DOI: https://doi.org/10.53040/cga12.10 UDC: 616-056.7:575.113

GENETIC MAPPING OF HUMAN DISEASES: A QUANTITATIVE AND FUNCTIONAL PERSPECTIVE ON GENE–PATHOLOGY RELATIONSHIPS

Maria DUCA[1], ORCID 0000-0002-5855-5194
mduca2000@yahoo.com
Alexandr ANTONOV[1], ORCID 0000-0001-8610-2654
antonov.alexandr@usm.md
Manuela SIDOROFF[2], ORCID: 0000-0001-8740-3803

[1]Moldova State University, Chisinau, Republic of Moldova
[2]Romanian Society of Bioengineering and Biotechnology, Bucharest, Romania

Abstract. *The classification of human diseases based on genetic contribution has significantly evolved with advancements in genetic research. This study explores the chromosomal distribution of disease-associated genes from a quantitative and functional perspective, aiming to identify genomic regions with a high density of pathogenic associations—so-called genetic hotspots. We integrate data from genome-wide association studies (GWAS) to evaluate genetic overlap among major disease categories, elucidating molecular connections between seemingly unrelated pathologies. The findings highlight the significance of genetic architecture in disease manifestation, offering valuable insights for precision medicine, genetic diagnostics, and targeted therapies.*

Keywords: genetic mapping, genetic hotspots, gene-pathology relationships, quantitative genetics, genetic overlap.

INTRODUCTION

The advancement of genetic research has enabled the classification of human diseases based on the extent and manner in which genetic factors contribute to their onset and progression. Two major perspectives have thus emerged: *qualitative genetics*, which focuses on monogenic diseases, and *quantitative genetics*, which addresses multifactorial or polygenic diseases. Both branches offer critical insights for diagnosis, treatment, and prevention.

Qualitative genetics investigates diseases caused by mutations in a single gene or by major chromosomal abnormalities. These are conditions characterized by clear and detectable genetic alterations. Such diseases often follow well-defined Mendelian inheritance patterns—autosomal dominant, autosomal recessive, or X-linked. Classic examples include cystic fibrosis, Duchenne muscular dystrophy, and sickle cell anemia, caused by mutations in the *CFTR* (Cystic Fibrosis Transmembrane Conductance Regulator) gene. This mutation disrupts chloride ion

transport in epithelial tissues, leading to the accumulation of thick mucus in the lungs and digestive tract (Riordan et al., 1989). As an autosomal recessive condition, both parents must be carriers of the mutation for the disease to manifest in their offspring. Similarly, Huntington's disease follows an autosomal dominant inheritance pattern, being caused by a CAG trinucleotide repeat expansion in the *HTT* gene (MacDonald et al., 1993). In such cases, the genotype-phenotype correlation is typically clear and predictable: a specific mutation is strongly associated with disease expression. Monogenic diseases are often severe and have early onset, but their predictable inheritance allows for genetic counseling and preventive testing.

In contrast, most common adult-onset disorders are influenced by numerous genes, each exerting a modest effect. In combination with environmental factors, these genetic contributions shape disease risk. These complex diseases include type 2 diabetes, cardiovascular diseases, obesity, and neuropsychiatric disorders. A typical example is type 2 diabetes, for which over 400 genetic variants have been associated (Mahajan et al., 2018). Genome-wide association studies (GWAS) have identified numerous genetic variants implicated in such diseases, each contributing only marginally to overall risk (Visscher et al., 2017). For these conditions, genetic research employs advanced statistical models, including polygenic risk scores (PRS), to evaluate individual susceptibility (Chatterjee et al., 2016).

Some conditions lie at the intersection of qualitative and quantitative genetics. For instance, hereditary breast cancer associated with BRCA1 and BRCA2 mutations has a clearly defined genetic basis (qualitative), yet the actual cancer risk is modulated by other genetic and environmental factors (quantitative). This overlap illustrates that while the distinction between qualitative and quantitative genetics is pedagogically useful, biological reality is far more nuanced (Mavaddat et al., 2019)

The distinction between these two branches of genetics has direct implications for precision medicine. In monogenic disorders, genetic testing can confirm the diagnosis and guide treatment—as seen in the case of gene therapies for spinal muscular atrophy (Mendell et al., 2017). Conversely, for multifactorial diseases, genetics mainly aids in risk stratification and prevention.

In the era of genomic medicine, understanding the genetic basis of disease is critical for the development of personalized treatments and preventive strategies. Identifying genes associated with various categories of disease offers important insights into the molecular mechanisms underlying these conditions.

This study aims to analyze, from a quantitative and functional perspective, the chromosomal distribution of disease-associated genes, and to identify genomic regions with a high density of pathogenic associations—so-called "genetic hotspots." Furthermore, we assess the degree of genetic overlap among major disease categories, with the goal of identifying molecular connections between

seemingly unrelated pathologies.

MATERIALS AND METHODS

The current investigations utilized publicly available genomic databases and curated disease-gene association datasets to analyze the chromosomal distribution of disease-associated genes. Statistical and bioinformatic tools were employed to identify genomic hotspots with high densities of pathogenic associations and to assess genetic overlap between major disease categories.

RESULTS AND DISCUSSION

By combining quantitative data on the number of implicated genes with a comparative analysis across different disease categories, we aimed to obtain an integrative view of the genetic complexity of human diseases. The results obtained can contribute to prioritizing research directions, optimizing genetic testing, and developing targeted therapies based on individual genomic profiles.

The bioinformatic analysis ***of gene distribution associated with various disease categories*** provided a comprehensive overview of the genetic landscape of human pathology. The data summarized in Table 1 reveal not only the extent of current knowledge in medical genetics but also the priority directions for research and intervention in precision medicine. Behind the numbers lie not only real biological differences but also echoes of the evolution of genetic technologies, funding priorities, and the intensity of scientific investigations.

Among all analyzed disease categories, nervous system and sensory organ diseases (NSD) stand out with the highest number of identified genes – 11,862. This impressive figure reflects both the structural and functional complexity of the central nervous system and the diversity of conditions included under this category, ranging from neurodegenerative diseases such as Alzheimer's or Parkinson's to neurological developmental disorders or sensory perception disorders. The pathogenic complexity of these conditions often translates into significant genetic heterogeneity, which justifies the scale of research in the field.

In second place, with 8366 associated genes, are mental and behavioral disorders (MaBD). This finding supports the idea of a consistent genetic contribution to the etiology of these diseases, often considered multifactorial and biologically difficult to define. The presence of an extensive genetic basis offers hope for the development of more effective methods for differential diagnosis and personalized treatment.

Digestive diseases (DSD) and musculoskeletal diseases (MSD) follow, with 7260 and 6688 identified genes respectively. The multiple involvements of these categories in various metabolic and inflammatory processes, as well as the variety

of organs and tissues affected, partially explain the associated genetic volume. At the same time, these figures reflect increased research interest, with direct implications for available treatments.

At the opposite end, infectious and parasitic diseases (IaPd) show the lowest number of associated genes – 3369. This lower value aligns with the nature of these diseases, where an external pathogenic agent plays a dominant role, and the host's genetic predispositions have a relatively secondary impact. However, the number is not negligible and suggests the existence of genetic susceptibility factors, especially in the context of immune response and systemic complications.

Beyond their descriptive value, these results have significant medical implications. Disease categories with a large number of identified genes emerge as promising targets for innovative therapies – including gene therapy, targeted molecular interventions, or predictive genetic testing. For example, treatment strategies in neurological and psychiatric disorders could be shaped based on individual genetic profiles. For diseases characterized by heterogeneous etiology – such as NSD or MaBD – the identification and cataloging of involved genes contribute to the definition of molecular subtypes and, implicitly, to the development of a more precise diagnostic and clinical classification framework. The gene distribution also reflects current scientific investments. Diseases such as Alzheimer's, cancer, or depression have received particular attention in genomic research projects, which has led to better genetic characterization and rapid data accumulation.

Table 1. Identified Genes in the GWAS Catalog Database for Various Disease Categories. *https://www.ebi.ac.uk/gwas/home (21.03.2025)*

Disease Category	**Abbreviation**	**Number of Identified Genes**
Diseases of the circulatory system	CVD	5,687
Diseases of the respiratory system	RSD	3,650
Endocrine, nutritional, and metabolic diseases	ESD	5,512
Diseases of the digestive system	DSD	7,260
Diseases of the nervous system and sensory organs	NSD	11,862
Diseases of the musculoskeletal system and connective tissue	MSD	6,688
Mental and behavioral disorders	MaBD	8,366
Infectious and parasitic diseases	IaPd	3,369
Neoplasms (tumors)	Neo	6,479

This quantitative perspective thus provides a solid starting point for the in-depth exploration of the molecular mechanisms involved in pathology and for

guiding research toward areas with the greatest potential for clinical impact. In addition to the quantitative analysis of genes involved in different disease categories, a further functional perspective can be obtained by evaluating ***their chromosomal distribution***. In this context, understanding chromosomal distribution and genetic overlaps has not only theoretical value but also direct implications in personalized medicine. Knowing the chromosomal location can aid in the development of targeted gene therapies, with minimal risk of side effects by avoiding damage to adjacent regions.

Based on the absolute number of genes per chromosome (Figure 1), the results highlight the intensity of genetic associations and reveal chromosomes 6 (5484 genes), 1 (4934 genes), 2 (4780 genes), 5 (3430 genes), and again 1 (3275 genes) as genomic regions with higher pathological loads — genetic "hotspots" for specific diseases or pathology in general.

Chromosome 6, ranking first, is already known for its region of highest genetic density within the immune system — the Major Histocompatibility Complex (MHC). This region plays a central role in antigen recognition and autoimmunity, justifying its constant presence in the pathogenesis of infectious, autoimmune, and even mental illnesses, where chronic inflammation is an emerging pathogenic mechanism. Conversely, chromosomes 13, 18, 21, and Y have the lowest number of genes associated with diseases. This can be explained by their smaller size and limited genetic content or low density of functional genes. For chromosome Y, its male-specific genetic content has lower relevance for common non-sex-linked diseases.

Although numerically less active, the X chromosome is critical in genetic pathology, especially in sex-linked diseases (e.g., hemophilia, Duchenne muscular dystrophy), due to its unique transmission and random inactivation in females.

The chromosomal distribution of disease-associated genes (Table 2) highlighted a disproportionate concentration on certain chromosomes — particularly those of larger size or containing gene-rich regions involved in nervous system development, metabolism, or immune regulation. For example, chromosomes 1, 6, 11, and X host a significant number of genes associated with neurological diseases, psychiatric disorders, and autoimmune conditions. These findings support the idea of genetic vulnerability “hotspots”.

For each group of conditions, certain chromosomes stand out as "genetic hotspots". For example, in CVD, chromosomes 1 (513 genes), 6 (454 genes), and 2 (445 genes) are predominant. These chromosomes encompass key genes involved in blood pressure regulation, lipid metabolism, and inflammatory response. Respiratory diseases are primarily associated with chromosome 6 (418 genes), reinforcing the hypothesis of an immunogenetic basis in these conditions, followed by chromosomes 1, 2, and 5. Endocrine and metabolic diseases (ESD) are linked to genes located on chromosome 6 (570 genes), followed by 1 (461 genes) and 2 (414

genes), reflecting the genetic importance in diabetes, thyroid disorders, and obesity. Neurological and sensory diseases are determined by genes with closely related values (~1000 genes each), located on chromosomes 1, 2, and 6, supporting the complexity of genetic networks involved in central nervous system function. Neoplasms are determined by oncogenes and tumor suppressor genes with a wide distribution across chromosomes 6 (611 genes), 2 (549 genes) and 1 (510 genes).

Table 2. Chromosomal Breakdown for Each Disease Group

Chromo-some ID	CVD	RSD	ESD	DSD	NSD	MSD	MaBD	IaPd	Neo	Total of affected genes
1	513	305	461	598	1023	561	680	283	510	4934
2	445	294	414	562	1018	524	743	231	549	4780
3	341	213	317	422	734	398	542	224	426	3617
4	310	178	310	357	628	379	435	199	302	3098
5	328	239	292	421	712	360	506	182	390	3430
6	454	418	570	659	1011	706	737	318	611	5484
7	284	182	303	397	615	335	437	179	293	3025
8	230	162	226	313	531	297	383	156	336	2634
9	189	170	212	279	444	292	304	134	311	2335
10	283	164	268	383	521	368	347	145	308	2787
11	364	186	317	423	648	380	468	162	327	3275
12	314	181	286	369	608	327	433	163	284	2965
13	133	83	144	201	336	179	238	96	167	1577
14	175	96	142	196	374	174	260	91	190	1698
15	181	146	161	217	376	184	269	99	199	1832
16	168	97	192	257	400	246	275	101	220	1956
17	270	156	225	286	456	240	309	116	238	2296
18	106	70	111	151	259	133	168	84	149	1231
19	207	105	162	254	404	171	296	114	195	1908
20	140	72	147	215	296	155	206	83	217	1531
21	64	36	52	78	150	84	104	54	76	698
22	97	55	100	138	205	119	136	57	125	1032
X	37	38	77	84	107	74	83	93	56	649
Y	0	4	0	0	6	1	6	4	0	21

Thus, understanding the cytogenetic topography of pathologies can be applied in targeted research, personalized medicine and in diagnosis and prevention. Chromosomal distribution can guide the development of sequencing panels focused on the "hot regions" of the genome. Integrating this knowledge into clinical practice will lead to faster diagnoses, targeted therapies, and effective prevention.

REFERENCES

1. Chatterjee, N., Shi, J., & García-Closas, M. (2016). *Developing and evaluating polygenic*

risk prediction models for stratified disease prevention. Nature Reviews Genetics, 17(7), 392–406.

2. MacDonald, M. E., Ambrose, C. M., Duyao, M. P., et al. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, 72(6), 971-983.
3. Mahajan, A., Taliun, D., Thurner, M., et al. (2018). *Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps*. Nature Genetics, 50(11), 1505–1513.
4. Mavaddat, N., Michailidou, K., Dennis, J., et al. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *The American Journal of Human Genetics*, 104(1), 21-34.
5. Mendell, J. R., Al-Zaidy, S. A., Shell, R., et al. (2017). *Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy*. New England Journal of Medicine, 377(18), 1713–1722.
6. Riordan, J. R., Rommens, J. M., Kerem, B., et al. (1989). *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA*. Science, 245(4922), 1066–1073.
7. The NHGRI-EBI Catalog of human genome-wide association studies. Available at: https://www.ebi.ac.uk/gwas/. Accessed: 21.03.2025.
8. Visscher, P. M., Wray, N. R., Zhang, Q., et al. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 101(1), 5-22