified as non-disease traits (e.g. 'alcohol drinking', 'hair color', 'response to statin') and were excluded from the categorization. Furthermore, to prevent non-specific cancer associations from inflating cancer disorders, a known limitation of broad gene-disease databases, the cancers flag was restricted exclusively to patients carrying a gene from the OncoKB Cancer Gene Census (a curated list of 389 genes with proven causal roles in cancer) [161].

## References

1. Gunderson, C. G., Bilan, V. P., Holleck, J. L., Nickerson, P., Cherry, B. M., Chui, P., Bastian, L. A., Grimshaw, A. A., & Rodwin, B. A. (2020). Prevalence of harmful diagnostic errors in hospitalised adults: a systematic review and meta-analysis. *BMJ quality & safety*, *29*(12), 1008–1018.
2. Newman-Toker, D. E., Nassery, N., Schaffer, A. C., Yu-Moe, C. W., Clemens, G. D., Wang, Z., Zhu, Y., Saber Tehrani, A. S., Fanai, M., Hassoon, A., & Siegal, D. (2024). Burden of serious harms from diagnostic error in the USA. *BMJ quality & safety*, *33*(2), 109–120.
3. Singh, H., Giardina, T. D., Meyer, A. N., Forjuoh, S. N., Reis, M. D., & Thomas, E. J. (2013). Types and origins of diagnostic errors in primary care settings. *JAMA internal medicine*, *173*(6), 418–425.
4. Makary, M. A., & Daniel, M. (2016). Medical error—the third leading cause of death in the US. *BMJ (Clinical research ed.)*, *353*, i2139.
5. Zeeshan, S., Xiong, R., Liang, B. T., & Ahmed, Z. (2020). 100 Years of evolving gene-disease complexities and scientific debutants. *Briefings in bioinformatics*, *21*(3), 885–905.
6. Abdelhalim, H., Berber, A., Lodi, M., Jain, R., Nair, A., Pappu, A., Patel, K., Venkat, V., Venkatesan, C., Wable, R., Dinatale, M., Fu, A., Iyer, V., Kalove, I., Kleyman, M., Koutsoutis, J., Menna, D., Paliwal, M., Patel, N., Patel, T., ... Ahmed, Z. (2022). Artificial Intelligence, Healthcare, Clinical Genomics, and Pharmacogenomics Approaches in Precision Medicine. *Frontiers in genetics*, *13*, 929736.

7. Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database: the journal of biological databases and curation*, 2020, baaa010.
8. Ahmed Z. (2022). Precision medicine with multi-omics strategies, deep phenotyping, and predictive analysis. *Progress in molecular biology and translational science*, 190(1), 101–125.
9. Shenouda, A., Senthilkumar, S., Mourad, Y., Xie, J., Peker, E., Zeeshan, S., & Ahmed, Z. (2026). Artificial intelligence to investigate metabolomics data for precision medicine. *Metabolomics: Official journal of the Metabolomic Society*, 22(2), 29.
10. Ahmed, Z., & Alam, M. N. (2026). Artificial intelligence and precision medicine. *Scientific reports*, 16(1), 2186.
11. Ahmed, Z., Wan, S., Zhang, F. & Zhong, W. (2024). Artificial intelligence for omics data analysis. *BMC Methods* 1, 4
12. Ahmed Z. (2020). Practicing precision medicine with intelligently integrative clinical and multi-omics data analysis. *Human genomics*, 14(1), 35.
13. Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature reviews. Genetics*, 20(8), 467–484.
14. Singh, K., Usman, S., Zeeshan, S., Yanamala, N., Bhise, V., Nichols, M., Bokhari, S., Sengupta, P., & Ahmed, Z. (2025). Bioinformatics and AI/ML approaches using multi-omics data to accelerate diagnosis and delivery of precision care for patients with rare diseases. *Methods in Cell Biology (MCB): 2D and 3D Cellular Screening Models and AI Guided Analysis*, vol 204. Editors: Oliver Kepp and Guido Kroemer. Academic Press.
15. Uffelmann, E., Huang, Q.Q., Munung, N.S. *et al.* (2021). Genome-wide association studies. *Nat Rev Methods Primers* 1, 59
16. Jiang, Z., Zhou, X., Li, R., Michal, J. J., Zhang, S., Dodson, M. V., Zhang, Z., & Harland, R. M. (2015). Whole transcriptome analysis with sequencing: methods, challenges and potential solutions. *Cellular and molecular life sciences: CMLS*, 72(18), 3425–3439.
17. Patel, K. K., Venkatesan, C., Abdelhalim, H., Zeeshan, S., Arima, Y., Linna-Kuosmanen, S., & Ahmed, Z. (2023). Genomic approaches to identify and investigate genes associated with atrial fibrillation and heart failure susceptibility. *Human genomics*, 17(1), 47.

18. Zech, M., & Winkelmann, J. (2024). Next-generation sequencing and bioinformatics in rare movement disorders. *Nature reviews. Neurology*, *20*(2), 114–126.
19. Kaya, C., Dorsaint, P., Mercurio, S., Campbell, A. M., Eng, K. W., Nikiforova, M. N., Elemento, O., Nikiforov, Y. E., & Sboner, A. (2021). Limitations of Detecting Genetic Variants from the RNA Sequencing Data in Tissue and Fine-Needle Aspiration Samples. *Thyroid: official journal of the American Thyroid Association*, *31*(4), 589–595.
20. Narayanan, R., DeGroat, W., Peker, E., Zeeshan, S., & Ahmed, Z. (2024). VAREANT: a bioinformatics application for gene variant reduction and annotation. *Bioinformatics advances*, *5*(1), vbae210.
21. Vadapalli, S., Abdelhalim, H., Zeeshan, S., & Ahmed, Z. (2022). Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine. *Briefings in bioinformatics*, *23*(5), bbac191.
22. Armoundas, A. A., Narayan, S. M., Arnett, D. K., Spector-Bagdady, K., Bennett, D. A., Celi, L. A., Friedman, P. A., Gollob, M. H., Hall, J. L., Kwitek, A. E., Lett, E., Menon, B. K., Sheehan, K. A., Al-Zaiti, S. S., & American Heart Association Institute for Precision Cardiovascular Medicine; Council on Cardiovascular and Stroke Nursing; Council on Lifelong Congenital Heart Disease and Heart Health in the Young; Council on Cardiovascular Radiology and Intervention; Council on Hypertension; Council on the Kidney in Cardiovascular Disease; and Stroke Council (2024). Use of Artificial Intelligence in Improving Outcomes in Heart Disease: A Scientific Statement From the American Heart Association. *Circulation*, *149*(14), e1028–e1050.
23. DeGroat, W., Mendhe, D., Bhusari, A., Abdelhalim, H., Zeeshan, S., & Ahmed, Z. (2023). IntelliGenes: a novel machine learning pipeline for biomarker discovery and predictive analysis using multi-genomic profiles. *Bioinformatics (Oxford, England)*, *39*(12), btad755.
24. Narayanan, R., DeGroat, W., Mendhe, D., Abdelhalim, H., & Ahmed, Z. (2024). *IntelliGenes*: Interactive and user-friendly multimodal AI/ML application for biomarker discovery and predictive medicine. *Biology methods & protocols*, *9*(1), bpae040.
25. Narayanan, R., Peker, E., Degroat, W., Mendhe, D., Zeeshan, S., & Ahmed, Z.\* (2025). 3D IntelliGenes: AI/ML application using multi-omics data for biomarker discovery and disease prediction with multi-dimensional visualization. medRxiv, preprint.

26. Degroat, W., Venkat, V., Pierre-Louis, W., Abdelhalim, H., & Ahmed, Z. (2023). Hygieia: AI/ML pipeline integrating healthcare and genomics data to investigate genes associated with targeted disorders and predict disease. *Software Impacts*, 100493.
27. DeGroat, W., Abdelhalim, H., Patel, K., Mendhe, D., Zeeshan, S., & Ahmed, Z. (2024). Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. *Scientific reports*, 14(1), 1.
28. Vahabi, N., & Michailidis, G. (2022). Unsupervised Multi-Omics Data Integration Methods: A Comprehensive Review. *Frontiers in genetics*, 13, 854752.
29. Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., & Stegle, O. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular systems biology*, 14(6), e8124.
30. DeGroat, W., Abdelhalim, H., Peker, E., Sheth, N., Narayanan, R., Zeeshan, S., Liang, B. T., & Ahmed, Z. (2024). Multimodal AI/ML for discovering novel biomarkers and predicting disease using multi-omics profiles of patients with cardiovascular diseases. *Scientific reports*, 14(1), 26503.
31. Srivastava, K., Khil, P. P., Sippert, E., Volkova, E., Dekker, J. P., Rios, M., & Flegel, W. A. (2020). ACKR1 Alleles at 5.6 kb in a Well-Characterized Renewable US Food and Drug Administration (FDA) Reference Panel for Standardization of Blood Group Genotyping. *The Journal of molecular diagnostics : JMD*, 22(10), 1272–1279.
32. Martini, R., Chen, Y., Jenkins, B. D., Elhussin, I. A., Cheng, E., Hoda, S. A., Ginter, P. S., Hanover, J., Zeidan, R. B., Oppong, J. K., Adjei, E. K., Jibril, A., Chitale, D., Bensenhaver, J. M., Awuah, B., Bekele, M., Abebe, E., Kyei, I., Aitpillah, F. S., Adinku, M. O., ... Davis, M. B. (2021). Investigation of triple-negative breast cancer risk alleles in an International African-enriched cohort. *Scientific reports*, 11(1), 9247.
33. Reyes, N., Benedetti, I., Rebollo, J., Correa, O., & Geliebter, J. (2017). Atypical chemokine receptor CCRL2 is overexpressed in prostate cancer cells. *Journal of biomedical research*, 33(1), 17–23. Advance online publication.
34. Bleckwehl, T., Babler, A., Tebens, M., Maryam, S., Nyberg, M., Bosteen, M., Halder, M., Shaw, I., Fleig, S., Pyke, C., Hvid, H., Voetmann, L. M., van Buul, J. D., Sluimer, J. C., Das, V., Baumgart, S., Kramann, R., & Hayat, S. (2025). Encompassing view of spatial and single-cell RNA sequencing renews the role of the microvasculature in human atherosclerosis. *Nature cardiovascular research*, 4(1), 26–44.

35. Hernández-Aguilera, A., Fibla, M., Cabré, N., Luciano-Mateo, F., Camps, J., Fernández-Arroyo, S., Martín-Paredero, V., Menendez, J. A., Sirvent, J. J., & Joven, J. (2020). Chemokine (C-C motif) ligand 2 and coronary artery disease: Tissue expression of functional and atypical receptors. *Cytokine*, *126*, 154923.
36. Tian, Z., Zhang, P., Li, X., & Jiang, D. (2023). Analysis of immunogenic cell death in ascending thoracic aortic aneurysms based on single-cell sequencing data. *Frontiers in immunology*, *14*, 1087978.
37. Dinardo, C. L., Oliveira, T. G. M., Kelly, S., Ashley-Koch, A., Telen, M., Schmidt, L. C., Castilho, S., Melo, K., Dezan, M. R., Wheeler, M. M., Johnsen, J. M., Nickerson, D. A., Jain, D., Custer, B., Pereira, A. C., Sabino, E. C., & NHLBI Recipient Epidemiology Donor Evaluation Study (REDS-III) International Component-Brazil, the Outcome Modifying Genes in SCD (OMG) study and the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program Sickle Cell Disease Working Group (2021). Diversity of variant alleles encoding Kidd, Duffy, and Kell antigens in individuals with sickle cell disease using whole genome sequencing data from the NHLBI TOPMed Program. *Transfusion*, *61*(2), 603–616.
38. Osborne, A., Phelan, J. E., Vanheer, L. N., Manjurano, A., Gitaka, J., Drakeley, C. J., Kaneko, A., Kita, K., Campino, S., & Clark, T. G. (2023). High throughput human genotyping for variants associated with malarial disease outcomes using custom targeted amplicon sequencing. *Scientific reports*, *13*(1), 12062.
39. Ford, A., Kepple, D., Abagero, B. R., Connors, J., Pearson, R., Auburn, S., Getachew, S., Ford, C., Gunalan, K., Miller, L. H., Janies, D. A., Rayner, J. C., Yan, G., Yewhalaw, D., & Lo, E. (2020). Whole genome sequencing of *Plasmodium vivax* isolates reveals frequent sequence and structural polymorphisms in erythrocyte binding genes. *PLoS neglected tropical diseases*, *14*(10), e0008234.
40. Zhang, S. T., Shi, M., Shao, L. N., Zhou, S. H., Yu, W. J., Chen, M., Xiao, N., Duan, Y., Pan, L. Z., Wang, N., Song, W. Q., Xia, Y. X., Zhang, L., Qi, N., & Liu, M. (2017). Atypical Chemokine Receptor 1 Polymorphism can not Affect Susceptibility to Hepatitis C Virus. *Balkan medical journal*, *34*(4), 308–312.
41. He, W., Neil, S., Kulkarni, H., Wright, E., Agan, B. K., Marconi, V. C., Dolan, M. J., Weiss, R. A., & Ahuja, S. K. (2008). Duffy antigen receptor for chemokines mediates trans-infection of HIV-1 from red blood cells to target cells and affects HIV-AIDS susceptibility. *Cell host & microbe*, *4*(1), 52–62.
42. Lipski, D. A., Foucart, V., Dewispelaere, R., Caspers, L. E., Defrance, M., Bruyns, C., & Willermain, F. (2020). Retinal endothelial cell phenotypic modifications during experimental autoimmune uveitis: a transcriptomic approach. *BMC ophthalmology*, *20*(1), 106.

43. Zhang, Z., Wang, P., Zhang, L., Huang, C., Gao, J., Li, Y., & Yang, B. (2020). Identification of Key Genes and Long Noncoding RNA-Associated Competing Endogenous RNA (ceRNA) Networks in Early-Onset Preeclampsia. *BioMed research international*, 2020, 1673486.
44. Zhou, D., Rudnicki, M., Chua, G. T., Lawrance, S. K., Zhou, B., Drew, J. L., Barbar-Smilely, F., Armstrong, T. K., Hilt, M. E., Birmingham, D. J., Passler, W., Auletta, J. J., Bowden, S. A., Hoffman, R. P., Wu, Y. L., Jarjour, W. N., Mok, C. C., Ardoin, S. P., Lau, Y. L., & Yu, C. Y. (2021). Human Complement C4B Allotypes and Deficiencies in Selected Cases With Autoimmune Diseases. *Frontiers in immunology*, 12, 739430.
45. Gu, X., Dou, M., Su, W., Jiang, Z., Duan, Q., Cao, B., & Chen, Y. (2022). Identifying novel proteins underlying schizophrenia via integrating pQTLs of the plasma, CSF, and brain with GWAS summary data. *BMC medicine*, 20(1), 474.
46. van Dongen, J., Zilhão, N. R., Sugden, K., BIOS Consortium, Hannon, E. J., Mill, J., Caspi, A., Agnew-Blais, J., Arseneault, L., Corcoran, D. L., Moffitt, T. E., Poulton, R., Franke, B., & Boomsma, D. I. (2019). Epigenome-wide Association Study of Attention-Deficit/Hyperactivity Disorder Symptoms in Adults. *Biological psychiatry*, 86(8), 599–607.
47. Li, Y., Qin, Z., Yang, M., Qin, Y., Lin, C., & Liu, S. (2011). Differential expression of complement proteins in cerebrospinal fluid from active multiple sclerosis patients. *Journal of cellular biochemistry*, 112(7), 1930–1937.
48. Odell, D., Maciulis, A., Cutler, A., Warren, L., McMahon, W. M., Coon, H., Stubbs, G., Henley, K., & Torres, A. (2005). Confirmation of the association of the C4B null allele in autism. *Human immunology*, 66(2), 140–145.
49. Zhang, X., Yin, X., Yu, H., Liu, X., Yang, F., Yao, J., Jin, H., & Yang, P. (2012). Quantitative proteomic analysis of serum proteins in patients with Parkinson's disease using an isobaric tag for relative and absolute quantification labeling, two-dimensional liquid chromatography, and tandem mass spectrometry. *The Analyst*, 137(2), 490–495.
50. Panitch, R., Hu, J., Chung, J., Zhu, C., Meng, G., Xia, W., Bennett, D. A., Lunetta, K. L., Ikezu, T., Au, R., Stein, T. D., Farrer, L. A., & Jun, G. R. (2021). Integrative brain transcriptome analysis links complement component 4 and HSPA2 to the APOE  $\epsilon$ 2 protective effect in Alzheimer disease. *Molecular psychiatry*, 26(10), 6054–6064.

51. Blaskó, B., Kolka, R., Thorbjornsdottir, P., Sigurdarson, S. T., Sigurdsson, G., Rónai, Z., Sasvári-Székely, M., Bödvarsson, S., Thorgeirsson, G., Prohászka, Z., Kovács, M., Füst, G., & Arason, G. J. (2008). Low complement C4B gene copy number predicts short-term mortality after acute myocardial infarction. *International immunology*, *20*(1), 31–37.
52. Mariaselvam, C. M., Seth, G., Kavadichanda, C., Boukouaci, W., Wu, C. L., Costes, B., Thabah, M. M., Krishnamoorthy, R., Leboyer, M., Negi, V. S., & Tamouza, R. (2024). Low C4A copy numbers and higher HERV gene insertion contributes to increased risk of SLE, with absence of association with disease phenotype and disease activity. *Immunologic research*, *72*(4), 697–706.
53. Clarkson, R., Bate, A. S., Grennan, D. M., Chattopadhyay, C., Sanders, P., Davis, M., & Kelly, C. (1990). DQw7 and the C4B null allele in rheumatoid arthritis and Felty's syndrome. *Annals of the rheumatic diseases*, *49*(12), 976–979.
54. Lintner, K. E., Patwardhan, A., Rider, L. G., Abdul-Aziz, R., Wu, Y. L., Lundström, E., Padyukov, L., Zhou, B., Alhomosh, A., Newsom, D., White, P., Jones, K. B., O'Hanlon, T. P., Miller, F. W., Spencer, C. H., & Yu, C. Y. (2016). Gene copy-number variations (CNVs) of complement C4 and C4A deficiency in genetic risk and pathogenesis of juvenile dermatomyositis. *Annals of the rheumatic diseases*, *75*(9), 1599–1606.
55. Lee, J., Min, H. K., Lim, J. Y., Song, Y. S., Jeon, J. H., Jang, S. G., Kim, M. J., Park, Y., Park, S. H., Kim, S. W., & Kwok, S. K. (2025). Human nasal turbinate stem cells with specific gene signatures (HAS2, CXCL1, KRTAP1-5, GSTT2B, and C4B) attenuate rheumatoid arthritis. *Scientific reports*, *15*(1), 6493.
56. Ahmed, E., Nityanand, S., Mustafa, A., Brismar, K., & Lefvert, A. K. (1999). Anti-cardiolipin antibodies and circulating immune complexes in type 1 diabetes mellitus: increased prevalence and relation to vascular complications. *Clinical and experimental immunology*, *115*(2), 255–259.
57. Liu, Y. H., Wan, L., Chang, C. T., Liao, W. L., Chen, W. C., Tsai, Y., Tsai, C. H., & Tsai, F. J. (2011). Association between copy number variation of complement component C4 and Graves' disease. *Journal of biomedical science*, *18*(1), 71.
58. Kim, J. H., Kim, G. H., Yoo, H. W., & Choi, J. H. (2023). Molecular basis and genetic testing strategies for diagnosing 21-hydroxylase deficiency, including CAH-X syndrome. *Annals of pediatric endocrinology & metabolism*, *28*(2), 77–86.

59. Hovingh, E. S., van den Broek, B., Kuipers, B., Pinelli, E., Rooijackers, S. H. M., & Jongerius, I. (2017). Acquisition of C1 inhibitor by Bordetella pertussis virulence associated gene 8 results in C2 and C4 consumption away from the bacterial surface. *PLoS pathogens*, *13*(7), e1006531.
60. Felberg, A., Bieńkowski, M., Stokowy, T., Myszczyński, K., Polakiewicz, Z., Kitowska, K., Sądej, R., Mohlin, F., Kuźniewska, A., Kowalska, D., Stasiłojć, G., Jongerius, I., Spaapen, R., Mesa-Guzman, M., Montuenga, L. M., Blom, A. M., Pio, R., & Okrój, M. (2024). Elevated expression of complement factor I in lung cancer cells associates with shorter survival-Potentially via non-canonical mechanism. *Translational research: the journal of laboratory and clinical medicine*, *269*, 1–13.
61. Jiao, M., Cui, Y., Qiu, X., Liang, X., Li, J., Guo, C., & Tian, H. (2024). Causal relationship between complement *C1QB* and colorectal cancer: a drug target Mendelian randomization study. *Frontiers in genetics*, *15*, 1403509.
62. Kretzschmar, G. C., Oliveira, L. C., Nisihara, R. M., Velavan, T. P., Stinghen, S. T., Stahlke, E. R. S., Petzl-Erler, M. L., Messias-Reason, I. J. T., & Boldt, A. B. W. (2018). Complement receptor 1 (CR1, CD35) association with susceptibility to leprosy. *PLoS neglected tropical diseases*, *12*(8), e0006705.
63. Rahmati, Y., Mollanoori, H., Najafi, S., Esmaeili, S., & Alivand, M. R. (2021). CASP5 and CR1 as potential biomarkers for Kawasaki disease: an Integrated Bioinformatics-Experimental Study. *BMC pediatrics*, *21*(1), 566.
64. Hu, Z., Zuo, C., Mao, C., Shi, C., & Xu, Y. (2023). Peripheral immune markers and amyotrophic lateral sclerosis: a Mendelian randomization study. *Frontiers in neuroscience*, *17*, 1269354.
65. Zeng, Y., Zhang, Z., Chen, H., Fan, J., Yuan, W., Li, J., Zhou, S., & Liu, W. (2022). Comprehensive Analysis of Immune Implication and Prognostic Value of *IFI44L* in Non-Small Cell Lung Cancer. *Frontiers in oncology*, *11*, 798425.
66. Cui, Y., Zhang, H., Wang, Z., Gong, B., Al-Ward, H., Deng, Y., Fan, O., Wang, J., Zhu, W., & Sun, Y. E. (2023). Exploring the shared molecular mechanisms between systemic lupus erythematosus and primary Sjögren's syndrome based on integrated bioinformatics and single-cell RNA-seq analysis. *Frontiers in immunology*, *14*, 1212330.
67. Raupov, R., Suspitsin, E., Preobrazhenskaya, E. V., & Kostik, M. (2024). Interferon type I signature associated with skin disease in juvenile dermatomyositis. *Frontiers in medicine*, *11*, 1214920.

68. Zhu, H., Wu, L. F., Mo, X. B., Lu, X., Tang, H., Zhu, X. W., Xia, W., Guo, Y. F., Wang, M. J., Zeng, K. Q., Wu, J., Qiu, Y. H., Lin, X., Zhang, Y. H., Liu, Y. Z., Yi, N. J., Deng, F. Y., & Lei, S. F. (2019). Rheumatoid arthritis-associated DNA methylation sites in peripheral blood mononuclear cells. *Annals of the rheumatic diseases*, *78*(1), 36–42.
69. Yang, Y., Zhang, H., Xiao, X., & Guo, M. (2024). Identification of EPST11 as a new potential biomarker for SLE based on GEO database. *Clinical rheumatology*, *43*(5), 1531–1540.
70. van Pul, L., van Dort, K. A., Girigorie, A. F., Maurer, I., Harskamp, A. M., & Kootstra, N. A. (2024). Human Immunodeficiency Virus-Induced Interferon-Stimulated Gene Expression Is Associated With Monocyte Activation and Predicts Viral Load. *Open forum infectious diseases*, *11*(8), ofae434.
71. Zhang, Z., Jin, H., Zhang, X., Bai, M., Zheng, K., Tian, J., Deng, B., Mao, L., Qiu, P., & Huang, B. (2024). Bioinformatics and system biology approach to identify the influences among COVID-19, influenza, and HIV on the regulation of gene expression. *Frontiers in immunology*, *15*, 1369311.
72. Liu, C., & Li, X. (2024). Identification of hub genes and establishment of a diagnostic model in tuberculosis infection. *AMB Express*, *14*(1), 36.
73. Huang, W. C., Tung, S. L., Chen, Y. L., Chen, P. M., & Chu, P. Y. (2018). IFI44L is a novel tumor suppressor in human hepatocellular carcinoma affecting cancer stemness, metastasis, and drug resistance via regulating met/Src signaling pathway. *BMC cancer*, *18*(1), 609.
74. Garau, J., Charras, A., Varesio, C., Orcesi, S., Dragoni, F., Galli, J., Fazzi, E., Gagliardi, S., Pansarasa, O., Cereda, C., & Hedrich, C. M. (2023). Altered DNA methylation and gene expression predict disease severity in patients with Aicardi-Goutières syndrome. *Clinical immunology (Orlando, Fla.)*, *249*, 109299.
75. Cecchi, I., Radin, M., Barinotti, A., Foddai, S. G., Menegatti, E., Roccatello, D., Suárez, A., Sciascia, S., & Rodríguez-Carrio, J. (2024). Type I interferon pathway activation across the antiphospholipid syndrome spectrum: associations with disease subsets and systemic antiphospholipid syndrome presentation. *Frontiers in immunology*, *15*, 1351446.
76. Friedlová, N., Zavadil Kokáš, F., Hupp, T. R., Vojtěšek, B., & Nekulová, M. (2022). IFITM protein regulation and functions: Far beyond the fight against viruses. *Frontiers in immunology*, *13*, 1042368.

77. Xu, W., Wang, Y., Li, L., Qu, X., Liu, Q., Li, T., Wu, S., Liao, M., Jin, N., Du, S., & Li, C. (2022). Transmembrane domain of IFITM3 is responsible for its interaction with influenza virus HA<sub>2</sub> subunit. *Virologica Sinica*, *37*(5), 664–675.
78. Xue, W., Chu, H., Wang, J., Sun, Y., Qiu, X., Song, C., Tan, L., Ding, C., & Liao, Y. (2024). Coronavirus nucleocapsid protein enhances the binding of p-PKC $\alpha$  to RACK1: Implications for inhibition of nucleocytoplasmic trafficking and suppression of the innate immune response. *PLoS pathogens*, *20*(11), e1012097.
79. Gobillot, T. A., Humes, D., Sharma, A., Kikawa, C., & Overbaugh, J. (2020). The Robust Restriction of Zika Virus by Type-I Interferon in A549 Cells Varies by Viral Lineage and Is Not Determined by IFITM3. *Viruses*, *12*(5), 503.
80. Wu, X., Yang, Y., Ru, Y., Hao, R., Zhao, D., Ren, R., Lu, B., Li, Y., Sun, S., Zheng, H., & Wang, W. (2024). Knockout of the WD40 domain of ATG16L1 enhances foot and mouth disease virus replication. *BMC genomics*, *25*(1), 796.
81. Wang, Z., Tuo, X., Zhang, J., Chai, K., Tan, J., & Qiao, W. (2022). Antiviral role of IFITM3 in prototype foamy virus infection. *Virology journal*, *19*(1), 195.
82. Volk, D. W., Chitrapu, A., Edelson, J. R., Roman, K. M., Moroco, A. E., & Lewis, D. A. (2015). Molecular mechanisms and timing of cortical immune activation in schizophrenia. *The American journal of psychiatry*, *172*(11), 1112–1121.
83. Feng, Y., Zhou, Q., Hu, B., Wang, S., Chen, L., Cai, W., Zhu, Q., Qin, X., Zhou, W., Wu, Y., & Song, W. (2025). Unveiling the role of KLF9-mediated IFITM3 regulation in amyloidogenesis. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *39*(4), e70403.
84. Andrade-Navarro, M. A., Mühlenberg, K., Spruth, E. J., Mah, N., González-López, A., Andreani, T., Russ, J., Huska, M. R., Muro, E. M., Fontaine, J. F., Amstislavskiy, V., Soldatov, A., Nietfeld, W., Wanker, E. E., & Priller, J. (2020). RNA Sequencing of Human Peripheral Blood Cells Indicates Upregulation of Immune-Related Genes in Huntington's Disease. *Frontiers in neurology*, *11*, 573560.
85. Barakat, L. A., Elsergany, A. R., Ghattas, M. H., Mahsoub, N., & Bondok, R. M. (2023). Relationship between interferon-induced transmembrane protein 3 and matrix metalloproteinase-9 gene polymorphisms in patients with hepatocellular carcinoma. *Clinics and research in hepatology and gastroenterology*, *47*(5), 102110.

86. Wang, Y. X., Ding, Z. J., Wang, Q. L., Zhao, C. C., Liu, S. Q., Du, S. L., Zhou, S., Zheng, L. Y., Gao, M., Shen, C. C., & Chen, X. D. (2024). CRISPR-Cas9 screening identified novel subtypes of cutaneous melanoma based on essential cancer genes. *Archives of dermatological research*, *317*(1), 86.
87. Zhao, Y., Wang, D., Liang, Y., Xu, C., Shi, L., & Tong, J. (2022). Expression profiles analysis identifies specific interferon-stimulated signatures as potential diagnostic and predictive indicators of *JAK2V617F*<sup>+</sup> myelofibrosis. *Frontiers in genetics*, *13*, 927018.
88. Wang, Z., Yang, C., Gao, W., Sun, W., Sun, J., Wang, H., Yan, S., & Xu, D. (2025). Systemic lupus erythematosus-specific CD14<sup>+</sup>IFITM3<sup>+</sup> monocyte: Implications for disease activity and progression. *International immunopharmacology*, *146*, 113916.
89. Liang, Z., Sun, R., Tu, P., Liang, Y., Liang, L., Liu, F., Bian, Y., Yin, G., Zhao, F., Jiang, M., Gu, J., & Tang, D. (2022). Immune-related gene-based prognostic index for predicting survival and immunotherapy outcomes in colorectal carcinoma. *Frontiers in immunology*, *13*, 944286.
90. Ozaki, K., Inoue, K., Sato, H., Iida, A., Ohnishi, Y., Sekine, A., Sato, H., Odashiro, K., Nobuyoshi, M., Hori, M., Nakamura, Y., & Tanaka, T. (2004). Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. *Nature*, *429*(6987), 72–75.
91. Yıldırım, C., Vogel, D. Y., Hollander, M. R., Baggen, J. M., Fontijn, R. D., Nieuwenhuis, S., Haverkamp, A., de Vries, M. R., Quax, P. H., Garcia-Vallejo, J. J., van der Laan, A. M., Dijkstra, C. D., van der Pouw Kraan, T. C., van Royen, N., & Horrevoets, A. J. (2015). Galectin-2 induces a proinflammatory, anti-arteriogenic phenotype in monocytes and macrophages. *PLoS one*, *10*(4), e0124347.
92. Jung, J. H., Kim, H. J., Yeom, J., Yoo, C., Shin, J., Yoo, J., Kang, C. S., & Lee, C. (2012). Lowered expression of galectin-2 is associated with lymph node metastasis in gastric cancer. *Journal of gastroenterology*, *47*(1), 37–48.
93. Li, H., Zhao, L., Lau, Y. S., Zhang, C., & Han, R. (2021). Genome-wide CRISPR screen identifies LGALS2 as an oxidative stress-responsive gene with an inhibitory function on colon tumor growth. *Oncogene*, *40*(1), 177–188.
94. Oğuz, A. K., Yılmaz, S. T., Oygür, Ç. Ş., Çandar, T., Sayın, I., Kılıçoğlu, S. S., Ergün, İ., Ateş, A., Özdağ, H., & Akar, N. (2016). Behçet's: A Disease or a Syndrome? Answer from an Expression Profiling Study. *PLoS one*, *11*(2), e0149052.

95. Li, Y., Yang, Q., Cai, D., Guo, H., Fang, J., Cui, H., Gou, L., Deng, J., Wang, Z., & Zuo, Z. (2021). Resistin, a Novel Host Defense Peptide of Innate Immunity. *Frontiers in immunology*, *12*, 699807.
96. Caso, F., Postiglione, L., Covelli, B., Ricciardone, M., Di Spigna, G., Formisano, P., D'Esposito, V., Girolimetto, N., Tasso, M., Peluso, R., Navarini, L., Ciccozzi, M., Margiotta, D. P. E., Oliviero, F., Afeltra, A., Punzi, L., Del Puente, A., Scarpa, R., & Costa, L. (2019). Pro-inflammatory adipokine profile in psoriatic arthritis: results from a cross-sectional study comparing PsA subset with evident cutaneous involvement and subset "sine psoriasis". *Clinical rheumatology*, *38*(9), 2547–2552.
97. Yang, Z., Ni, J., Kuang, L., Gao, Y., & Tao, S. (2020). Identification of genes and pathways associated with subchondral bone in osteoarthritis via bioinformatic analysis. *Medicine*, *99*(37), e22142.
98. Almedhed, K., d'Elia, H. F., Bokarewa, M., & Carlsten, H. (2008). Role of resistin as a marker of inflammation in systemic lupus erythematosus. *Arthritis research & therapy*, *10*(1), R15.
99. Pellicano, C., Leodori, G., Colalillo, A., Navarini, L., Gigante, A., & Rosato, E. (2022). Serum resistin is predictive marker of development of new digital ulcers in systemic sclerosis. *Clinical and experimental medicine*, *22*(3), 421–426.
100. Kim, W., Jin Oh, S., Thi Trinh, N., Yeon Gil, J., Ah Choi, I., Hyoun Kim, J., Hee Kim, J., Jung, J. Y., Kim, J., Kim, H. A., & Eun Lee, K. (2022). Effects of RETN polymorphisms on treatment response in rheumatoid arthritis patients receiving TNF- $\alpha$  inhibitors and utilization of machine-learning algorithms. *International immunopharmacology*, *111*, 109094.
101. Gao, S., Ma, W., Lin, X., Huang, S., & Yu, M. (2021). Identification of Key Genes and Underlying Mechanisms in Acute Kawasaki Disease Based on Bioinformatics Analysis. *Medical science monitor : international medical journal of experimental and clinical research*, *27*, e930547.
102. Li, Y., Zhang, W., Hu, Y., Xu, Z., Huo, Q., Qi, H., Liu, Q., & Xing, Y. (2024). Pericoronary adipose tissue radiomics features as imaging markers for coronary artery disease risk assessment: insights from gene expression analysis. *Cardiovascular diabetology*, *23*(1), 444.
103. Quispe, R., Sweeney, T., Martin, S. S., Jones, S. R., Allison, M. A., Budoff, M. J., Ndumele, C. E., Elshazly, M. B., & Michos, E. D. (2024). Associations of Adipokine Levels With Levels of Remnant Cholesterol: The Multi-Ethnic Study of Atherosclerosis. *Journal of the American Heart Association*, *13*(18), e030548.

104. Zhao, B., Bouchareb, R., & Lebeche, D. (2022). Resistin deletion protects against heart failure injury by targeting DNA damage response. *Cardiovascular research*, *118*(8), 1947–1963.
105. Gao, L., Skinner, J., Nath, T., Lin, Q., Griffiths, M., Damico, R. L., Pauciulo, M. W., Nichols, W. C., Hassoun, P. M., Everett, A. D., & Johns, R. A. (2024). Resistin predicts disease severity and survival in patients with pulmonary arterial hypertension. *Respiratory research*, *25*(1), 235.
106. Liu, L., Zhai, W., Wang, F., Yu, L., Zhou, F., Xiang, Y., Huang, S., Zheng, C., Yuan, Z., He, Y., Yu, Z., & Ji, J. (2022). Using machine learning to identify gene interaction networks associated with breast cancer. *BMC cancer*, *22*(1), 1070.
107. Wang, C. Q., Wang, Y., Huang, B. F., Tang, C. H., Du, Z., Zeng, Y., Wang, Q., Shao, J. K., & Jin, L. L. (2020). High Expression of Both Resistin and Fascin-1 Predicts a Poor Prognosis in Patients with Colorectal Cancer. *BioMed research international*, *2020*, 8753175.
108. Koshiol, J., Castro, F., Kemp, T. J., Gao, Y. T., Roa, J. C., Wang, B., Nogueira, L., Araya, J. C., Shen, M. C., Rashid, A., Hsing, A. W., Hildesheim, A., Ferreccio, C., Pfeiffer, R. M., & Pinto, L. A. (2016). Association of inflammatory and other immune markers with gallbladder cancer: Results from two independent case-control studies. *Cytokine*, *83*, 217–225.
109. Chiodi, V., Rappa, F., Lo Re, O., Chaldakov, G. N., Lelouvier, B., Micale, V., Domenici, M. R., & Vinciguerra, M. (2023). Deficiency of histone variant macroH2A1.1 is associated with sexually dimorphic obesity in mice. *Scientific reports*, *13*(1), 19123.
110. Ibrahim, S. M., & Bastawy, A. A. (2020). The Relevance of Single-nucleotide Polymorphism +62 G>A to the Expression of Resistin Gene Affecting Serum Resistin Levels in Metabolic Syndrome in the Egyptian Population. *Current pharmaceutical biotechnology*, *21*(7), 626–634.
111. Hedayati, M., Nezhadali, M., Karimi, H., Daneshpour, M. S., Najd-Hassan-Bonab, L., & Sayyar, N. (2025). Association of resistin rs1862513, rs10401670 and rs3745367 polymorphisms with resistin level and insulin resistance in an Iranian type 2 diabetic population. *Journal of diabetes and metabolic disorders*, *24*(1), 38.
112. Chimponda, T. N., Mushayi, C., Osakunor, D. N. M., Vengesai, A., Enwono, E., Amanfo, S., Murray, J., Tshuma, C., Mutapi, F., & Mduluzza, T. (2019). Elevation of C-reactive protein, P-selectin and Resistin as potential inflammatory biomarkers of urogenital Schistosomiasis exposure in preschool children. *BMC infectious diseases*, *19*(1), 1071.

113. Wruck, W., Boima, V., Erichsen, L., Thimm, C., Koranteng, T., Kwakyi, E., Antwi, S., Adu, D., & Adjaye, J. (2022). Urine-Based Detection of Biomarkers Indicative of Chronic Kidney Disease in a Patient Cohort from Ghana. *Journal of personalized medicine*, *13*(1), 38.
114. Chen, X., Min, S., Chen, C., Lin, X., Wang, D., & Jiang, G. (2022). Influence of RETN, IL-1, and IL-6 gene polymorphisms on the risk of acne vulgaris in the Chinese population. *Journal of cosmetic dermatology*, *21*(10), 4965–4973.
115. Rehman, H., Ang, T. F. A., Tao, Q., Espenilla, A. L., Au, R., Farrer, L. A., Zhang, X., Qiu, W. Q., & Alzheimer's Disease Neuroimaging Initiative (2024). Comparison of Commonly Measured Plasma and Cerebrospinal Fluid Proteins and Their Significance for the Characterization of Cognitive Impairment Status. *Journal of Alzheimer's disease: JAD*, *97*(2), 621–633.
116. Flegel W. A. (2007). The genetics of the Rhesus blood group system. *Blood transfusion = Trasfusione del sangue*, *5*(2), 50–57.
117. Isakson, P., & Pardi, C. (2023). Evaluation of an automated platform for non-invasive single-exon fetal RHD genotyping early in pregnancy. *Blood transfusion = Trasfusione del sangue*, *21*(6), 472–478.
118. Thomsen, W. C., & Afify, Z. A. (2021). Late Hyporegenerative Anemia Following Successful Management of Rh-D Alloimmunization: Still an Underappreciated Entity. *Clinical pediatrics*, *60*(9-10), 399–402.
119. Chai, T., Wang, Z., Yang, X., Qiu, Z., & Chen, L. (2022). *PSRC1* May Affect Coronary Artery Disease Risk by Altering *CELSR2*, *PSRC1*, and *SORT1* Gene Expression and Circulating Granulin and Apolipoprotein B Protein Levels. *Frontiers in cardiovascular medicine*, *9*, 763015.
120. Takasaki, K., Friedman, D. F., Uter, S., Vege, S., Westhoff, C. M., & Chou, S. T. (2023). Variant RHD alleles and Rh immunization in patients with sickle cell disease. *British journal of haematology*, *201*(6), 1220–1228.
121. Chérif-Zahar, B., Bony, V., Steffensen, R., Gane, P., Raynal, V., Goosens, D., Laursen, J. S., Varming, K., Jersild, C., & Cartron, J. P. (1998). Shift from Rh-positive to Rh-negative phenotype caused by a somatic mutation within the RHD gene in a patient with chronic myelocytic leukaemia. *British journal of haematology*, *102*(5), 1263–1270.

122. Palmer, C. G., Mallery, E., Turunen, J. A., Hsieh, H. J., Peltonen, L., Lonnqvist, J., Woodward, J. A., & Sinsheimer, J. S. (2008). Effect of Rhesus D incompatibility on schizophrenia depends on offspring sex. *Schizophrenia research*, *104*(1-3), 135–145.
123. Huang, C. H., Chen, Y., Reid, M. E., & Seidl, C. (1998). Rhnull disease: the amorph type results from a novel double mutation in RhCe gene on D-negative background. *Blood*, *92*(2), 664–671.
124. Salzer, U., Bacchelli, C., Buckridge, S., Pan-Hammarström, Q., Jennings, S., Lougaris, V., Bergbreiter, A., Hagen, T., Birmelin, J., Plebani, A., Webster, A. D., Peter, H. H., Suez, D., Chapel, H., McLean-Tooke, A., Spickett, G. P., Anover-Sombke, S., Ochs, H. D., Urschel, S., Belohradsky, B. H., ... Grimbacher, B. (2009). Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes. *Blood*, *113*(9), 1967–1976.
125. Went, M., Duran-Lozano, L., Halldorsson, G. H., Gunnell, A., Ugidos-Damboriena, N., Law, P., Ekdahl, L., Sud, A., Thorleifsson, G., Thodberg, M., Olafsdottir, T., Lamarca-Arrizabalaga, A., Cafaro, C., Niroula, A., Ajore, R., Lopez de Lapuente Portilla, A., Ali, Z., Pertesi, M., Goldschmidt, H., Stefansdottir, L., ... Nilsson, B. (2024). Deciphering the genetics and mechanisms of predisposition to multiple myeloma. *Nature communications*, *15*(1), 6644.
126. Kyrtsolis, M. C., Sarris, K., Koulieris, E., Maltezas, D., Nikolaou, E., Angelopoulou, M. K., Bartzis, V., Tzenou, T., Dimou, M., Siakandaris, M. P., Viniou, N. A., Sachanas, S., Kalpadakis, C., Sfikakis, P. P., Pangalis, G. A., & Panayiotidis, P. (2014). Serum soluble TACI, a BlyS receptor, is a powerful prognostic marker of outcome in chronic lymphocytic leukemia. *BioMed research international*, *2014*, 159632.
127. Hinterleitner, C., Zhou, Y., Tandler, C., Heitmann, J. S., Kropp, K. N., Hinterleitner, M., Koch, A., Hartkopf, A. D., Zender, L., Salih, H. R., & Maurer, S. (2021). Platelet-Expressed TNFRSF13B (TACI) Predicts Breast Cancer Progression. *Frontiers in oncology*, *11*, 642170.
128. Bisgin, A., Sonmezler, O., Boga, I., & Yilmaz, M. (2021). The impact of rare and low-frequency genetic variants in common variable immunodeficiency (CVID). *Scientific reports*, *11*(1), 8308.
129. Ortiz-Aljaro, P., Montes-Cano, M. A., García-Lozano, J. R., Aquino, V., Carmona, R., Perez-Florida, J., García-Hernández, F. J., Dopazo, J., & González-Escribano, M. F. (2022). Protein and functional isoform levels and genetic variants of the BAFF and APRIL pathway components in systemic lupus erythematosus. *Scientific reports*, *12*(1), 11219.

130. Moura, R. A., Canhão, H., Polido-Pereira, J., Rodrigues, A. M., Navalho, M., Mourão, A. F., Resende, C., Campanilho-Marques, R., Madruga Dias, J., da Silva, J. A., Graca, L., & Fonseca, J. E. (2013). BAFF and TACI gene expression are increased in patients with untreated very early rheumatoid arthritis. *The Journal of rheumatology*, *40*(8), 1293–1302.
131. Verstappen, G. M., Ice, J. A., Bootsma, H., Pringle, S., Haacke, E. A., de Lange, K., van der Vries, G. B., Hickey, P., Vissink, A., Spijkervet, F. K. L., Lessard, C. J., & Kroese, F. G. M. (2020). Gene expression profiling of epithelium-associated FcRL4<sup>+</sup> B cells in primary Sjögren's syndrome reveals a pathogenic signature. *Journal of autoimmunity*, *109*, 102439.
132. Khalid, Z., Huan, M., Sohail Raza, M., Abbas, M., Naz, Z., Kombe Kombe, A. J., Zeng, W., He, H., & Jin, T. (2022). Identification of Novel Therapeutic Candidates Against SARS-CoV-2 Infections: An Application of RNA Sequencing Toward mRNA Based Nanotherapeutics. *Frontiers in microbiology*, *13*, 901848.
133. Zou, J., Zhang, G., Li, H., Zhao, Z., Zhang, Q., Pyykkö, I., & Mäkitie, A. (2023). Multiple genetic variants involved in both autoimmunity and autoinflammation detected in Chinese patients with sporadic Meniere's disease: a preliminary study. *Frontiers in neurology*, *14*, 1159658.
134. Kiryluk, K., Sanchez-Rodriguez, E., Zhou, X. J., Zanoni, F., Liu, L., Mladkova, N., Khan, A., Marasa, M., Zhang, J. Y., Balderes, O., Sanna-Cherchi, S., Bomback, A. S., Canetta, P. A., Appel, G. B., Radhakrishnan, J., Trimarchi, H., Sprangers, B., Cattran, D. C., Reich, H., Pei, Y., ... Gharavi, A. G. (2023). Genome-wide association analyses define pathogenic signaling pathways and prioritize drug targets for IgA nephropathy. *Nature genetics*, *55*(7), 1091–1105.
135. Ahmed, Z., Renart, E. G., Zeeshan, S., & Dong, X. (2021). Advancing clinical genomics and precision medicine with GViZ: FAIR bioinformatics platform for variable gene-disease annotation, visualization, and expression analysis. *Human genomics*, *15*(1), 37.
136. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, *27*(15), 2156–2158.
137. Ahmed, Z., Renart, E. G., Mishra, D., & Zeeshan, S. (2021). JWES: a new pipeline for whole genome/exome sequence data processing, management, and gene-variant discovery, annotation, prediction, and genotyping. *FEBS open bio*, *11*(9), 2441–2452.

138. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–1760.
139. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297–1303.
140. Cingolani P. (2022). Variant Annotation and Functional Prediction: SnpEff. *Methods in molecular biology* (Clifton, N.J.), 2493, 289–314.
141. Ahmed, Z., Zeeshan, S., & Liang, B. T. (2021). RNA-seq driven expression and enrichment analysis to investigate CVD genes with associated phenotypes among high-risk heart failure patients. *Human genomics*, 15(1), 67.
142. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research Res.* 12, 2825–2830.
143. Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H., Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L., Lander, E. S., & Engreitz, J. M. (2019). Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nature genetics*, 51(12), 1664–1669.
144. Ding, C., & Peng, H. (2003). Minimum redundancy feature selection from microarray gene expression data. Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003, 523–528.
145. Radovic, M., Ghalwash, M., Filipovic, N., & Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC bioinformatics*, 18(1), 9.
146. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
147. Muzellec, B., Teleńczuk, M., Cabeli, V., & Andreux, M. (2023). PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics (Oxford, England)*, 39(9), btad547.

148. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550.
149. Becht, E., McInnes, L., Healy, J., Dutertre, C. A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., & Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature biotechnology*, 10.1038/nbt.4314. Advance online publication.
150. Zhang, X., Liu, H., & Zhang, X. (2017). Novel density-based and hierarchical density-based clustering algorithms for uncertain data. *Neural networks: the official journal of the International Neural Network Society*, *93*, 240–255.
151. Al-Shammari, A., Zhou, R., Naseriparsaa, M., & Liu, C. (2019). An effective density-based clustering and dynamic maintenance framework for evolving medical data streams. *International journal of medical informatics*, *126*, 176–186.
152. Fang, Z., Liu, X., & Peltz, G. (2023). GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics (Oxford, England)*, *39*(1), btac757.
153. Xie, Z., Bailey, A., Kuleshov, M. V., Clarke, D. J. B., Evangelista, J. E., Jenkins, S. L., Lachmann, A., Wojciechowicz, M. L., Kropiwnicki, E., Jagodnik, K. M., Jeon, M., & Ma'ayan, A. (2021). Gene Set Knowledge Discovery with Enrichr. *Current protocols*, *1*(3), e90.
154. Aung, N., Vargas, J. D., Yang, C., Cabrera, C. P., Warren, H. R., Fung, K., Tzanis, E., Barnes, M. R., Rotter, J. I., Taylor, K. D., Manichaikul, A. W., Lima, J. A. C., Bluemke, D. A., Piechnik, S. K., Neubauer, S., Munroe, P. B., & Petersen, S. E. (2019). Genome-Wide Analysis of Left Ventricular Image-Derived Phenotypes Identifies Fourteen Loci Associated With Cardiac Morphogenesis and Heart Failure Development. *Circulation*, *140*(16), 1318–1330.
155. Xiao, W., Ren, L., Chen, Z., Fang, L. T., Zhao, Y., Lack, J., Guan, M., Zhu, B., Jaeger, E., Kerrigan, L., Blomquist, T. M., Hung, T., Sultan, M., Idler, K., Lu, C., Scherer, A., Kusko, R., Moos, M., Xiao, C., Sherry, S. T., ... Shi, L. (2021). Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing. *Nature biotechnology*, *39*(9), 1141–1150.
156. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, *29*(1), 308–311.

157. Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., ... Maglott, D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic acids research*, 46(D1), D1062–D1067.

158. Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S. K., Boddu, S., Branco Lins, P. R., Brooks, L., Ramaraju, S. B., Campbell, L. I., Martinez, M. C., Charkhchi, M., Chougule, K., Cockburn, A., ... Yates, A. D. (2024). Ensembl 2024. *Nucleic acids research*, 52(D1), D891–D899.

160. Zhao, Y., Li, M. C., Konaté, M. M., Chen, L., Das, B., Karlovich, C., Williams, P. M., Evrard, Y. A., Doroshow, J. H., & McShane, L. M. (2021). TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. *Journal of translational medicine*, 19(1), 269.

160. Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., Pierleoni, A., Pignatelli, M., Platt, T., Rowland, F., Wankar, P., Bento, A. P., Burdett, T., Fabregat, A., Forbes, S., Gaulton, A., ... Dunham, I. (2017). Open Targets: a platform for therapeutic target identification and validation. *Nucleic acids research*, 45(D1), D985–D994.

161. Chakravarty, D., Gao, J., Phillips, S. M., Kundra, R., Zhang, H., Wang, J., Rudolph, J. E., Yaeger, R., Soumerai, T., Nissan, M. H., Chang, M. T., Chandarlapaty, S., Traina, T. A., Paik, P. K., Ho, A. L., Hantash, F. M., Grupe, A., Baxi, S. S., Callahan, M. K., Snyder, A., ... Schultz, N. (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO precision oncology*, 2017, 10.1200/PO.17.00011.

## Declarations

### Ethical Approval and Consent to participate:

Informed consent was obtained from all subjects. All human samples were used in accordance with relevant guidelines and regulations, and all experimental protocols were approved by the Institutional Review Board (IRB). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institution and the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

### Funding:

No funding was received.

### **Availability of data and material:**

The data that supports the findings of this study are available from the corresponding author on reasonable request.

### **Competing interests:**

The Authors declare no Competing Financial or Non-Financial Interests.

## **Acknowledgments**

This research was conducted with the great support of Department of Medicine, Robert Wood Johnson Medical School (RWJMS); and Rutgers Institute for Health, Health Care Policy, and Aging Research (IFH), Rutgers Health. We express sincere gratitude to the current and former members of Ahmed Lab at Rutgers RWJMS/IFH, and appreciate all colleagues, collaborators and institutions who provided direct and indirect insight and expertise that greatly assisted the research and development of this project, most importantly UMKC School of Medicine and UConn School of Medicine. We acknowledge Office of Advanced Research Computing (OARC) for having access to the High-Performance Computing environment and associated computing resources at Rutgers that have contributed to data analysis.

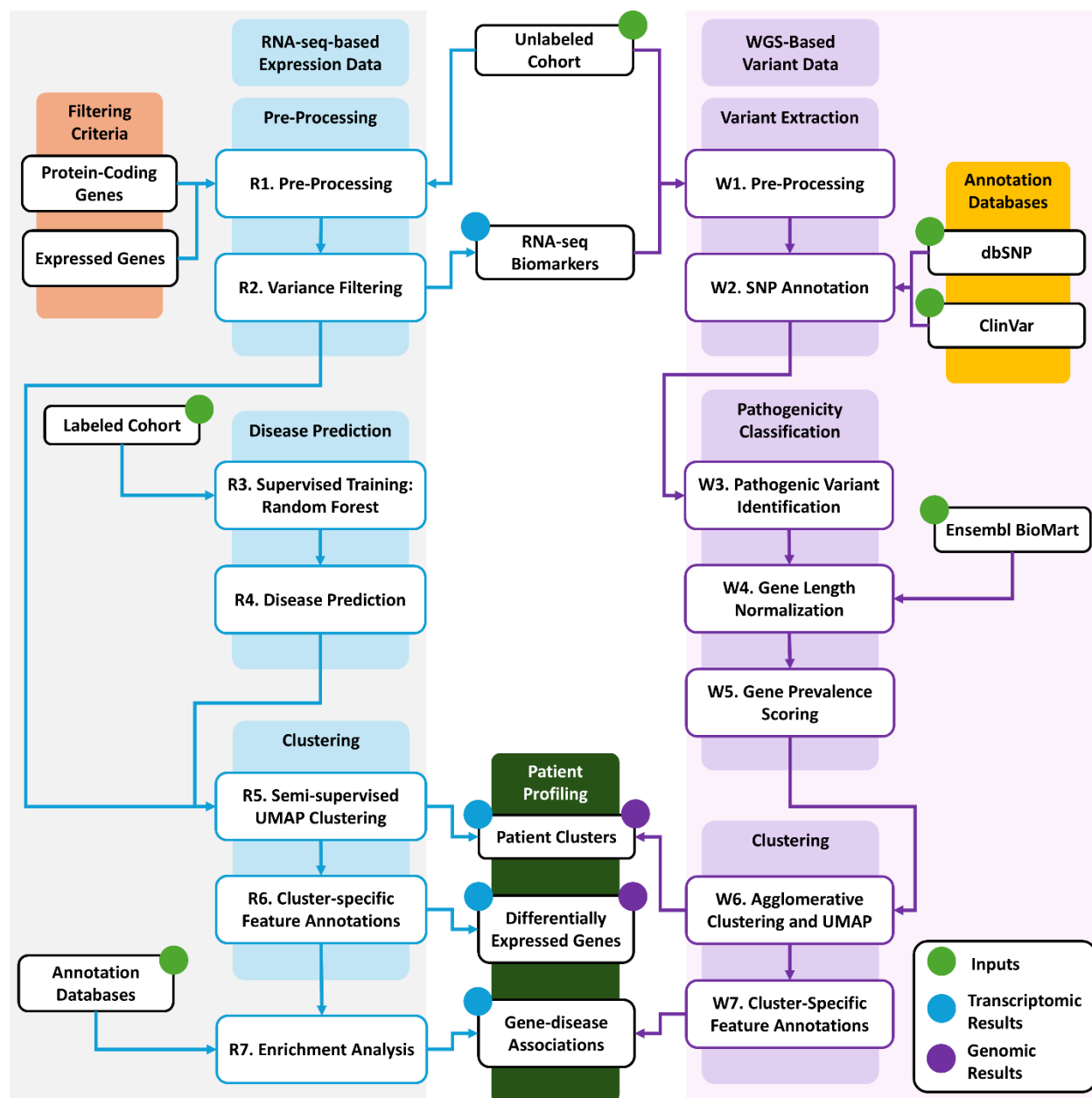
## **Author contributions**

Z.A. proposed, led, supervised, and drafted this study. Z.A. performed overall project conceptualization, administration, methodology, investigation, data duration, and required funds and resource allocation. Z.A. generated RNA-seq and WGS data, and performed multi-omics data processing, quality checking, and downstream analysis, including gene expression and variant discovery. P.G., W.D., and R.N., participated in predictive analysis, gene-disease annotation, and visualization. E.P. assisted with validation of results using available clinical data and state-of-science literature. S.Z. provided collaborative support and guided the study. All authors have contributed to writing and approved manuscript for publication.

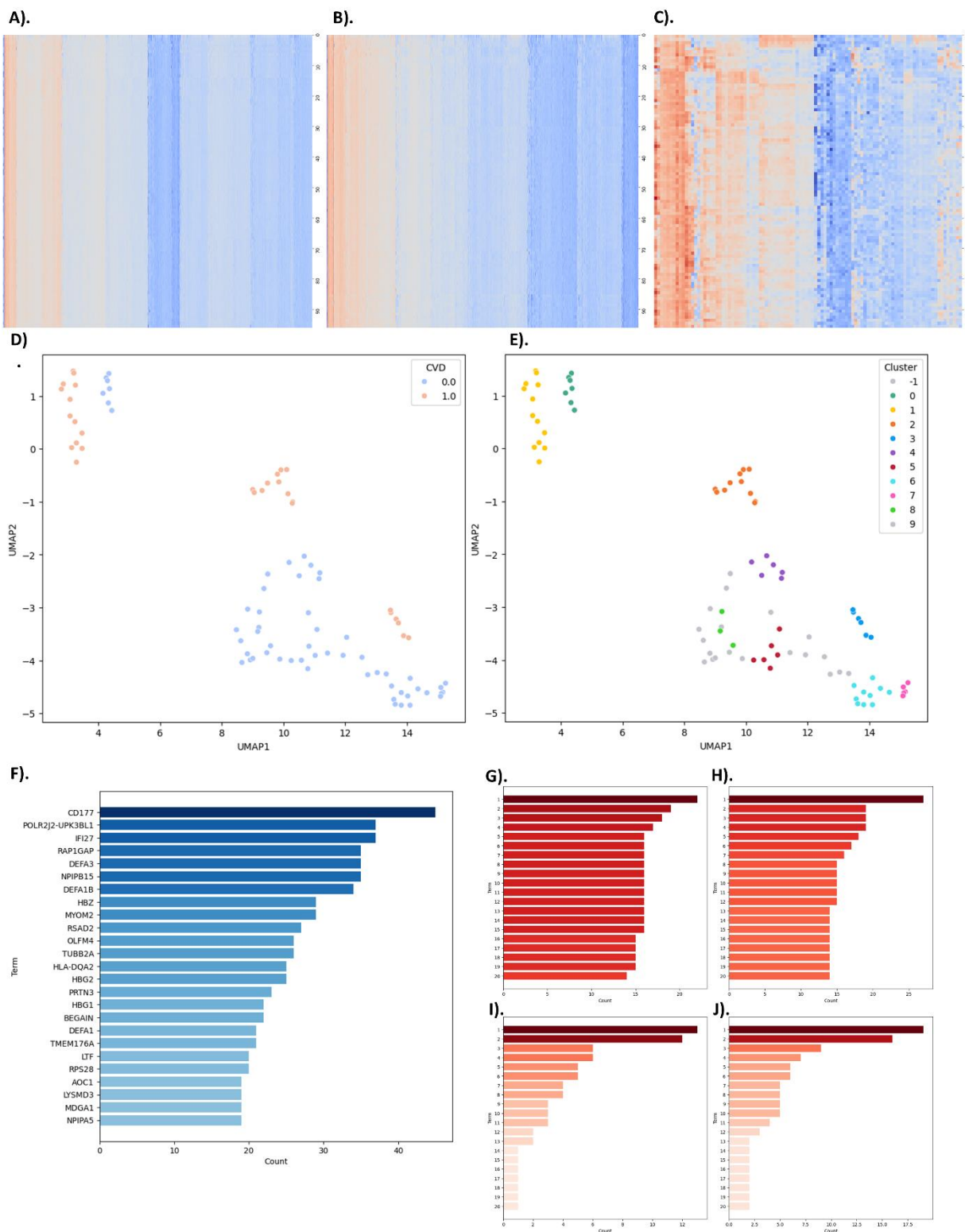
## **Biographical note**

Z.A. is the Assistant Professor at the Department of Medicine, Rutgers Robert Wood Johnson Medical School, and Core Faculty Member at the Rutgers Institute for Health, Health Care Policy and Aging Research, Rutgers Health. NJ. P.G., W.D., R.N., and E.P., are the Research Assistants/Students at the Ahmed lab, Rutgers IFH/RWJMS. SZ is the Director of UMKC Lab for Investigating Oncology Networks and Systems (LIONS), and Research Assistant Professor at the Department of Biomedical and Health Informatics, at UMKC School of Medicine.

## Figures

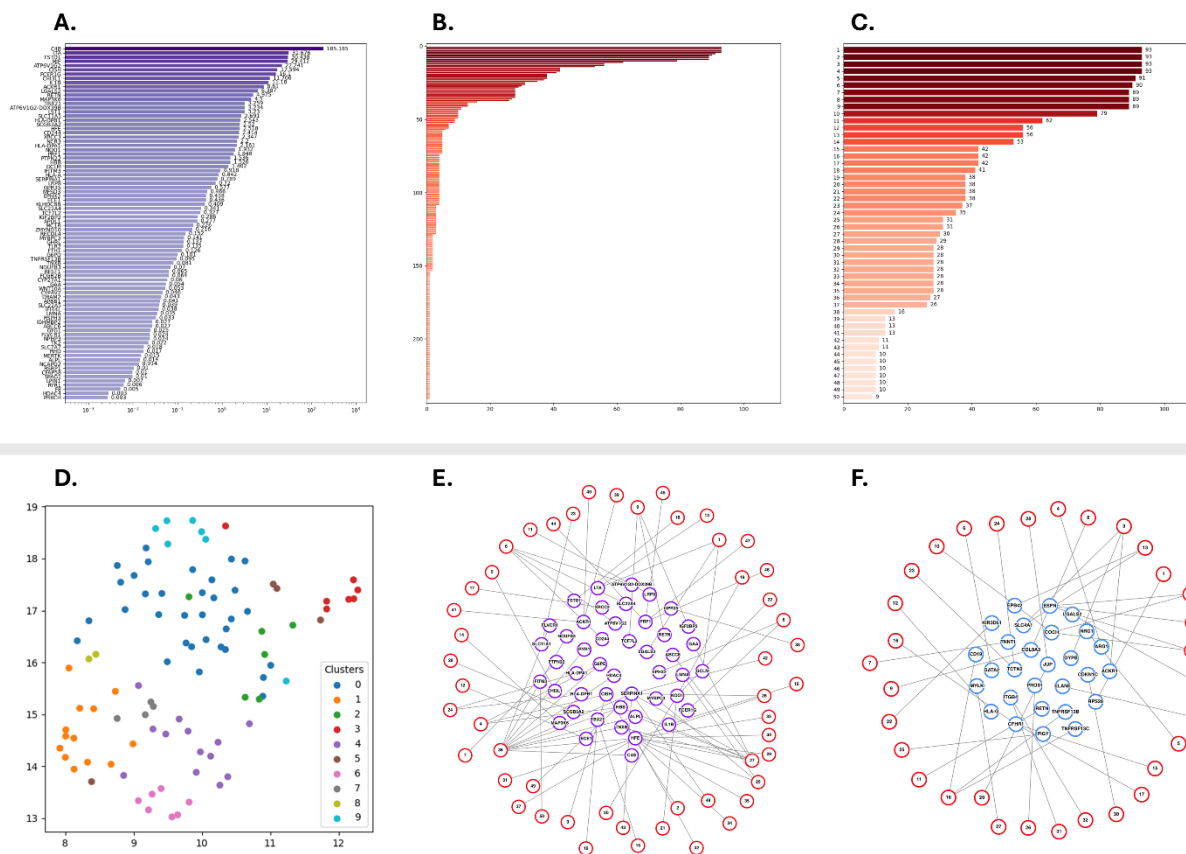


**Figure 1. Data analysis methodology, divided into two sections.** RNA-seq driven expression data analysis and WGS based variant data analysis.



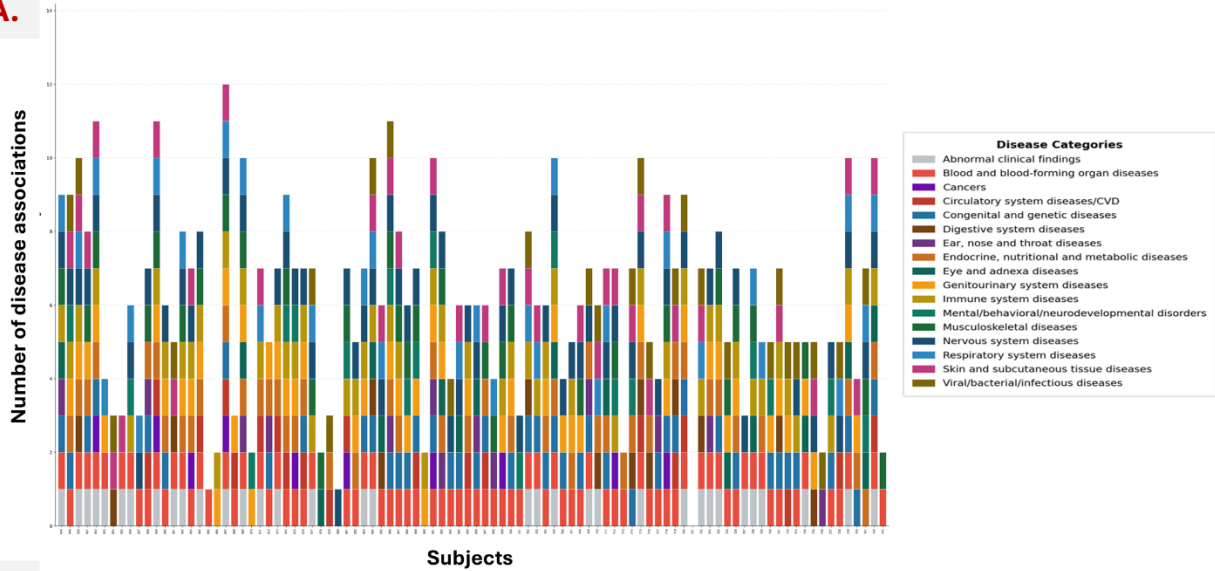
**Figure 2. RNA-seq driven gene-expression data analysis results. (A)** Heatmap containing all genes that passed RNA-seq preprocessing. **(B)** Heatmap of 5,000 variance-filtered genes utilized in clustering analysis.

**(C)** Heatmap of 100 variance-filtered genes. **(D)** UMAP embedding shows patients predicted to have CVD. **(E)** UMAP embedding showing HDBSCAN-generated clusters of patients. **(F)** Most occurrent genes in RNA-seq-based subject-specific profiles. **(G)** Most frequent annotations in GWAS Catalog. **(H)** Most frequent annotations in DisGeNET. **(I)** Most frequent annotations in OMIM Disease. **(J)** Most frequent annotations in ClinVar.

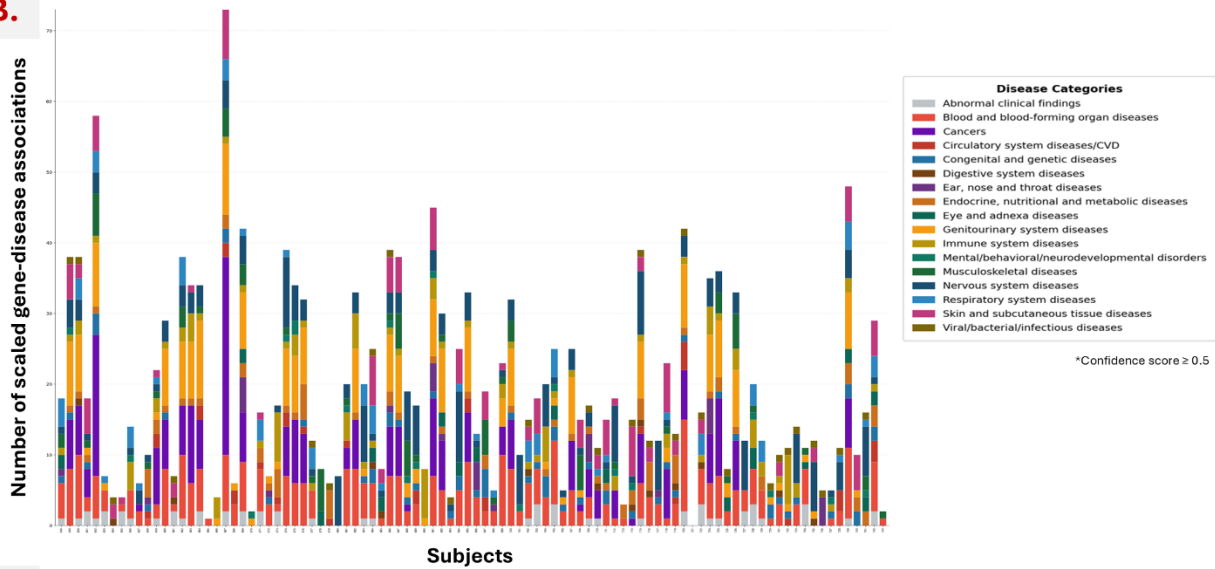


**Figure 3. WGS based gene-variant data analysis results. (A) Normalized gene counts. (B) Disease prevalence counts. (C) Top 50 Disease counts. (D) UMAP clustering. (E) WGS-based gene-disease network. (F) RNA-seq based gene-disease network.**

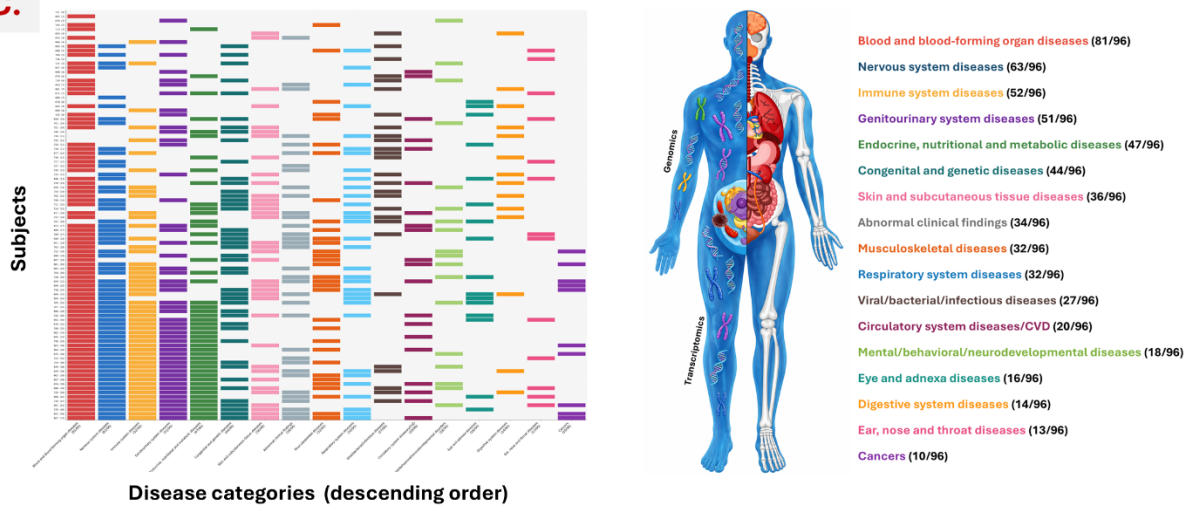
**A.**



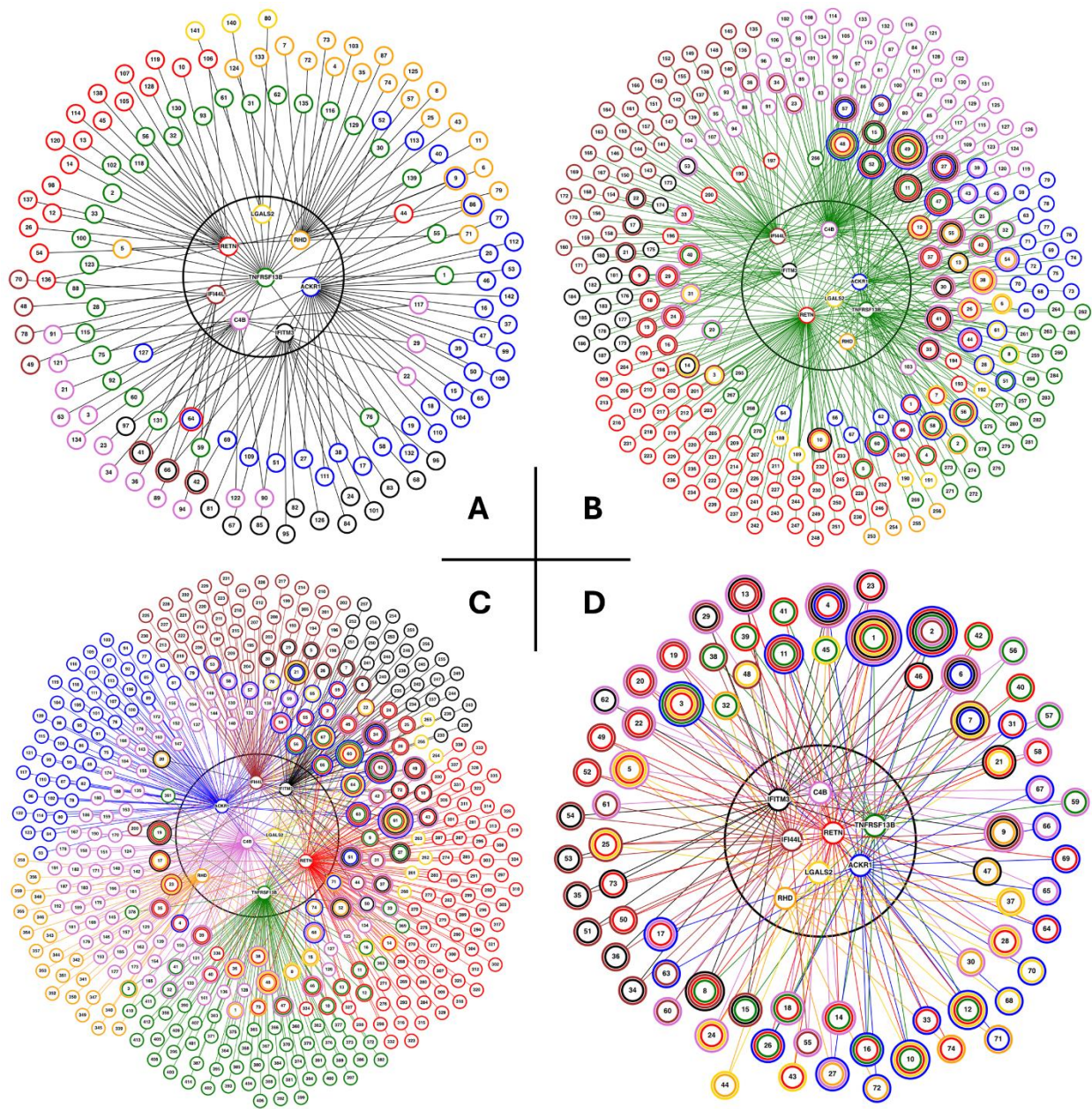
**B.**



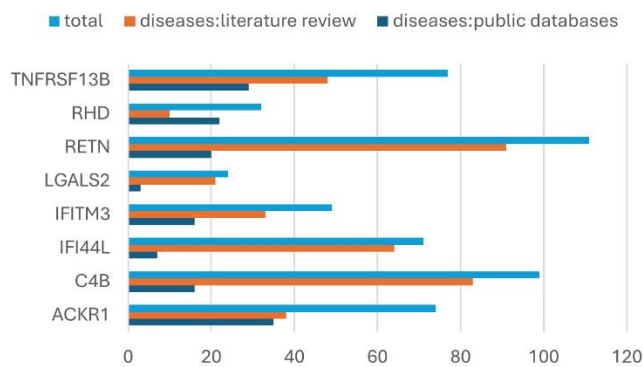
**C.**



**Figure 4.** Patient profiles, gene-disease associations. A) number of diseases in patients, B) number of scaled gene-diseases associations in descending order, and C) disease distribution across patient cohort. Each bar 4 represents individual patient (A and B), whereas each color box represents its association with a disease (A, B, and C).



Gene	Gene Disease Annotations Based on Public Databases	Gene Disease Annotations Based on Literature Review	Total
ACKR1	35	39	74
C4B	16	83	99
IFI44L	7	64	71
IFITM3	16	33	49
LGALS2	3	21	24
RETN	20	91	111
RHD	22	10	32
TNFRSF13B	29	48	77
Total	148	389	537



**Figure 5. Gene-disease association network.** (A) Gene-disease association network from the public annotation databases integrated in the methodology. (B) Gene-disease association network from the literature review conducted using NIH and PubMed databases. (C) Gene-disease associations from both annotation types. (D) Gene-disease associations only for diseases that had at least one and at most seven genes overlap of the eight mutually predictive genes.

## Tables

SID	Genes	Diseases	Annot-ations
648	<i>POLR2J2-UPK3BL1; HLA-DQB2; COL9A3; HLA-DQA2; RAP1GAP; HBG2; ELANE; SPP1; DEFA1; CEACAM6; CFAP298-TCP10L; CTSG; MARCO; PRTN3; ESPN; ATOH8; DEFA3; DEFA4; CD177; AOC1</i>	Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Stickler syndrome; Stickler syndrome, type 6; alpha 1-antitrypsin deficiency; bone disease; cyanosis, transient neonatal; cyclic hematopoiesis; emphysema; epiphyseal dysplasia, multiple, 3; hearing loss, autosomal recessive; neurodegenerative disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; osteoporosis; primary ciliary dyskinesia; severe congenital neutropenia	19
649	<i>RNASE4; OLFM4; LYSMD3; MDGA1; TUBB2A; NPIPA5; CD177; RAB4B-EGLN2; IFI44L; HDC; ZNF888; LTF; RSAD2; KRT1; DEFA4; MS4A2; TMTC1; IFITM3; CFHR1; HBG1</i>	Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Non-epidermolytic palmoplantar keratoderma; Transitional Cell Carcinoma; adenocarcinoma; amyotrophic lateral sclerosis; anemia; anemia (phenotype); annular epidermolytic ichthyosis; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; chronic kidney disease; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; ichthyosis hystrix of Curth-Macklin; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; schizophrenia; stomach neoplasm; tuberculosis	45
650	<i>DEFA1B; CD177; EIF3CL; MMP8; PNRC2; CEACAM6; KIR2DS4; PRTN3; DEFA3; LTF; SERPINB10; LGALS9C; ZDHHC11B; OLFM4; TUBB2A; ELANE; DEFA1; LCN2; HLA-DQB1; TNFRSF17</i>	Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; acne; adenocarcinoma; alpha 1-antitrypsin deficiency; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyclic hematopoiesis; diffuse large B-cell lymphoma; emphysema; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; infection; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; non-Hodgkins lymphoma; non-small cell	46

		lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; periodontitis; peritoneal neoplasm; prostate cancer; prostate neoplasm; rosacea; severe congenital neutropenia; stomach neoplasm; tuberculosis	
651	<i>HBZ; PVALB; OLFM4; MYOM2; BEGAIN; SMIM1; MYEOV; PRTN3; ETV7; NRG1; RPS28; TNNC2; ATOH8; RAP1GAP; NECTIN2; DEFA1B; RPH3A; DAAM2; KRT1; NAV3</i>	Abnormality of the skeletal system; Blackfan-Diamond anemia; Diamond-Blackfan anemia; Non-epidermolytic palmoplantar keratoderma; Umbilical hernia; annular epidermolytic ichthyosis; cancer; congenital myopathy 15; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; hypothyroidism; ichthyosis hystrix of Curth-Macklin; lung adenocarcinoma; nephrotic syndrome, type 24; neurodegenerative disease; papillary thyroid carcinoma; thyroid cancer; thyroid carcinoma	19
652	<i>RIOX1; LRRN3; KRT73; KIR2DL1; KRT72; KAZN; CD248; NOG; EDAR; ADTRP; FLT4; DAAM2; AK5; DEFA1B; NECTIN2; GCNT4; TUBB2A; PRRT1; RBM11; HOOK1</i>	Abnormality of the skeletal system; Genetic central nervous system malformation; Hodgkins lymphoma; Milroy disease; Transitional Cell Carcinoma; adenocarcinoma; autosomal dominant hypohidrotic ectodermal dysplasia; autosomal recessive hypohidrotic ectodermal dysplasia; brachydactyly type B2; breast cancer; breast carcinoma; breast neoplasm; cancer; capillary infantile hemangioma; cervical cancer; colorectal cancer; colorectal neoplasm; complex cortical dysplasia with other brain malformations 5; congenital heart defects, multiple types, 7; diffuse large B-cell lymphoma; ectodermal dysplasia 10A, hypohidrotic/hair/nail type, autosomal dominant; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gastrointestinal stromal tumor; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hepatocellular carcinoma; hypohidrotic ectodermal dysplasia; idiopathic pulmonary fibrosis; interstitial lung disease; liposarcoma; lymphatic malformation 1; lymphoma; medullary thyroid gland carcinoma; metastatic colorectal cancer; multiple myeloma; multiple synostoses syndrome 1; neoplasm; nephrotic syndrome, type 24; neurodegenerative disease; neuroendocrine neoplasm; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; proximal symphalangism 1A; pulmonary fibrosis; renal cell carcinoma; sarcoma; soft tissue sarcoma; stapes ankylosis with broad thumbs and toes; stomach neoplasm; systemic scleroderma; tarsal-carpal coalition syndrome; thyroid cancer; thyroid carcinoma; thyroid neoplasm; tooth agenesis	65
653	<i>DEFA1B; VMO1; TMEM176B; TMEM176A; LYSMD3; CD177; PNRC2; LYPD2; GP1BB; CFAP298-TCP10L; PCSK1N; RAB4B-EGLN2; SPATC1L; LGALS2;</i>	Bernard-Soulier syndrome; Macrothrombocytopenia; Thrombocytopenia; anemia; anemia (phenotype); chronic kidney disease; primary ciliary dyskinesia	7

	<i>AOC1; CDC42EP1; FFAR3; NICOL1; PTGDS; IFI27</i>		
654	<i>DEFA1B; PRTN3; RSPH10B2; CEACAM6; NPIP15; MMP8; CTSG; OLFM4; LTF; JCHAIN; RAP1GAP; DEFA1; RBM11; CD248; MACROD2; PTPRK; DEFA4; IFI27; DEFA3; POLR2J2-UPK3BL1</i>	acne; infection; periodontitis; rosacea; tuberculosis	5
655	<i>RIOX1; H2AC7; AOC1; NPIP15; IFITM3; IFI27; OLIG2; PCSK1N; MYOM2; C19orf33; KREMEN1; SMIM1; CD177; PI3; RSAD2; OLFM4; CD200; GP1BB; SLPI; SIGLEC8</i>	Bernard-Soulier syndrome; Macrothrombocytopenia; Thrombocytopenia; autosomal recessive hypohidrotic ectodermal dysplasia; infectious disease; viral disease	6
656	<i>DEFA1B; RIOX1; RNASE4; PCSK1N; DEFA3; CEACAM6; MARCO; PTPRK; ORM2; CYP2S1; SCAMP5; SIGLEC1; TMTC1; PMEPA1; KLRC3; ELANE; JCHAIN; C1QB; IFI44L; RSAD2</i>	Alpha-1-antitrypsin deficiency; C1Q deficiency; C1Q deficiency 1; Cyclic neutropenia; Decreased total neutrophil count; alpha 1-antitrypsin deficiency; amyotrophic lateral sclerosis; cyclic hematopoiesis; emphysema; immunodeficiency due to a classical component pathway complement deficiency; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; schizophrenia; severe congenital neutropenia	14
657	<i>POLR2J2-UPK3BL1; CFAP298-TCP10L; MDGA1; AOC1; C19orf33; BEGAIN; EPHB2; CD177; S100P; RNASE4; MYOM2; RSAD2; S100B; IFI27; RPS28; ATOH8; IFI44L; ETV7; TMEM176A; NLRP2</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; amyotrophic lateral sclerosis; medullary thyroid gland carcinoma; neurodegenerative disease; primary ciliary dyskinesia	6
658	<i>EIF3CL; CALHM6; HBG2; HBZ; HBG1; RAP1GAP; NPIP15; CBS; SMIM1; IFI27; ESPN; NPIPA5; XCL2; ATOH8; CRACR2B; PCSK1N; NICOL1; ELAPOR1; ORM1; TNNC2</i>	Classical homocystinuria; Hereditary persistence of fetal hemoglobin - beta-thalassemia; classic homocystinuria; congenital myopathy 15; cyanosis, transient neonatal; familial thoracic aortic aneurysm and aortic dissection; hearing loss, autosomal recessive; homocystinuria; hyperhomocysteinemia, thrombotic, cbs-related; neurodegenerative disease	10

659	<p><i>GOLGA80; DEFA3; NPIP15; HDC; CLEC18A; NRG1; CD177; HLA-DQB2; SPP1; CPA3; GATA2; DAAM2; MS4A2; KLC3; HBG1; ORM2; DNAAF11; IFI27; MYZAP; MGAM2</i></p>	<p>Abnormality of the skeletal system; Deafness - lymphedema - leukemia; GATA2 deficiency with susceptibility to MDS/AML; Hereditary persistence of fetal hemoglobin - beta-thalassemia; acute myeloid leukemia; atrial fibrillation; bone disease; cancer; cardiomyopathy, dilated, 2K; deafness-lymphedema-leukemia syndrome; genetic disorder; hypothyroidism; leukemia, acute myeloid, susceptibility to; lung adenocarcinoma; monocytopenia with susceptibility to infections; myelodysplastic syndrome; nephrotic syndrome, type 24; neurodegenerative disease; osteoporosis; papillary thyroid carcinoma; primary ciliary dyskinesia; prostate carcinoma; thyroid cancer; thyroid carcinoma; type 2 diabetes mellitus</p>	25
660	<p><i>HLA-DQA2; HLA-DQB2; DEFA3; NPIPA5; MDGA1; FCRL5; RGPLD1; LYSMD3; TUBB2A; DERL3; DEFA1B; MARCO; RGPLD2; TCL1A; TMOD4; DDX11; NECTIN2; ORM1; RPS28; TNFRSF17</i></p>	<p>Blackfan-Diamond anemia; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; Warsaw breakage syndrome; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; clonal hematopoiesis; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm</p>	36
661	<p><i>NPIP15; NPIPA5; OLFM4; ZC2HC1A; ELAPOR1; RAP1GAP; NPIP9; KLRC3; C19orf33; PRSS21; LTF; HLA-DQA2; MMP8; GP1BB; CEACAM6; CEACAM8; PGM5; APOBEC3B; DEFA1; LCN2</i></p>	<p>Bernard-Soulier syndrome; Macrothrombocytopenia; Thrombocytopenia; acne; infection; periodontitis; rosacea; tuberculosis</p>	8

662	<p><i>DEFA1B; RIOX1; CD177; CFAP298-TCP10L; PRTN3; PRSS33; AOC1; ORM2; IFI27; TNNT1; OLFM4; ORM1; TMEM176B; TMEM176A; TUBB2A; B3GAT1; ELANE; SLPI; CTSG; DEFA1</i></p>	<p>Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; alpha 1-antitrypsin deficiency; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyclic hematopoiesis; diffuse large B-cell lymphoma; emphysema; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; multiple myeloma; nemaline myopathy 5; nemaline myopathy 5B, autosomal recessive, childhood-onset; nemaline myopathy 5C, autosomal dominant; neoplasm; neurodegenerative disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; primary ciliary dyskinesia; prostate cancer; prostate neoplasm; severe congenital neutropenia; stomach neoplasm</p>	45
663	<p><i>H2AC7; NPIP15; TMEM176B; TMEM176A; LYPD2; MYOM2; HDC; TUBB2A; CD177; CLEC18B; AOC1; CFD; RSAD2; MS4A2; GATA2; IFI27; CPA3; H2AC25; B3GAT1; NECTIN2</i></p>	<p>Deafness - lymphedema - leukemia; GATA2 deficiency with susceptibility to MDS/AML; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; acute myeloid leukemia; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; deafness-lymphedema-leukemia syndrome; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; genetic disorder; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; infectious disease; leukemia, acute myeloid, susceptibility to; liposarcoma; lymphoma; monocytopenia with susceptibility to infections; multiple myeloma; myelodysplastic syndrome; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate carcinoma; prostate neoplasm; recurrent Neisseria infections due to factor D deficiency; stomach neoplasm; viral disease</p>	44

664	<p><i>GP1BB; DEFA1B; PF4V1; EGF; GP9; SPOCD1; MYEOV; ALOX12; MYL9; PEAR1; TUBB2A; TREML1; GYPB; CMTM5; SPX; ENKUR; ABLIM3; SPARC; PPBP; MYLK</i></p>	<p>Bernard-Soulier syndrome; Genetic central nervous system malformation; Hodgkins lymphoma; Macrothrombocytopenia; Thrombocytopenia; Transitional Cell Carcinoma; adenocarcinoma; aortic aneurysm, familial thoracic 7; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; familial thoracic aortic aneurysm and aortic dissection; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; megacystis-microcolon-intestinal hypoperistalsis syndrome; megacystis-microcolon-intestinal hypoperistalsis syndrome 1; megacystis-microcolon-intestinal hypoperistalsis syndrome 4; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; osteogenesis imperfecta; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm</p>	41
665	<p><i>DEFA1B; RIOX1; POLR2J2-UPK3BL1; HLA-DQB2; PRTN3; CES1; ARPIN; MDGA1; PDK4; AOC1; BEGAIN; NLRP2; JCHAIN; KIR2DS4; PTPRM; DEFA3; HLA-DQA2; OLFM4; TNFRSF17; IGLL5</i></p>	<p>multiple myeloma; neoplasm</p>	2
666	<p><i>POLR2J2-UPK3BL1; NPIPA5; TMEM176B; NPIP15; LYSM3; TMEM176A; C1QB; VMO1; NECTIN2; RGP1; DAAM2; CMBL; CES1; IL1R2; DEFA3; CTSG; ZDHHC11B; NPIP6; MYOM2; FKBP5</i></p>	<p>C1Q deficiency; C1Q deficiency 1; immunodeficiency due to a classical component pathway complement deficiency; nephrotic syndrome, type 24</p>	4
667	<p><i>NPIP15; RAB4B-EGLN2; AOC1; TUBB2A; MYOM2; PVALB; CLEC18A; KIR2DS4; RAP1GAP; MYZAP; NRG1; DAAM2; KIR3DL1; ITGB4; SPX; IDO1; TPST1; PDGFRB; KIR2DL1; LGALS9B</i></p>	<p>Abnormality of the skeletal system; Basal ganglia calcification; Generalized junctional epidermolysis bullosa, non-Herlitz type; Genetic central nervous system malformation; Hodgkins lymphoma; Hypereosinophilic syndrome; Junctional epidermolysis bullosa - pyloric atresia; Localized epidermolysis bullosa simplex; Myelodysplastic/Myeloproliferative Neoplasm; Transitional Cell Carcinoma; acroosteolysis-keloid-like lesions-premature aging syndrome; acute lymphoblastic leukemia; acute myeloid leukemia; adenocarcinoma; anemia; anemia (phenotype); atrial fibrillation; bilateral striopallidodentate calcinosis; breast cancer; breast carcinoma; breast neoplasm; cancer; cardiomyopathy, dilated, 2K; cervical cancer; chronic kidney disease; chronic myelogenous leukemia; chronic</p>	80

		myeloproliferative disorder; colorectal cancer; colorectal neoplasm; complex cortical dysplasia with other brain malformations 5; dermatofibrosarcoma protuberans; diffuse large B-cell lymphoma; epidermolysis bullosa simplex 1C, localized; epidermolysis bullosa, junctional 5A, intermediate; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gastrointestinal stromal tumor; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hepatocellular carcinoma; hypothyroidism; idiopathic pulmonary fibrosis; infantile myofibromatosis; interstitial lung disease; junctional epidermolysis bullosa; junctional epidermolysis bullosa with pyloric atresia; junctional epidermolysis bullosa, non-Herlitz type; liposarcoma; lung adenocarcinoma; lymphoma; metastatic colorectal cancer; multiple myeloma; myelodysplastic/myeloproliferative disease; myofibromatosis, infantile, 1; neoplasm; nephrotic syndrome, type 24; neurodegenerative disease; neuroendocrine neoplasm; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; papillary thyroid carcinoma; peritoneal neoplasm; prostate cancer; prostate neoplasm; pulmonary fibrosis; renal cell carcinoma; sarcoma; skeletal overgrowth-craniofacial dysmorphism-hyperelastic skin-white matter lesions syndrome; soft tissue sarcoma; stomach neoplasm; systemic mastocytosis; systemic scleroderma; thyroid cancer; thyroid carcinoma	
668	<i>ZNF888; NPIP15; NPIPA5; HBZ; LRRN3; KLRC4; ZNF683; CD200; RGPD6; BEGAIN; TCL1A; PDE9A; IL9R; PTPRM; ELAPOR1; RSAD2; RAB4B-EGLN2; CLEC17A; CD248; MACROD2</i>	anemia; anemia (phenotype); chronic kidney disease; clonal hematopoiesis; coronary artery disease; stroke	6
669	<i>KIR2DL1; TUBB2A; PI3; CD177; ALAS2; CHST13; GYPB; CA1; SMIM1; SLPI; HBZ; GP1BB; TNNT1; RAP1GAP; NECTIN2; ESPN; COCH; RGPD1; APOBEC3B; CLEC18B</i>	Bernard-Soulier syndrome; Genetic central nervous system malformation; Hodgkins lymphoma; Macrothrombocytopenia; Seizure; Thrombocytopenia; Transitional Cell Carcinoma; X-linked erythropoietic protoporphyria; X-linked sideroblastic anemia 1; adenocarcinoma; altitude sickness; angle-closure glaucoma; autosomal dominant nonsyndromic hearing loss; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; deafness; diffuse large B-cell lymphoma; edema; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; glaucoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hearing loss, autosomal recessive; hearing loss, autosomal recessive 110; liposarcoma; lymphoma; multiple myeloma; nemaline myopathy 5; nemaline myopathy 5B, autosomal recessive, childhood-onset; nemaline myopathy 5C, autosomal dominant;	50

		neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; nonsyndromic genetic hearing loss; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm	
670	<i>LYSMD3; MDGA1; GOLGA80; CD177; LGALS9C; S100B; HBZ; IFI44L; DNAJC25; CLEC12B; TMTC1; DAAM2; DEFA1B; KLRC4; RAP1GAP; CHIT1; GPR84; TMEM176A; CLEC12A; NUTM2B</i>	nephrotic syndrome, type 24; schizophrenia	2
671	<i>H2AC7; CD177; PNRC2; PVALB; S100B; KLRC2; SMIM1; HP; BAIAP3; ELANE; TMEM176A; NECTIN2; PRTN3; TMEM176B; RAP1GAP; OLFM4; RPS28; KREMEN1; DEFA1; GOS2</i>	Alpha-1-antitrypsin deficiency; Blackfan-Diamond anemia; Cyclic neutropenia; Decreased total neutrophil count; Diamond-Blackfan anemia; Hypercholesterolemia; alpha 1-antitrypsin deficiency; autosomal recessive hypohidrotic ectodermal dysplasia; coronary artery disease; cyclic hematopoiesis; emphysema; hyperlipidemia; infectious disease; metabolic disease; metabolic syndrome; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; severe congenital neutropenia; viral disease	19
672	<i>MYOM2; PNRC2; NPIP15; NPIP5; HBG1; IFI27; CES1; HBG2; TCL1A; ESPN; HBZ; HLA-DQB1; ALAS2; CLEC17A; KRT72; ST6GALNAC1; PLEK2; DAAM2; SMIM1; GPR15</i>	Hereditary persistence of fetal hemoglobin - beta-thalassemia; X-linked erythropoietic protoporphyria; X-linked sideroblastic anemia 1; clonal hematopoiesis; cyanosis, transient neonatal; hearing loss, autosomal recessive; nephrotic syndrome, type 24	7
673	<i>GP1BB; POLR2J2-UPK3BL1; MDGA1; CBS; NPIP15; VMO1; ATOH8; RAP1GAP; BEGAIN; SMIM1; DEFA1B; IFI27; AOC1; TGIF2-RAB5IF; RNASE3; NPIP5; RSAD2; PF4V1; SERPING1; DEFA3</i>	Bartholin gland disease; Bernard-Soulier syndrome; C1 inhibitor deficiency; Classical homocystinuria; Macrothrombocytopenia; Thrombocytopenia; angioedema; classic homocystinuria; familial thoracic aortic aneurysm and aortic dissection; genetic disorder; hereditary angioedema; hereditary angioedema type 1; hereditary angioedema type 2; hereditary angioedema type 3; hereditary angioedema with C1Inh deficiency;	18

		homocystinuria; hyperhomocysteinemia, thrombotic, cbs-related; neurodegenerative disease	
674	<i>TGIF2-RAB51F; KIR2DS4; TFF3; HLA-DQA2; MYOM2; BEGAIN; ATOH8; TUBB2A; CACNG6; GGT5; IFI44L; OLIG2; PRSS33; IL5RA; SMPD3; PTGDR2; RSAD2; RPL9; SLC29A1; RPH3A</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hodgkins lymphoma; Seizure; Transitional Cell Carcinoma; adenocarcinoma; anxiety disorder; asthma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; coronary artery disease; diffuse large B-cell lymphoma; epilepsy; familial Mediterranean fever; fibromyalgia; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; multiple myeloma; neoplasm; neuralgia; neurodegenerative disease; neuropathic pain; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; postherpetic neuralgia; prostate cancer; prostate neoplasm; restless legs syndrome; spinal cord injury; stomach neoplasm; stroke	46
675	<i>GOLGA8Q; TMTTC1; RAP1GAP; POLR2J2-UPK3BL1; C4B; JCHAIN; HLA-DQA2; TMCC2; TMEM176B; C4A; TMEM176A; SLC6A9; HBG1; IGLL5; CLEC17A; H2AC25; TUBB2A; RETN; PAX5; BCAM</i>	Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; T-cell acute lymphoblastic leukemia; Transitional Cell Carcinoma; acute lymphoblastic leukemia; adenocarcinoma; atypical glycine encephalopathy; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complement component 4a deficiency; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; glycine encephalopathy; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; immunodeficiency due to a classical component pathway complement deficiency; infectious disease; liposarcoma; lymphoma; major depressive disorder; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; schizophrenia; stomach neoplasm; viral disease	43

676	<i>POLR2J2-UPK3BL1; CBS; NPIP15; HBG1; RAP1GAP; HBG2; TUBB2A; ATOH8; DEFA3; ORM2; ELAPOR1; APOBEC3B; CD177; RSAD2; FCRL5; NECTIN2; IFITM3; HLA-DQB1; ENTPD2; KIR2DL1</i>	Classical homocystinuria; Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; classic homocystinuria; complex cortical dysplasia with other brain malformations 5; cyanosis, transient neonatal; diffuse large B-cell lymphoma; familial Mediterranean fever; familial thoracic aortic aneurysm and aortic dissection; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; homocystinuria; hyperhomocysteinemia, thrombotic, cbs-related; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm	39
677	<i>DEFA1B; SIGLEC14; RNASE4; CEACAM6; SPP1; CD177; NPIP15; RAP1GAP; DEFA3; CLEC18A; DEFA4; DEFA1; CEACAM8; LTF; MYOM2; CTSG; ELANE; KAZN; CHIT1; ALOX15</i>	Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; alpha 1-antitrypsin deficiency; amyotrophic lateral sclerosis; bone disease; cyclic hematopoiesis; emphysema; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; osteoporosis; severe congenital neutropenia; tuberculosis	13
678	<i>H2AC7; DEFA3; COL9A3; MYOM2; EIF3CL; MDGA1; IFI27; PNRC2; PCSK1N; HBZ; CD248; ORM2; RSAD2; NOG; HES4; IFI44L; RORC; SERPINE2; LGALS9C; IGFBP4</i>	Stickler syndrome; Stickler syndrome, type 6; brachydactyly type B2; epiphyseal dysplasia, multiple, 3; infectious disease; multiple synostoses syndrome 1; proximal symphalangism 1A; stapes ankylosis with broad thumbs and toes; tarsal-carpal coalition syndrome; viral disease	10
679	<i>HLA-DQB2; CD177; NPIPA5; GOLGA8Q; LYSMD3; NPIP15; DEFA3; BPI; LTF; CTSG; CEACAM6; LCN2; CBS; BEGAIN; RAP1GAP; HLA-DQA2; S100B; DEFA1; SERPINB10; AZU1</i>	Classical homocystinuria; classic homocystinuria; familial thoracic aortic aneurysm and aortic dissection; homocystinuria; hyperhomocysteinemia, thrombotic, cbs-related; tuberculosis	6
680	<i>HLA-DQA2; DEFA1B; POLR2J2-UPK3BL1; DEFA3; HLA-DQB2; SMN2; CTSG; OLFM4; S100P; CEACAM6; DEFA4; NPIP15; CALHM6; TMEM176B; DEFA1; BPI; PRTN3; TMEM176A; RNASE3; LCN2</i>	Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; neurodegenerative disease; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III	7

681	<i>DEFA1B; RIOX1; PRTN3; TNFRSF13C; PNRC2; TCL1A; CD200; FCRL5; HLA-DQA2; FCER2; CD19; LCN10; CD22; PAX5; FCRLA; BLK; VPREB3; CD79A; RPS28; FCRL2</i>	B-cell acute lymphoblastic leukemia; Blackfan-Diamond anemia; Diamond-Blackfan anemia; Mantle cell lymphoma; T-cell acute lymphoblastic leukemia; acute lymphoblastic leukemia; chronic myelogenous leukemia; clonal hematopoiesis; diffuse large B-cell lymphoma; follicular lymphoma; hairy cell leukemia; immunodeficiency, common variable, 3; immunodeficiency, common variable, 4; isolated agammaglobulinemia; major depressive disorder; mucocutaneous lymph node syndrome; neoplasm; neurodegenerative disease; neuromyelitis optica; rheumatoid arthritis; systemic lupus erythematosus	21
682	<i>EIF3CL; FCRL5; ARPIN; LGALS9C; HBZ; TUBB2A; MACROD2; RPS28; TCL1A; SERF1A; IGLL5; C1QB; DAAM2; TNFRSF17; ORM2; JCHAIN; CD200; IFI27; CD79A; FCER2</i>	Blackfan-Diamond anemia; C1Q deficiency; C1Q deficiency 1; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; clonal hematopoiesis; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; immunodeficiency due to a classical component pathway complement deficiency; isolated agammaglobulinemia; liposarcoma; lymphoma; multiple myeloma; neoplasm; nephrotic syndrome, type 24; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm	40
683	<i>DEFA1B; POLR2J2-UPK3BL1; RNASE4; DEFA1; EIF3CL; ORM1; OLFM4; PRTN3; DEFA4; RAP1GAP; RNASE3; PGM5; MYOM2; CEACAM6; DAAM2; SMN2; HBG2; NPIPA5; CD177; ELANE</i>	Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; alpha 1-antitrypsin deficiency; amyotrophic lateral sclerosis; cyanosis, transient neonatal; cyclic hematopoiesis; emphysema; nephrotic syndrome, type 24; neurodegenerative disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; severe congenital neutropenia; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III	20
684	<i>H2AC7; DEFA3; OLFM4; LTF; MMP8; DEFA1; CEACAM6; DEFA1B; DEFA4; NPIP15; ORM2; CD177; CEACAM8; CTSG; TMEM176B; KRT1; CRISP3; ELANE; CA1; HBG2</i>	Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Non-epidermolytic palmoplantar keratoderma; Seizure; acne; alpha 1-antitrypsin deficiency; altitude sickness; angle-closure glaucoma; annular epidermolytic ichthyosis; cyanosis, transient neonatal; cyclic hematopoiesis; diffuse nonepidermolytic palmoplantar keratoderma; edema; emphysema; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; glaucoma; ichthyosis hystrix of Curth-Macklin; infection; infectious disease;	29

		neutropenia; neutropenia, severe congenital, 1, autosomal dominant; periodontitis; rosacea; severe congenital neutropenia; tuberculosis; viral disease	
685	<i>RSPH10B2; NPIP15; CD177; BEGAIN; MYOM2; LYPD2; ZDHHC19; CA1; PNRC2; HLA-DQA2; NME2; KIR2DL1; MMP8; ORM1; SPATC1L; MCEMP1; BPGM; ZDHHC11B; S100A12; DEFA1</i>	Seizure; acne; altitude sickness; angle-closure glaucoma; edema; glaucoma; hemolytic anemia due to diphosphoglycerate mutase deficiency; infection; periodontitis; rosacea	10
686	<i>DEFA1B; HBG2; AOC1; HBG1; KRT1; TUBB2A; TMTC1; TFF3; ALAS2; RAP1GAP; PNRC2; KIR2DS4; EIF3CL; ESPN; GGT5; TMEM176A; TMEM176B; LTF; GYPB; MYL11</i>	Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Non-epidermolytic palmoplantar keratoderma; Transitional Cell Carcinoma; X-linked erythropoietic protoporphyria; X-linked sideroblastic anemia 1; adenocarcinoma; annular epidermolytic ichthyosis; arthrogryposis, distal, type 1C; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyanosis, transient neonatal; diffuse large B-cell lymphoma; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hearing loss, autosomal recessive; ichthyosis hystrix of Curth-Macklin; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; schizophrenia; stomach neoplasm; tuberculosis	46

687	<p><i>RIOX1; EIF3CL; LRRN3; PNR2; ARPIN; S100P; RPS28; NOG; TUBB2A; TMEM176B; GOLGA8Q; NPIPA5; CD248; HLA-DQA2; TMEM176A; KRT73; PPT2-EGFL8; EFCAB13; EDAR; LLCFC1</i></p>	<p>Blackfan-Diamond anemia; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; autosomal dominant hypohidrotic ectodermal dysplasia; autosomal recessive hypohidrotic ectodermal dysplasia; brachydactyly type B2; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; ectodermal dysplasia 10A, hypohidrotic/hair/nail type, autosomal dominant; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hypohidrotic ectodermal dysplasia; liposarcoma; lymphoma; multiple myeloma; multiple synostoses syndrome 1; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; proximal symphalangism 1A; psoriasis; stapes ankylosis with broad thumbs and toes; stomach neoplasm; tarsal-carpal coalition syndrome; tooth agenesis</p>	45
688	<p><i>CLEC4C; CD177; HBG1; HBG2; TMEM45B; MYL9; ENTPD2; PRSS33; VMO1; ZDHC19; TFF3; TCL1A; MYOM2; SPP1; ALOX15; OTX1; CCR3; TNNT1; CACNG6; SIGLEC8</i></p>	<p>Hereditary persistence of fetal hemoglobin - beta-thalassemia; Seizure; anxiety disorder; bone disease; clonal hematopoiesis; cyanosis, transient neonatal; epilepsy; fibromyalgia; megacystis-microcolon-intestinal hypoperistalsis syndrome 4; nemaline myopathy 5; nemaline myopathy 5B, autosomal recessive, childhood-onset; nemaline myopathy 5C, autosomal dominant; neuralgia; neuropathic pain; osteoporosis; postherpetic neuralgia; prostate cancer; prostate carcinoma; restless legs syndrome; spinal cord injury</p>	20
689	<p><i>DEFA3; HBG2; RPS28; PNR2; SMIM1; PVALB; PRSS33; CACNG6; ALOX15; ZNF888; RPL9; DEFA1B; CLC; RGPD1; MYZAP; SIGLEC8; HRH4; LYPD2; VSIG10; LAMB2</i></p>	<p>Blackfan-Diamond anemia; Diamond-Blackfan anemia; Hereditary persistence of fetal hemoglobin - beta-thalassemia; LAMB2-related infantile-onset nephrotic syndrome; Pierson syndrome; Seizure; anxiety disorder; atrial fibrillation; cardiomyopathy, dilated, 2K; cyanosis, transient neonatal; epilepsy; fibromyalgia; neuralgia; neuropathic pain; postherpetic neuralgia; restless legs syndrome; spinal cord injury</p>	17
690	<p><i>CD177; KRT73; KRT72; IFI27; IFI44L; RSAD2; SIGLEC1; PRTN3; IFIT1; RAP1GAP; MS4A2; DEFA3; RNASE3; FCRL5; FCRL1; ZBED6; DNAJC25; CMPK2; AGAP9; SERPING1</i></p>	<p>Bartholin gland disease; C1 inhibitor deficiency; angioedema; genetic disorder; hereditary angioedema; hereditary angioedema type 1; hereditary angioedema type 2; hereditary angioedema type 3; hereditary angioedema with C1Inh deficiency</p>	9

691	<p><i>NPIPA5; DEFA3; EIF3CL; MS4A2; PVALB; IGFBP4; HDC; IFI27; RAP1GAP; KRT1; TMTC1; GATA2; TUBB2A; APOBEC3B; COCH; HBG1; PRSS21; SLC45A3; FCER2; CPA3</i></p>	<p>Deafness - lymphedema - leukemia; GATA2 deficiency with susceptibility to MDS/AML; Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Non-epidermolytic palmoplantar keratoderma; Transitional Cell Carcinoma; acute myeloid leukemia; adenocarcinoma; annular epidermolytic ichthyosis; autosomal dominant nonsyndromic hearing loss; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; deafness; deafness-lymphedema-leukemia syndrome; diffuse large B-cell lymphoma; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; genetic disorder; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hearing loss, autosomal recessive 110; ichthyosis hystrix of Curth-Macklin; leukemia, acute myeloid, susceptibility to; liposarcoma; lymphoma; monocytopenia with susceptibility to infections; multiple myeloma; myelodysplastic syndrome; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; nonsyndromic genetic hearing loss; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate carcinoma; prostate neoplasm; schizophrenia; stomach neoplasm</p>	53
692	<p><i>ACKR1; LYSMD3; AOC1; SIGLEC8; COL9A3; PRSS33; OLIG2; TMEM176A; TMEM176B; ALOX15; ARPIN; CNTNAP3; EIF3CL; RAP1GAP; BEGAIN; SMIM1; TUBB2A; ESPN; SIGLEC5; DAAM2</i></p>	<p>Genetic central nervous system malformation; Hodgkins lymphoma; Stickler syndrome; Stickler syndrome, type 6; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; epiphyseal dysplasia, multiple, 3; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hearing loss, autosomal recessive; liposarcoma; lymphoma; multiple myeloma; neoplasm; nephrotic syndrome, type 24; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm</p>	37
693	<p><i>RIOX1; TGIF2-RAB5IF; H2AC7; NPIP15; NPIPA5; ZNF683; HBG2; HBG1; CD8B; DEFA3; IFI27; LRRN3; NFXL1; RAB6B; GOLGA8Q; CD8A; TGFB11; CMTM5; CD248; PPBP</i></p>	<p>HIV infection; Hereditary persistence of fetal hemoglobin - beta-thalassemia; cyanosis, transient neonatal; infectious disease; neurodegenerative disease; viral disease</p>	6

694	<p><i>POLR2J2-UPK3BL1; CFHR1; ZNF888; SMN2; ARPIN; MDGA1; DEFA3; NPIP15; CA1; HBD; KRT1; AHSP; ACKR1; APOOL; SLC6A9; HBG2; ITLN1; KLF1; MYOM2; GOLGA80</i></p>	<p>Congenital dyserythropoietic anemia type IV; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Non-epidermolytic palmoplantar keratoderma; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; Seizure; altitude sickness; angle-closure glaucoma; annular epidermolytic ichthyosis; atypical glycine encephalopathy; congenital dyserythropoietic anemia type 4; cyanosis, transient neonatal; delta-beta-thalassemia; diffuse nonepidermolytic palmoplantar keratoderma; edema; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; glaucoma; glycine encephalopathy; hereditary persistence of fetal hemoglobin-sickle cell disease syndrome; ichthyosis hystrix of Curth-Macklin; neurodegenerative disease; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III</p>	26
695	<p><i>RNASE4; DEFA3; NPIP15; MYZAP; MARCO; RPS28; LYSMD3; AOC1; LGALS9C; NSG1; TMEM176B; TMEM176A; ELAPOR2; ENTPD2; TMEM158; TUBB2A; RAB4B-EGLN2; APOBEC3B; RETN; ZNF683</i></p>	<p>Blackfan-Diamond anemia; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; amyotrophic lateral sclerosis; anemia; anemia (phenotype); atrial fibrillation; breast cancer; breast carcinoma; breast neoplasm; cancer; cardiomyopathy, dilated, 2K; cervical cancer; chronic kidney disease; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm</p>	40
696	<p><i>HBZ; TCL1A; OLFM4; IL9R; CD177; HBG1; RAP1GAP; AOC1; RSPH10B; FCER2; RPS28; HBG2; IFI27; CD200; CA1; APOBEC3B; LRRN3; GPR15; COCH; PVALB</i></p>	<p>Blackfan-Diamond anemia; Diamond-Blackfan anemia; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Seizure; altitude sickness; angle-closure glaucoma; autosomal dominant nonsyndromic hearing loss; clonal hematopoiesis; cyanosis, transient neonatal; deafness; edema; glaucoma; hearing loss, autosomal recessive 110; nonsyndromic genetic hearing loss</p>	14
697	<p><i>DEFA1B; POLR2J2-UPK3BL1; HLA-DQB2; NOG; CD177; TMEM176B; GOLGA8Q; TMEM176A; ARG1; LRRN3; CDRT4; AOC1; FAM153A; EDAR; RPS28; ZDHHC19; PDE9A; DACT1; BEGAIN; GALNT14</i></p>	<p>Argininemia; Blackfan-Diamond anemia; Diamond-Blackfan anemia; Townes-Brocks syndrome; Townes-Brocks syndrome 2; arginase deficiency; autosomal dominant hypohidrotic ectodermal dysplasia; autosomal recessive hypohidrotic ectodermal dysplasia; brachydactyly type B2; coronary artery disease; ectodermal dysplasia 10A, hypohidrotic/hair/nail type, autosomal dominant; hypohidrotic ectodermal dysplasia; multiple synostoses syndrome 1; proximal symphalangism 1A; spina bifida; stapes ankylosis with broad thumbs and toes; stroke; tarsal-carpal coalition syndrome; tooth agenesis</p>	19

698	<i>HLA-DQB2; MYL11; KIR2DL1; MDGA1; BEGAIN; CFAP298-TCP10L; BCAM; JCHAIN; KLRC3; PDZK1IP1; HLA-DQA2; RAB3IL1; POLR2J2-UPK3BL1; ESPN; SMIM24; ORM2; RPS28; KLC3; CD160; LRRN3</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; arthrogryposis, distal, type 1C; hearing loss, autosomal recessive; primary ciliary dyskinesia	5
699	<i>GOLGA80; PVALB; PNRC2; RHD; S100B; DAAM2; IFI27; NPIP15; GATA2; MYOM2; HLA-DQA2; BEGAIN; HDC; CPA3; PAGE2B; KLC3; CD248; SPP1; SLC4A1; PF4V1</i>	Congenital hemolytic anemia; Deafness - lymphedema - leukemia; Distal renal tubular acidosis with anemia; GATA2 deficiency with susceptibility to MDS/AML; Hereditary cryohydrocytosis with normal stomatin; acute myeloid leukemia; autosomal dominant distal renal tubular acidosis; bone disease; cryohydrocytosis; deafness-lymphedema-leukemia syndrome; distal renal tubular acidosis; genetic disorder; hematologic disease; hemolytic anemia; hereditary spherocytosis; hereditary spherocytosis type 4; inherited distal renal tubular acidosis; leukemia, acute myeloid, susceptibility to; monocytopenia with susceptibility to infections; myelodysplastic syndrome; nephrotic syndrome, type 24; neurodegenerative disease; osteoporosis; prostate carcinoma; renal tubular acidosis, distal, 4, with hemolytic anemia; southeast Asian ovalocytosis	26
700	<i>POLR2J2-UPK3BL1; CD177; KIR3DL1; MYOM2; ATOH8; RSAD2; RPS28; HBZ; IFI44L; HLA-DQA2; TNNT1; SIGLEC1; HBG2; KIR3DL2; IL32; IGFBP3; TUBB2A; IFI27; PF4V1; ERAP2</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyanosis, transient neonatal; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; multiple myeloma; nemaline myopathy 5; nemaline myopathy 5B, autosomal recessive, childhood-onset; nemaline myopathy 5C, autosomal dominant; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm	39
701	<i>HLA-DQB2; SMN2; HLA-DQA2; KRT73; KRT72; NPIP15; ZNF674; ENTPD2; TMTC1; MDGA1; HBG1; ATOH8; H2AC25; TKTL1; IFI27; JCHAIN; OLFM1; MS4A2; TNFRSF17; RAP1GAP</i>	Hereditary persistence of fetal hemoglobin - beta-thalassemia; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; infectious disease; multiple myeloma; neoplasm; neurodegenerative disease; schizophrenia; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III; viral disease	13

702	<i>DEFA1B; CD177; NPIP15; ORM2; HLA-DQB2; HBG1; HLA-DQA2; OLFM4; HBG2; ELANE; ORM1; PRTN3; IFI27; LTF; DEFA4; DEFA1; CRISP3; PF4V1; LCN2; MMP8</i>	Alpha-1-antitrypsin deficiency; Cyclic neutropenia; Decreased total neutrophil count; Hereditary persistence of fetal hemoglobin - beta-thalassemia; acne; alpha 1-antitrypsin deficiency; cyanosis, transient neonatal; cyclic hematopoiesis; emphysema; infection; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; periodontitis; rosacea; severe congenital neutropenia; tuberculosis	16
703	<i>GP1BB; DEFA1B; RIOX1; H2AC7; MDGA1; HBZ; S100B; BEGAIN; PNRC2; CD177; ELANE; ORM2; ORM1; CTSG; HLA-DQA2; BPI; CEACAM6; DEFA1; KRT1; DEFA3</i>	Alpha-1-antitrypsin deficiency; Bernard-Soulier syndrome; Cyclic neutropenia; Decreased total neutrophil count; Macrothrombocytopenia; Non-epidermolytic palmoplantar keratoderma; Thrombocytopenia; alpha 1-antitrypsin deficiency; annular epidermolytic ichthyosis; cyclic hematopoiesis; diffuse nonepidermolytic palmoplantar keratoderma; emphysema; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; ichthyosis hystrix of Curth-Macklin; infectious disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; severe congenital neutropenia; viral disease	20
704	<i>DEFA1B; POLR2J2-UPK3BL1; MDGA1; VMO1; LYPD2; KIR2DS4; S100P; NME2; B3GAT1; CD70; ATOH8; HBG2; MYOM2; AOC1; TMEM63C; TNFRSF13B; CTSG; PRTN3; RAB4B-EGLN2; CDKN1C</i>	Alzheimer disease; Beckwith-Wiedemann syndrome; Hereditary persistence of fetal hemoglobin - beta-thalassemia; IMAGE syndrome; Parkinson disease; anemia; anemia (phenotype); chronic kidney disease; common variable immunodeficiency; cyanosis, transient neonatal; hereditary spastic paraplegia; immunodeficiency, common variable, 1; immunodeficiency, common variable, 2; immunoglobulin A deficiency 2; lysosomal storage disease; multiple myeloma; multiple sclerosis; neurodegenerative disease; severe combined immunodeficiency due to CD70 deficiency; spastic paraplegia 87, autosomal recessive	20
705	<i>GP1BB; DEFA1B; HBZ; NPIP15; TMTC1; ELANE; IFI27; PCSK1N; PRTN3; RAB3IL1; SELENBP1; HBQ1; HBG2; CA1; ALAS2; DAAM2; MYOM2; EPB42; HBG1; NICOL1</i>	Alpha-1-antitrypsin deficiency; Bernard-Soulier syndrome; Cyclic neutropenia; Decreased total neutrophil count; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Macrothrombocytopenia; Seizure; Thrombocytopenia; X-linked erythropoietic protoporphyria; X-linked sideroblastic anemia 1; alpha 1-antitrypsin deficiency; altitude sickness; angle-closure glaucoma; cyanosis, transient neonatal; cyclic hematopoiesis; edema; emphysema; extraoral halitosis due to methanethiol oxidase deficiency; glaucoma; hereditary spherocytosis; nephrotic syndrome, type 24; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; recessive spherocytosis; schizophrenia; severe congenital neutropenia	26
706	<i>MYOM2; DEFA3; NPIPA5; SMIM1; RPS28; DEFA1B; ERAP2; IFI44L; CD177; DAAM2; PI3; RSAD2; DDX11; NPRL3; ISG15;</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; Warsaw breakage syndrome; familial focal epilepsy with variable foci; nephrotic syndrome, type 24	5

	<i>MSR1; IFIT1; MYEOV; RAP1GAP; GYPB</i>		
707	<i>POLR2J2-UPK3BL1; NPIP15; MDGA1; HBZ; TMEM176B; RSAD2; IFI27; TMEM176A; IFI44L; CD248; SIGLEC1; TUBB2A; CLEC17A; RAP1GAP; CD177; IFIT1; CD200; FCRL1; FCRL5; PTPRK</i>	Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm	32
708	<i>POLR2J2-UPK3BL1; LYSMD3; NOG; FAM153A; CES1; LRRN3; DEFA3; EDAR; NPIP15; KRT72; KRT73; CBLN3; RAB4B-EGLN2; PCSK1N; S100B; ADTRP; FAM153CP; AQP10; EPHA1; PRRT1</i>	Alzheimer disease; anemia; anemia (phenotype); autosomal dominant hypohidrotic ectodermal dysplasia; autosomal recessive hypohidrotic ectodermal dysplasia; brachydactyly type B2; chronic kidney disease; ectodermal dysplasia 10A, hypohidrotic/hair/nail type, autosomal dominant; hypohidrotic ectodermal dysplasia; medullary thyroid gland carcinoma; multiple synostoses syndrome 1; proximal symphalangism 1A; stapes ankylosis with broad thumbs and toes; tarsal-carpal coalition syndrome; tooth agenesis	15
709	<i>POLR2J2-UPK3BL1; HBZ; KIR2DS4; BEGAIN; AP1M2; CD177; COCH; MYOM2; IFI27; RPS28; PGM5; HBG1; NRG1; SMIM1; TCTN2; TFF3; VNN1; ERAP2; HLA-DQA2; PVALB</i>	Abnormality of the skeletal system; Blackfan-Diamond anemia; Diamond-Blackfan anemia; HIV infection; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Joubert syndrome; Joubert syndrome and related disorders; Meckel syndrome; autosomal dominant nonsyndromic hearing loss; cancer; deafness; hearing loss, autosomal recessive 110; hypothyroidism; lung adenocarcinoma; nonsyndromic genetic hearing loss; papillary thyroid carcinoma; thyroid cancer; thyroid carcinoma	18
710	<i>DEFA1B; RIOX1; ACKR1; PRTN3; H2AC7; DEFA1; DEFA4; OLFM4; DEFA3; CEACAM6; MMP8; BPI; LTF; NRG1; LGALS4; CEACAM8; RAP1GAP; ARPIN; SIGLEC1; DNAAF11</i>	Abnormality of the skeletal system; acne; cancer; hypothyroidism; infection; infectious disease; lung adenocarcinoma; papillary thyroid carcinoma; periodontitis; primary ciliary dyskinesia; rosacea; thyroid cancer; thyroid carcinoma; tuberculosis; viral disease	15

711	<i>RIOX1; POLR2J2-UPK3BL1; NPIP15; RAP1GAP; CALHM6; KRT73; KRT72; KRT1; HDC; CA1; HBZ; HBD; IFIT1B; MYOM2; NME2; AHSP; ALAS2; SELENBP1; OSBP2; BPGM</i>	Non-epidermolytic palmoplantar keratoderma; Seizure; X-linked erythropoietic protoporphyria; X-linked sideroblastic anemia 1; altitude sickness; angle-closure glaucoma; annular epidermolytic ichthyosis; delta-beta-thalassemia; diffuse nonepidermolytic palmoplantar keratoderma; edema; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; extraoral halitosis due to methanethiol oxidase deficiency; glaucoma; hemolytic anemia due to diphosphoglycerate mutase deficiency; ichthyosis hystrix of Curth-Macklin	16
712	<i>TFF3; NPIP15; TMEM176B; TMEM176A; HDC; MS4A2; CD177; AOC1; CPA3; CACNG6; MDGA1; GATA2; PRSS33; CCR3; S100B; HBZ; DEFA1B; CLC; PGM5; SIGLEC8</i>	Deafness - lymphedema - leukemia; GATA2 deficiency with susceptibility to MDS/AML; Seizure; acute myeloid leukemia; anxiety disorder; deafness-lymphedema-leukemia syndrome; epilepsy; fibromyalgia; genetic disorder; leukemia, acute myeloid, susceptibility to; monocytopenia with susceptibility to infections; myelodysplastic syndrome; neuralgia; neurodegenerative disease; neuropathic pain; postherpetic neuralgia; prostate carcinoma; restless legs syndrome; spinal cord injury	19
713	<i>RIOX1; POLR2J2-UPK3BL1; OLIG2; ALOX15; SIGLEC8; PRSS33; AOC1; IDO1; KLRC3; HBM; RTN1; LGALS9B; LYPD2; PGM5; CD177; BCAM; HBD; ERAP2; ARG1; AQP1</i>	Argininemia; arginase deficiency; delta-beta-thalassemia	3
714	<i>SLC25A43; DEFA1B; SIGLEC14; TMTC1; ORM2; OLFM4; KRT1; ARG1; IFI27; ATOH8; MYOM2; S100B; EPHB2; NPIPA5; PNRC2; CRISP3; LTF; GYPB; MMP8; PRTN3</i>	Argininemia; Non-epidermolytic palmoplantar keratoderma; acne; annular epidermolytic ichthyosis; arginase deficiency; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; ichthyosis hystrix of Curth-Macklin; infection; medullary thyroid gland carcinoma; neurodegenerative disease; periodontitis; rosacea; schizophrenia; tuberculosis	16
715	<i>NPIP15; SMN2; BEGAIN; PRTN3; IFI27; TFF3; OLFM4; HP; CTSG; CLEC18A; TUBB2A; RNASE3; RSAD2; LTF; C19orf33; GOLGA80; MMP8; KAZN; ETV7; DEFA4</i>	Genetic central nervous system malformation; Hodgkins lymphoma; Hypercholesterolemia; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; Transitional Cell Carcinoma; acne; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; coronary artery disease; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hyperlipidemia; infection; liposarcoma; lymphoma; metabolic disease; metabolic syndrome; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic	48

		neoplasm; periodontitis; peritoneal neoplasm; prostate cancer; prostate neoplasm; rosacea; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III; stomach neoplasm; tuberculosis	
716	<i>POLR2J2-UPK3BL1; NPIP15; OLFM4; LYSMD3; LYPD2; CD177; MMP8; MYOM2; CRISP3; LCN2; MS4A2; IL9R; LTF; RPS28; NECTIN2; RETN; CEACAM8; CAMP; TCN1; HDC</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; acne; deficiency anemia; infection; megaloblastic anemia; periodontitis; rosacea; transcobalamin I deficiency; tuberculosis; vitamin B deficiency; vitamin B12 deficiency; vitamin deficiency disorder	13
717	<i>RIOX1; RSPH10B; SMN2; KIR2DS4; HBG2; HBG1; MYOM2; AOC1; ALOX15; OLFM4; RPL9; UPK3A; ESPN; HLA-G; TMEM176A; SERF1A; RAP1GAP; JCHAIN; TMEM176B; KLRC2</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; cyanosis, transient neonatal; hearing loss, autosomal recessive; neurodegenerative disease; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III	12
718	<i>ORM2; HKDC1; HBZ; BEGAIN; DAAM2; DNAAF11; ORM1; CD177; KRT1; MS4A2; HDC; CPA3; MMP9; ZNF683; CES1; GATA2; BMX; OTX1; TGM3; DNAJC25</i>	Deafness - lymphedema - leukemia; GATA2 deficiency with susceptibility to MDS/AML; Non-epidermolytic palmoplantar keratoderma; acute myeloid leukemia; alopecia areata; annular epidermolytic ichthyosis; basal cell carcinoma; deafness-lymphedema-leukemia syndrome; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; genetic disorder; ichthyosis hystrix of Curth-Macklin; leukemia, acute myeloid, susceptibility to; metaphyseal anadysplasia; monocytopenia with susceptibility to infections; myelodysplastic syndrome; nephrotic syndrome, type 24; neurodegenerative disease; primary ciliary dyskinesia; prostate cancer; prostate carcinoma; skin cancer; skin neoplasm	24
719	<i>ORM2; DEFA3; DEFA1B; ORM1; AOC1; RAP1GAP; PRTN3; BEGAIN; FCRL5; HBG2; CD177; RPS28; OLFM4; MYOM2; ETV7;</i>	Blackfan-Diamond anemia; Classical homocystinuria; Diamond-Blackfan anemia; Hereditary persistence of fetal hemoglobin - beta-thalassemia; acne; classic homocystinuria; cyanosis, transient neonatal; familial thoracic aortic aneurysm and aortic dissection; homocystinuria; hyperhomocysteinemia,	14

	<i>MMP8; ZNF683; CBS; CRISP3; LTF</i>	thrombotic, cbs-related; infection; periodontitis; rosacea; tuberculosis	
720	<i>TMED7-TICAM2; NPIP15; HBG2; RHD; HBZ; GOLGA8Q; TUBB2A; IFI27; ALOX15; TMOD4; MYL9; S100B; RPL9; GP1BB; RSAD2; PPPBP; IFI44L; HLA-DQA2; C19orf33; PROS1</i>	Bernard-Soulier syndrome; Blackfan-Diamond anemia; Diamond-Blackfan anemia; Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Macrothrombocytopenia; Thrombocytopenia; Thromboembolism; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyanosis, transient neonatal; deep vein thrombosis; dengue disease; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hereditary thrombophilia due to congenital protein S deficiency; liposarcoma; lymphoma; megacystis-microcolon-intestinal hypoperistalsis syndrome 4; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; protein S deficiency; pulmonary embolism; stomach neoplasm; thrombophilia due to protein S deficiency, autosomal dominant; thrombophilia due to protein S deficiency, autosomal recessive; venous thromboembolism	49
721	<i>ETV7; RSAD2; PNRC2; KIR3DL2; KIR2DS4; IFI44L; IFIT1; RGP2; ISG15; HLA-DQA2; CRYZ; RGP1; IFI27; IFI6; JCHAIN; MX1; POLR2J2-UPK3BL1; MCOLN2; CEACAM6; APOBEC3B</i>		0
722	<i>GP1BB; DEFA1B; RIOX1; H2AC7; EIF3CL; IFI27; DEFA3; SMIM1; MMP8; OLFM4; SERPINB10; LTF; RHD; ETV7; RSAD2; ELANE; DEFA1; CEACAM6; GOLGA80; LCN2</i>	Alpha-1-antitrypsin deficiency; Bernard-Soulier syndrome; Cyclic neutropenia; Decreased total neutrophil count; Macrothrombocytopenia; Thrombocytopenia; acne; alpha 1-antitrypsin deficiency; cyclic hematopoiesis; emphysema; infection; infectious disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; periodontitis; rosacea; severe congenital neutropenia; tuberculosis; viral disease	19

723	<p><i>KIR2DS4; SETD9; TMEM176B; TUBB2A; TMEM176A; ESPN; RNASE4; ORM2; GOS2; ETV7; OLFM4; NME2; COCH; RSAD2; ZDHHC19; C1QB; SMIM1; CHST13; GPX3; IFI44L</i></p>	<p>Abnormality of the skeletal system; C1Q deficiency; C1Q deficiency 1; Genetic central nervous system malformation; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; amyotrophic lateral sclerosis; autosomal dominant nonsyndromic hearing loss; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; deafness; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hearing loss, autosomal recessive; hearing loss, autosomal recessive 110; immunodeficiency due to a classical component pathway complement deficiency; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; nonsyndromic genetic hearing loss; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm</p>	42
724	<p><i>POLR2J2-UPK3BL1; CD177; RAP1GAP; IFI27; TUBB2A; ORM2; HBZ; RSAD2; NRG1; MACROD2; LYSMD3; PRSS33; VMO1; ADAMTS1; APOBEC3B; TNNT1; HBG2; IDO1; CLEC12B; NUTM2B</i></p>	<p>Abnormality of the skeletal system; Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Transitional Cell Carcinoma; adenocarcinoma; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyanosis, transient neonatal; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; hypothyroidism; liposarcoma; lung adenocarcinoma; lymphoma; multiple myeloma; nemaline myopathy 5; nemaline myopathy 5B, autosomal recessive, childhood-onset; nemaline myopathy 5C, autosomal dominant; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; papillary thyroid carcinoma; peritoneal neoplasm; prostate cancer; prostate neoplasm; stomach neoplasm; thyroid cancer; thyroid carcinoma</p>	43
725	<p><i>POLR2J2-UPK3BL1; NPIP15; DEFA3; BEGAIN; COL9A3; PNRC2; KIR2DS4; RAB4B-EGLN2; PRSS21; DAAM2; PVALB; NPIPA5; CXCL8; DEFA4; LTF; LCN2; RETN; RAP1GAP; IFI27; KLRC1</i></p>	<p>Stickler syndrome; Stickler syndrome, type 6; anemia; anemia (phenotype); chronic kidney disease; epiphyseal dysplasia, multiple, 3; nephrotic syndrome, type 24; tuberculosis</p>	8

726	<i>POLR2J2-UPK3BL1; CHURC1-FNTB; BEGAIN; C4A; TUBB2A; RAP1GAP; NOG; CD248; NPIP815; CES1; SMIM1; HLA-DQB1; LRRN3; S100B; IFI27; OLFM4; LYSMD3; HLA-G; ALOX15; TMEM176A</i>	Genetic central nervous system malformation; Hodgkins lymphoma; Hutchinson-Gilford progeria syndrome; Transitional Cell Carcinoma; adenocarcinoma; brachydactyly type B2; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complement component 4a deficiency; complex cortical dysplasia with other brain malformations 5; diffuse large B-cell lymphoma; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; immunodeficiency due to a classical component pathway complement deficiency; liposarcoma; lymphoma; multiple myeloma; multiple synostoses syndrome 1; neoplasm; neurodegenerative disease; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; proximal symphalangism 1A; stapes ankylosis with broad thumbs and toes; stomach neoplasm; tarsal-carpal coalition syndrome	40
727	<i>GP1BB; RIOX1; MDGA1; SMN2; ATOH8; NECTIN2; JCHAIN; GOLGA80; RGPD1; LYSMD3; CD177; HBZ; LGALS2; CRIP2; HBG1; IFI27; TNFRSF17; HLA-DQA2; RSAD2; SMIM24</i>	Bernard-Soulier syndrome; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Macrothrombocytopenia; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; Thrombocytopenia; multiple myeloma; neoplasm; neurodegenerative disease; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III	13
728	<i>GP1BB; RIOX1; CD177; MYOM2; HBZ; PNRC2; DEFA1; C1QB; TMTC1; IFI27; CTSG; PRTN3; C1QA; DEFA3; COL18A1; DEFA4; ELANE; CEACAM6; CEACAM8; KIR2DS4</i>	Alpha-1-antitrypsin deficiency; Bernard-Soulier syndrome; C1Q deficiency; C1Q deficiency 1; Cyclic neutropenia; Decreased total neutrophil count; Dupuytren Contracture; Knobloch syndrome; Knobloch syndrome 1; Macrothrombocytopenia; Thrombocytopenia; alpha 1-antitrypsin deficiency; cyclic hematopoiesis; emphysema; hereditary glaucoma, primary closed-angle; immunodeficiency due to a classical component pathway complement deficiency; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; schizophrenia; severe congenital neutropenia	20
729	<i>DEFA1B; POLR2J2-UPK3BL1; H2AC7; MDGA1; BEGAIN; ARG1; PNRC2; HBZ; PRTN3; H2AC25; SMIM1; DEFA4; OLFM4; SERPINB10; CD70; DEFA1; JCHAIN; KIR2DS4; ELANE; EXOC3L1</i>	Alpha-1-antitrypsin deficiency; Argininemia; Cyclic neutropenia; Decreased total neutrophil count; alpha 1-antitrypsin deficiency; arginase deficiency; cyclic hematopoiesis; emphysema; infectious disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; severe combined immunodeficiency due to CD70 deficiency; severe congenital neutropenia; viral disease	14
730	<i>RIOX1; PRSS21; S100B; POLR2J2-UPK3BL1; HBG1; MYOM2; LYSMD3; AP1M2; CD177; HBG2; VMO1; TMEM158; NECTIN2;</i>	HIV infection; Hereditary persistence of fetal hemoglobin - beta-thalassemia; cyanosis, transient neonatal; heterotaxy, visceral, 9, autosomal, with male infertility; nephrotic syndrome, type 24; schizophrenia	6

	<i>MNS1; SERPINB10; LYPD2; DAAM2; SMIM1; FCRL5; TMTC1</i>		
731	<i>POLR2J2-UPK3BL1; DEFA4; HBZ; CD177; HLA-DQA2; CTSG; DEFA1; S100P; DEFA3; NPIP15; RPS28; CFD; HBG2; RGPLD6; KIR2DS4; LTF; DAAM2; MMP8; HBG1; CEACAM6</i>	Blackfan-Diamond anemia; Diamond-Blackfan anemia; Hereditary persistence of fetal hemoglobin - beta-thalassemia; acne; cyanosis, transient neonatal; infection; nephrotic syndrome, type 24; periodontitis; recurrent Neisseria infections due to factor D deficiency; rosacea; tuberculosis	11
732	<i>DEFA1B; POLR2J2-UPK3BL1; ST6GALNAC1; UQCRHL; TMEM176A; ISG15; RSAD2; TMEM176B; USP18; IFIT1; CXCL10; SPATS2L; CMPK2; OAS3; SERPING1; SIGLEC1; IFI44L; MX1; HERC5; EXOC3L1</i>	Bartholin gland disease; C1 inhibitor deficiency; COVID-19; Congenital intrauterine infection-like syndrome; angioedema; atrial fibrillation; genetic disorder; hereditary angioedema; hereditary angioedema type 1; hereditary angioedema type 2; hereditary angioedema type 3; hereditary angioedema with C1Inh deficiency	12
733	<i>SMN2; NPIPA5; S100B; KLRC4; C4A; TFF3; TCL1A; FCRL5; FCRL1; GOLGA80; RAP1GAP; CD177; FCRL2; CD8B; JCHAIN; GOLGA8B; HBG1; PIGY; HLA-G; CD248</i>	HIV infection; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; clonal hematopoiesis; complement component 4a deficiency; hypercoagulability syndrome due to glycosylphosphatidylinositol deficiency; hyperphosphatasia-intellectual disability syndrome; immunodeficiency due to a classical component pathway complement deficiency; neurodegenerative disease; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III	14
734	<i>HLA-DQA2; H2AC7; HLA-DQB2; GP1BB; TMEM176B; HBZ; GOLGA8Q; TMEM176A; SMIM1; RPS28; NPIPA5; PRDM5; IFI27; POLR2J2-UPK3BL1; PVALB; RAB4B-EGLN2; CD177; CENPK; PF4V1; LGALS9C</i>	Abnormality of the cardiovascular system; Bernard-Soulier syndrome; Blackfan-Diamond anemia; Connective tissue disease with eye involvement; Diamond-Blackfan anemia; Macrothrombocytopenia; Thrombocytopenia; anemia; anemia (phenotype); brittle cornea syndrome; chronic kidney disease; infectious disease; viral disease	13
735	<i>DEFA1B; POLR2J2-UPK3BL1; HLA-DQB2; SMN2; LYSMD3; RAP1GAP; OLFM4; PRTN3; DEFA3; CEACAM6; DEFA1; DEFA4; CRISP3; MDGA1; RSAD2; ETV7; LTF; DAAM2; MMP8; CEACAM8</i>	Proximal spinal muscular atrophy type 2; Proximal spinal muscular atrophy type 3; acne; infection; nephrotic syndrome, type 24; neurodegenerative disease; periodontitis; rosacea; spinal muscular atrophy; spinal muscular atrophy, type 1; spinal muscular atrophy, type II; spinal muscular atrophy, type III; tuberculosis	13

736	<i>H2AC7; MDGA1; S100P; BEGAIN; PNRC2; MYOM2; LYSMD3; NPIPA5; RHD; POLR2J2-UPK3BL1; APOBEC3B; HBZ; CD177; ZDHHC11B; RAP1GAP; COCH; LTF; ORM1; NECTIN2; DEFA1B</i>	autosomal dominant nonsyndromic hearing loss; deafness; hearing loss, autosomal recessive 110; infectious disease; nonsyndromic genetic hearing loss; tuberculosis; viral disease	7
737	<i>DEFA3; KRT72; BEGAIN; CD177; H2AC7; KRT73; NPIP15; KIR2DS4; TMTC1; HBZ; HBG2; NSG1; HBG1; LYSMD3; RSAD2; TGIF2-RAB5IF; POLR2J2-UPK3BL1; LAIR2; APOBEC3B; CFD</i>	Hereditary persistence of fetal hemoglobin - beta-thalassemia; cyanosis, transient neonatal; infectious disease; neurodegenerative disease; recurrent Neisseria infections due to factor D deficiency; schizophrenia; viral disease	7
738	<i>POLR2J2-UPK3BL1; CD177; IFI27; NME1-NME2; RSAD2; LY6E; IFI44L; SIGLEC1; OAS3; CMPK2; RPL9; IFIT1; KRT72; DEFA3; ISG15; EPST1; JUP; OAS1; AGRN; IFI44</i>	Arrhythmogenic right ventricular dysplasia; Blackfan-Diamond anemia; COVID-19; Congenital myasthenic syndromes; Diamond-Blackfan anemia; Naxos disease; Postsynaptic congenital myasthenic syndromes; Presynaptic congenital myasthenic syndromes; arrhythmogenic right ventricular dysplasia 12; congenital myasthenic syndrome 8; pulmonary alveolar proteinosis with hypogammaglobulinemia	11
739	<i>POLR2J2-UPK3BL1; DEFA3; HBZ; CTSG; RHD; ARG1; HBG1; HBG2; IFI27; DEFA1; TUBB2A; RAP1GAP; DEFA4; GYPB; PRTN3; RSAD2; CEACAM6; KRT1; ELANE; CA1</i>	Alpha-1-antitrypsin deficiency; Argininemia; Cyclic neutropenia; Decreased total neutrophil count; Genetic central nervous system malformation; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hodgkins lymphoma; Non-epidermolytic palmoplantar keratoderma; Seizure; Transitional Cell Carcinoma; adenocarcinoma; alpha 1-antitrypsin deficiency; altitude sickness; angle-closure glaucoma; annular epidermolytic ichthyosis; arginase deficiency; breast cancer; breast carcinoma; breast neoplasm; cancer; cervical cancer; complex cortical dysplasia with other brain malformations 5; cyanosis, transient neonatal; cyclic hematopoiesis; diffuse large B-cell lymphoma; diffuse nonepidermolytic palmoplantar keratoderma; edema; emphysema; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; familial Mediterranean fever; gastric cancer; gastroesophageal junction adenocarcinoma; glaucoma; gout; head and neck malignant neoplasia; head and neck neoplasia; head and neck squamous cell carcinoma; ichthyosis hystrix of Curth-Macklin; liposarcoma; lymphoma; multiple myeloma; neoplasm; neurodegenerative disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; non-Hodgkins lymphoma; non-small cell lung carcinoma; ovarian cancer; ovarian neoplasm; pancreatic neoplasm; peritoneal neoplasm; prostate cancer; prostate neoplasm; severe congenital neutropenia; stomach neoplasm	56
740	<i>HLA-DQA2; POLR2J2-UPK3BL1; HLA-DQB2; IFI27;</i>	Argininemia; Congenital intrauterine infection-like syndrome; Non-epidermolytic palmoplantar keratoderma; annular	11

	<i>S100B; DAAM2; CNTNAP3; MDGA1; CD177; LYSMD3; ARG1; ORM1; VMO1; IFI44L; GGT5; SIGLEC1; USP18; KRT1; LAG3; ENTPD2</i>	epidermolytic ichthyosis; arginase deficiency; diffuse nonepidermolytic palmoplantar keratoderma; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; ichthyosis hystrix of Curth-Macklin; melanoma; nephrotic syndrome, type 24	
741	<i>PRSS33; SIGLEC8; CD177; ALOX15; KRT72; CLC; IL5RA; MSR1; OLIG2; KRT73; ARG1; RNASE3; RPH3A; TFF3; S100B; XCL2; COL9A3; LTF; CACNG6; SIGLEC6</i>	Argininemia; Seizure; Stickler syndrome; Stickler syndrome, type 6; anxiety disorder; arginase deficiency; asthma; epilepsy; epiphyseal dysplasia, multiple, 3; fibromyalgia; neuralgia; neuropathic pain; postherpetic neuralgia; restless legs syndrome; spinal cord injury; tuberculosis	16
742	<i>TGIF2-RAB5IF; MYZAP; LYSMD3; HLA-DQA2; MYOM2; HBZ; PRTN3; HBG2; RPS28; ELANE; AZU1; ELAPOR1; IFI44L; KRT1; CTSG; RHD; MSR1; RSAD2; S100P; ABCA1</i>	Alpha-1-antitrypsin deficiency; Blackfan-Diamond anemia; Cyclic neutropenia; Decreased total neutrophil count; Diamond-Blackfan anemia; Hereditary persistence of fetal hemoglobin - beta-thalassemia; Hypercholesterolemia; Non-epidermolytic palmoplantar keratoderma; Tangier disease; alpha 1-antitrypsin deficiency; annular epidermolytic ichthyosis; atrial fibrillation; cardiomyopathy, dilated, 2K; coronary artery disease; cyanosis, transient neonatal; cyclic hematopoiesis; diffuse nonepidermolytic palmoplantar keratoderma; emphysema; epidermolytic ichthyosis; epidermolytic palmoplantar keratoderma, 1; glaucoma; hypoalphalipoproteinemia, primary, 1; ichthyosis hystrix of Curth-Macklin; metabolic disease; metabolic syndrome; neurodegenerative disease; neutropenia; neutropenia, severe congenital, 1, autosomal dominant; open-angle glaucoma; severe congenital neutropenia	30
743	<i>POLR2J2-UPK3BL1; GOLGA8Q; NPIP15; MS4A2; DEFA3; OLFM4; RGP1; SIGLEC6; SMIM1; HBZ; SPP1; HDC; CLEC4C; CLEC18A; BCAM; PCSK1N; KLC3; RAP1GAP; ZNF573; HBD</i>	bone disease; delta-beta-thalassemia; osteoporosis	3

**Table 1.** Patient profiles, and gene-disease associations.

Gene	Diseases From Annotation Databases	Diseases From Literature Review
<b>ACKR1</b>	<p>Basophil percentage of white blood cells; C-C motif chemokine levels (14,17,11); C-X-C motif chemokine 6 levels in chronic kidney disease with hypertension and no diabetes; C-X-C motif chemokine levels (6,11,2,3,5); Chronic periodontitis; DNA methylation - estimated granulocyte proportions; DNA methylation phenoage acceleration; Duffy blood group system; Duffy blood group system FY (a-b-) phenotype; Duffy blood group system FY (b-wk) phenotype; Eotaxin level in chronic kidney disease with hypertension and no diabetes; Eotaxin levels (CCL 11.5301.7.3); Gro-beta level in chronic kidney disease with hypertension and no diabetes (3148 49); Gro-gamma level in chronic kidney disease with hypertension and no diabetes (22973 8); Gro-gamma level in chronic kidney disease with hypertension and no diabetes; Growth-regulated protein alpha levels; IgE levels; Inflammatory biomarkers; Interleukin-8 levels; Monocyte percentage of white blood cells; Resistance to plasmodium vivax infection; Serum levels of protein CDSN (corneodesmosin); Sickle cell anemia; Susceptibility to malaria; Ulcerative colitis; White blood cell count quantitative trait locus 1</p>	<p>Alcohol-induced osteonecrosis of the femoral head (ONFH); Alpha thalassemia; Ascending thoracic aortic aneurysm; Asthma; Atherosclerosis; Behcet's disease (BD); Benign constitutional neutropenia (BCN); Benign ethnic neutropenia (BEN); Breast cancer; Cardiovascular disease (all-cause mortality; oronary heart disease, heart failure); Clozapine-related neutropenia in treatment-resistant schizophrenia; Coronary artery disease due to atherosclerosis; Endometriosis (EMS); Gallbladder cancer; Hashimoto's thyroiditis (HT); Hepatitis C virus; HIV-1; Lichen sclerosus urethral stricture disease (LS USD); Liver cirrhosis; Multiple sclerosis (MS); Non-infectious Posterior uveitis; Non-small cell lung cancer; Obesity-related disease; Preeclampsia (PE); Progressive lung fibrosis; Prostate cancer; Ptegyrium; Renal disease; Rheumatoid arthritis (RA); Sarcoma; Sepsis-associated encephalopathy (SAE); Sjogren's syndrome (SS); Skin fibrosis; Systemic lupus erythematosus (SLE); Thyroid associated ophthalmopathy (TAO); Triple-negative breast cancer; Type A aortic dissection (TAAD); Urothelial carcinoma (UC); Vascular dysfunction after COVID-19 infection</p>
<b>C4B</b>	<p>Adenoma, cortisol-producing; Adrenocortical carcinoma, Androgen-secreting; Cdeficiency; Classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency; Complement and coagulation cascades; Complement component 2 deficiency; Congenital adrenal hyperplasia; CYP21A2 (cytochrome P450 family 21 subfamily A member 2) - related disease; Inborn genetic disease; Pertussis; Positive regulation of apoptotic cell clearance (GO:2000427); Positive regulation of phagocytosis (GO: 0050766); Regulation of apoptotic cell clearance (GO:200425); Sillicosis; Staphylococcus aureus infection; Systemic lupus erythematosus (SLE)</p>	<p>Acute central serous chorioretinopathy (aCSC); Acute coronary syndrome (ACS); Acute myocardial infarction; Adult attention-deficit/hyperactivity disorder (ADHD); Adult rhinosinusitis; Adult sepsis; Age-related macular degeneration; Allergic rhinitis (AR); Alzheimer's (AD); Amyotropic lateral sclerosis (ALS); Atypical hemolytic uremic syndrome (aHUS); Autism spectrum disorder (ASD); Autoimmune chronic hepatitis; Autoimmune related kidney diseases (ARKDs); C3 glomerulonephritis; Celiac disease; Cervical cancer development due to human papillomavirus; Chronic central serous chorioretinopathy (cCSC); Chronic kidney disease (CKD); Colorectal cancer; Coronary artery disease (CAD); Coronary atherosclerotic heart disease (CAD); COVID-19; Coxsackievirus Bs-infected myocarditis; Crosslink between Alzheimer's and periodontitis; Cytomegalovirus (CMV); Dementia with Lewy bodies (DLB); Diabetic kidney disease (DKD); Diabetic retinopathy; Early rheumatoid arthritis (eRA); Ehlers-Danlos syndrome; Felty's syndrome; Graves' disease (GD); Hashimoto's</p>

		<p>thyroiditis; Henoch-schonlein purpura (HSP); Hepatitis B virus (HBV); Hepatitis C virus (HCV); Hepatocellular carcinoma (HCC); Hereditary angioedema (HAE); Herpes gestationis (HG); Human immunodeficiency virus type 1 (HIV-1); Immunoglobulin A nephropathy (IgAN); Juvenile dermatomyositis (JDM); Juvenile rheumatoid arthritis; Kawasaki disease (KD); Lafora progressive myoclonus epilepsy (Lafora disease, LD); Legg-Calve-Perthes disease; Leprosy; Link between chlamydia pneumoniae infection and coronary artery disease (CAD); Liver cirrhosis; Luminal breast cancer (BCa); Lung squamous cell carcinoma (LUSC); Membranoproliferative glomerulonephritis (MPGN) type III; Metastatic renal cell carcinoma (RCC); Monkeypox virus (MPXV); Multiple sclerosis (MS); Myasthenia gravis (MG); Neuronal ceroid lipofuscinoses (NCLs or Batten disease); Non-small cell lung cancer (NSCLC); Osteoarthritis; Paediatric inflammatory bowel disease (PIBD); Paracoccidioidomycosis (PCM); Parkinson's disease; Parkinson's disease dementia (PDD); Pediatric acute-onset neuropsychiatric syndrome (PANS); Periodontitis in coronary artery disease (CAD) patients; Premature atherosclerotic peripheral vascular disease; Primary biliary cirrhosis (PBC); Primary ovarian failure; Psoriasis vulgaris; Pulmonary arterial hypertension (PAH); Pulmonary embolism (PE); Recurrent lymphocytic meningitis (RLM); Rheumatic heart disease (RHD); Rheumatoid arthritis (RA); Sarcoidosis; Schizophrenia; Sepsis-associated encephalopathy (SAE); Subclinical atherosclerosis in rheumatoid arthritis (RA) patients; Systemic sclerosis (SSc); Type 1 diabetes (T1DM); Type 2 diabetes mellitus; Vascular calcification (VC) in different VC conditions</p>
<p><b>IFI44L</b></p>	<p>Defense response to symbiont (GO:0140546); Defense response to virus (GO:0051607); Febrile seizures; Febrile seizures (MMR vaccine-related); Influenza; Multisystem inflammatory syndrome in children; Systemic lupus erythematosus (SLE)</p>	<p>Acute febrile illness; Acute lymphoblastic leukemia (ALL); Acute myeloid leukemia (AML); Aicardi-goutieres syndrome (AGS); Alzheimer's disease (AD); Antiphospholipid syndrome (APD); Behcet's disease (BD); Bipolar disorder; Bloom syndrome (BS); Childhood-onset systemic lupus erythematosus (cSLE); Chronic cough; Chronic hepatitis B (CHB); Chronic kidney disease (CKD); COVID-19; Cutaneous leishmaniasis; Dengue fever virus (DENV); Diabetic foot ulcer (DFU); Early rheumatoid arthritis (eRA); Febrile neutropenia; Fetal diarrhea caused by cryptosporidium parvum infection; Gastrointestinal (GI) cancer; Graves' disease (GD); Head and neck cancer; Head and</p>

		<p>neck squamous cell carcinoma; Hepatitis C virus (HCV); Hepatocellular carcinoma (HCC); Histiocytic necrotizing lymphadenitis (HNL); Human immunodeficiency virus (HIV); Interconnection of membranous nephropathy (MN) and lupus nephritis (LN); Ischemic cardiomyopathy; Japanese encephalitis virus (JEV); Juvenile dermatomyositis (JDM); Juvenile myositis; Large-vessel involvement in giant cell arteritis (LV-GCA); Lupus nephritis (LN); Mixed connective tissue disease (MCTD); Multiple sclerosis (MS); Noncystic fibrosis bronchiectasis; Obesity-related traits; Oral squamous cell carcinoma; Osteonecrosis of the femoral head (ONFH); Pancreatic ductal adenocarcinoma (PDAC); Periodontitis and rheumatoid arthritis connection; Peripheral vascular atherosclerosis (PVA); Photoaging; Post-acute sequelae of COVID-19; Preeclampsia (PE); Premature ovarian failure (POF); Proliferative diabetic retinopathy (PDR); Psoriasis; Respiratory syncytial virus (RSV); Respiratory tract infections (RTIs) in children; Rheumatoid arthritis (RA); Rhinovirus infection; Schizophrenia; Severe fever with thrombocytopenia syndrome (SFTS); Sjogren's syndrome (SS); Skin cancer (cutaneous melanoma and non melanoma skin cancer); SSA/Ro autoantibody-associated congenital heart block (CHB); Stevens-Johnson syndrome (SJS); Systemic sclerosis (SSc); Tuberculosis (TB); Ulcerative colitis (UC); Zika virus (ZIKV)</p>
<p><b>IFITM3</b></p>	<p>Cellular response to type I interferon; Defense response to symbiont; Defense response to virus; Influenza; Influenza, severe susceptibility to; Interferon-mediated signaling pathway (GO: 0140888); Negative regulation of viral entry into host cell (GO:0046597); Negative regulation of viral genome replication (GO:0045071); Negative regulation of viral life cycle (GO:1902901); Negative regulation of viral process (GO:004825); Negative regulation of viral transcription (GO:0032897); Regulation of viral entry into host cell (GO:0046596); Regulation of viral transcription (GO:0046782); Regulation of viral genome replication (GO: 0045069); Response to interferon-alpha (GO:0035455); Type I interferon-mediated signaling pathway (GO:0060337)</p>	<p>Acute myeloid leukemia (AML); African swine fever virus (ASFV); Alzheimer's disease (AD); Autism; Bronchial asthma; Chron's disease; Chronic obstructive pulmonary disease (COPD); Coronaravirus; Cutaneous melanoma; Dengue virus (DENV); Foamy viruses (FVs); Foot and mouth disease; Hepatitis C virus (HCV); Hepatocellular carcinoma; Hemorrhagic fever with renal syndrome (HFRS) caused by hantaan virus (HTNV); HIV-associated neurocognitive disorder (HAND); Human genital herpes; Huntington's disease (HD); Infected joint replacements; Influenza virus A; Ischemic stroke; JAK2V617F+ myelofibrosis; Kawasaki disease (KD); Lupus nephritis (LN); Multiple sclerosis (MS); Rheumatoid arthritis (RA); Schizophrenia; Systemic lupus erythematosus; Tick-borne encephalitis virus (TBEV); Tongue squamous cell carcinoma; Tuberculosis (TB); Ulcerative colitis; West Nile virus (WNV); Zika virus (ZIKV)</p>

<b>LGALS2</b>	Myocardial infarction (MI); Myocardial infarction 1 (MI); Susceptibility to myocardial infarction	Acute-on-chronic liver failure (ACLF); Antiphospholipid syndrome; Atherosclerosis; Atherothrombotic cerebral infarction among patients with metabolic syndrome; Behcet's disease (BD); Breast cancer; Colorectal cancer; Coronary artery disease (CAD); Coronary atherosclerosis; Coronary heart disease (CHD); Diabetes mellitus; Diffuse large B-cell lymphoma (DLBC); Gastric cancer; Ischaemic stroke; Psoriasis; Rheumatoid arthritis (RA); Severe malaria (SM) syndromes; Triple negative breast cancer; Type 2 diabetes (T2D); Ulcerative colitis; Urothelial carcinoma
<b>RETN</b>	Aneurysm; Asthma; Azurophil granule; Azurophil granule lumen; Chronic kidney insufficiency; Diabetes mellitus type 2; Diabetes mellitus type 2 susceptibility; Endometriosis; Hypertension; Hypertension, insulin resistance-related susceptibility to; Inflammation; Renal cell carcinoma; Secretory granule lumen; Sepsis; Septicemia; Severe sepsis; Specific granule; Specific granule lumen; Vacuolar lumen (GO:0005775)	Abdominal aortic aneurysm (AAA); Acne vulgaris; Acute aortic dissection; Acute cerebral infarction; Acute kidney injury due to Puumala hantavirus (PUUV); Acute myeloid leukemia (AML); Acute-on-chronic liver failure (ACLF); Alcohol-induced osteonecrosis of femoral head; Alzheimer's disease (AD); Anorexia nervosa (AN); Aortic valve stenosis; Aplastic anemia (AA); Atherosclerosis (AS); Atopic dermatitis (AD); Bacterial pneumonia; Behcet's disease (BD); Beta-thalassemia in children; Breast cancer; Chagas disease; Chronic hepatitis B; Chronic kidney disease in children; Chronic myeloid leukemia; Chronic obstructive pulmonary disease (COPD); Chronic Renal failure; Colon cancer; Colorectal cancer; Coronary artery disease (CAD); COVID-19; Crohn's disease; Cystic echinococcosis (CE); Cystic fibrosis in children; Dementia; Depression due to alcohol dependence; Diabetic nephropathy (DN); Gallbladder cancer; Graves' disease (GD); H1N1 influenza; Hand osteoarthritis (HOA); Heart failure (HF); Hepatitis C virus; Hidradenitis suppurativa (HS); Hyperthyroidism; Hypertrophic cardiomyopathy; Hypothyroidism; Idiopathic dilated cardiomyopathy; Inflammatory bowel disease (IBD); Intervertebral disc degeneration (IDD); Ischemic stroke; Kawasaki disease (KD); Kidney failure (KF); Knee joint osteoarthritis; Liver cirrhosis; Long-COVID (LC) caused by Epstein-Barr virus (EBV); Lung cancer; Major depressive disorder (MDD); Metabolic syndrome (Mets) in antisynthetase syndrome (ASS); Metabolic syndrome (MS); Multiple organ dysfunction syndrome (MODS) caused by pediatric influenza; Neonatal sepsis (NS); Non-ST segment elevation myocardial infarction; Nonalcoholic fatty liver disease; Obesity; Obstructive sleep apnoea; Oral

		<p>cancer; Oral squamous cell carcinoma; Osteoarthritis; Osteoporosis; Pancreatic ductal adenocarcinoma (PDAC); Periodontitis; Peripheral artery disease (PAD); Perivascular adipose tissue (PVAT) dysfunction in rheumatoid arthritis (RA); Polycystic ovary syndrome (PCOS); Polymyalgia rheumatica; Prostate cancer; Psoriasis; Psoriatic arthritis (PsA); Pulmonary tuberculosis (PTB); Pulmonary arterial hypertension; Receptor-associated periodic syndrome (TRAPS); Rheumatoid arthritis (RA); Severe acute pancreatitis; Sjogren's syndrome (SS); Spontaneous bacterial peritonitis in cirrhosis; Steroid-induced osteonecrosis of the femoral head; Systemic lupus erythematosus; Systemic sclerosis (SSc); Testicular cancer; Type 1 diabetes; Urogenital Schistosomiasis; Vitiligo</p>
<b>RHD</b>	<p>Ammonium homeostasis (GO:0097272); Ammonium transmembrane transport (GO:0072488); Ammonium transmembrane transporter activity; Anemia; Anti-D isoimmunization affecting pregnancy; Blood; Cholesterol levels in small LDL; Chronic myeloid leukemia; Concentration of small LDL particles; Diabetes mellitus; Hemolytic congenital anemia; Hemolytic disease of fetus or newborn due to RhD isoimmunization; Mean corpuscular hemoglobin concentration; Mean corpuscular volume; Mean reticulocyte volume; Mean spheric corpuscular volume; Neoplasm metastasis; Omega-6 fatty acid levels; Sick cell anemia; Total lipid levels in small LDL</p>	<p>Cardiac resynchronisation therapy (CRT) prognosis in heart failure (HF) patients; Coronary artery disease (CAD); Dyslipidemia; Gastric cancer; Influenza A virus; Neonatal haemolytic disease caused by clinical RhD blood group incompatibility; Non-small cell lung cancer (NSCLC); SARS-CoV-2; Schizophrenia; Stomatocytosis and chronic hemolytic anemia caused by Rhnull disease</p>
<b>TNFRSF13B</b>	<p>Abnormal pulmonary interstitial morphology; Absent epiphyses; Chronic lung disease; Cleft palate; Clubfoot; Coat hanger sign of ribs; Common variable immunodeficiency; Hematology traits; Hemivertebrae; Immune deficiency familial variable; Immunodeficiency; Immunodeficiency common variable 2; Immunodeficiency, common variable 1; Immunoglobulin A deficiency 2; Meningitis; Micrognathia; Patent ductus arteriosus; Preaxial foot polydactyly; Pseudoarthrosis; Respiratory failure; Restrictive lung disease (HP:0002091); Severe SARS-CoV-2 infection, susceptibility to; Short femur; Skeletal dysplasia; TNFRSF13B-related disorder; Vasculitis (HP:0002633); Vertebral hypoplasia; Vertebral segmentation defect</p>	<p>Acute malaria; Acute myeloid leukemia (AML); Anti-N-methyl-D-aspartate receptor encephalitis (NMDAR-E); Asthma; Atherosclerosis; Autoimmune neutropenia; Behcet's disease; Breast cancer; Chronic lymphocytic leukemia (CLL); Chronic nonsuppurative otitis media (OM); Chronic rhinosinusitis with nasal polyps (CRSwNP); Coronary artery lesions in Kawasaki patients; Dengue virus; Diffuse large B-cell lymphoma (DLBCL); Diffuse pulmonary lymphangiomas accompanied by thrombocytopenia; Ehlers-Danlos syndrome (EDS); Epstein-Barr virus (EBV) associated lymphoproliferative diseases; Febrile infection-related epilepsy syndrome (FIRES); Hairy cell leukemia variant (HCL-v); Hidradenitis suppurativa (HS); How CVD associated biomarkers in Rheumatoid arthritis are impacted by</p>

		periodontal disease; Idiopathic pulmonary fibrosis (IPF); IgA nephropathy (IgAN); Immune thrombocytopenia (ITP); Inborn errors of immunity-related lymphoma (IEI); Intracranial aneurysms (IAs); Leishmania infection; Melanoma; Meniere's disease (MD); Metabolic syndrome (MetS); Multiple myeloma (MM); Multisystem inflammatory syndrome in children (MIS-C); Non-Hodgkin lymphoma; Osteoarthritis; Pediatric lymphoma; Predominantly antibody deficiencies (PAD); Primary central nervous system lymphoma (PCNSL); Primary progressive multiple sclerosis (PPMS); Prostate cancer; Pulmonary arterial hypertension (PAH); Rectal cancer; Renal cell carcinoma (RCC); Rheumatoid arthritis (RA); Sarcoidosis; SARS-CoV-2; Sepsis; Sjogren's syndrome (SS); Systemic lupus erythematosus (SLE)
--	--	---

**Table 2.** Gene-disease associations derived from annotations databases and discovered through scientific literature review of genes, mutually identified by multi-omics data analyses.