





Article

Genes and Diseases: Insights from Transcriptomics Studies

Dmitry S. Kolobkov ¹ , Darya A. Sviridova ¹ , Serikbai K. Abilev ¹ , Artem N. Kuzovlev ²
and Lyubov E. Salnikova ^{1,2,3,*} 

- ¹ The Laboratory of Ecological Genetics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119991, Russia; dmitry.s.kolobkov@gmail.com (D.S.K.); daria_sv11@mail.ru (D.A.S.); abilev@vigg.ru (S.K.A.)
- ² The Laboratory of Clinical Pathophysiology of Critical Conditions, Federal Research and Clinical Center of Intensive Care Medicine and Rehabilitology, Moscow 107031, Russia; artem_kuzovlev@fnkcr.ru
- ³ The Laboratory of Molecular Immunology, Rogachev National Research Center of Pediatric Hematology, Oncology and Immunology, Moscow 117997, Russia
- * Correspondence: salnikovalyubov@gmail.com

Abstract: Results of expression studies can be useful to clarify the genotype-phenotype relationship. However, according to data from recent literature, there is a large group of genes that are revealed as differentially expressed (DE) in many studies, regardless of the biological context. Additional analyses could shed more light on the relationships between genes, their differential expression, and diseases. We generated a set of 9972 disease genes from five gene-phenotype databases (OMIM, ORPHANET, DDG2P, DisGeNet and MalaCards) and a report of the International Union of Immunological Societies. To study transcriptomics of disease and non-disease genes in healthy tissues, we obtained data from the Human Protein Atlas (HPA) website. We analyzed the dependency between expression in healthy tissues and gene occurrence in Gene Expression Omnibus series using tools within the Enrichr libraries. The results of expression studies were annotated with Gene Ontology (GO) and Human Phenotype Ontology (HPO) terms. Using transcriptomics analysis of healthy tissues, we validated the previous findings of higher expression levels of disease genes in pathologically linked tissues compared to other tissues. Preferentially DE genes were generally highly expressed in one or multiple tissues and were enriched for disease genes. According to the results of GO enrichment analyses, both down- and up-regulated DE genes most often took part in immune response, translation and tissue-specific processes. A connection between DE-related pathology and the diversity of HPO terms was found. Investigating a link between expression and phenotype contributes to understanding the mode of development and progression of human diseases.

Keywords: differential expression (DE); disease genes; gene expression omnibus (GEO); gene ontology (GO); transcriptomics analysis of healthy tissues; human phenotype ontology (HPO)



Citation: Kolobkov, D.S.; Sviridova, D.A.; Abilev, S.K.; Kuzovlev, A.N.; Salnikova, L.E. Genes and Diseases: Insights from Transcriptomics Studies. *Genes* **2022**, *13*, 1168. <https://doi.org/10.3390/genes13071168>

Academic Editor: Stefano Lonardi

Received: 6 May 2022

Accepted: 23 June 2022

Published: 28 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Large-scale studies of gene expression are widely used to explore molecular mechanisms of human diseases. Most of these studies aim to identify genes with differential expression (DE) between diseased and healthy groups. It is generally accepted that an integrated analysis of DE genes can provide valuable information to identify disease markers and describe a targeted disease landscape in general [1–3]. However, recent studies demonstrated predictability of DE for a large group of genes, independent of the biological conditions in which they were investigated [4,5]. An analysis of overlapping genes within a framework of more than 600 expression studies showed that preferentially variable genes can be attributed to several sets of DE genes implicated in a common biological function, which were called DE modules. These modules included genes relating to sex, extracellular matrix, immune system and inflammation, cell cycle, and response to stress [4]. Since this finding is of a general nature, a number of questions arise. Could it be that main DE

modules do not differ for different tissues and directions of effect (down- or up-regulation)? Is there a link between gene expression in normal tissues and DE in pathology affected tissues? Are there differences in DE predictability between genes associated with a disease and other genes? Does the number of DE findings of a gene in transcriptome-wide studies correlate with the number of phenotypic terms for a gene within and between phenotypic categories?

Some of these issues have been partially addressed in the literature. Disease-associated genes tend to be expressed at higher levels in those healthy tissues where alterations of these genes cause pathology [6]. Gene expression profiles across healthy organs correlated with organ-specific cancer incidence, in particular, a positive correlation with cancer incidence was demonstrated for highly expressed genes involved in transcription, translation, and protein synthesis [7]. Both down and up-regulated DE genes may be enriched in the same biological processes [8,9], possibly due to compensatory mechanisms [10,11]. In multifactorial disorders and drug-induced perturbations, characteristics of DE profiles correlated with the broadness of phenotypic manifestations [12].

Due to swift development of molecular genetic technologies and rapid accumulation of gene expression data in public repositories, progressive refinement of expression studies becomes possible. In this work, we sought to use modern resources to address relationships between gene expression in healthy tissues and DE in pathology studies. With this objective, we generated a large set of disease genes from five gene-phenotype databases (OMIM, ORPHANET, DDG2P, DisGeNet and MalaCards) and the International Union of Immunological Societies (IUIS) report [13]. Due to the role of the immune system in the maintenance of tissue homeostasis and organism integrity [14], special attention was paid to the genes associated with immune diseases. In subsequent analyses, immune disease genes were compared with genes associated with respiratory and nervous system diseases. We analyzed transcriptomic profiles of disease and non-disease genes in healthy tissues and blood cells and assessed their expression levels and tissue specificity. We investigated the dependency between expression in healthy tissues and gene occurrence as DE in the Gene Expression Omnibus (GEO) series within the Enrichr libraries Disease Perturbations from GEO down studies and Disease Perturbations from GEO up studies. Structured Gene Ontology (GO) and Human Phenotype Ontology (HPO) vocabularies were used to incorporate the data of expression studies in the context with translational impact.

2. Materials and Methods

2.1. Creating a Series of Disease Genes

OMIM, ORPHANET, DDG2P, DisGeNet, and MalaCards were used to generate a list of disease genes. OMIM phenotypes related to a particular gene were obtained from the Gene Map retrieval set <https://www.omim.org/search/advanced/geneMap> (accessed on 1 August 2021). and annotated with Clinical Synopsis data (<https://www.omim.org/search/advanced/clinicalSynopsis>) with the MIMmatch service (accessed on 20 August 2021). ORPHANET gene-disease pairs were downloaded from <http://www.orphadata.org/cgi-bin/index.php> (accessed on 1 August 2021). Rare disease classifications were available from http://www.orphadata.org/cgi-bin/rare_free.html (accessed on 8 August 2021). DDG2P gene diseases and attributes were retrieved from <https://www.ebi.ac.uk/gene2phenotype/downloads> (accessed on 8 June 2021). DisGeNet curated gene-disease associations (GDA) with GDA scores ≥ 0.3 were obtained from <https://www.disgenet.org/downloads>. The list of GDAs was further restricted to the evidence index (EI) ≥ 0.9 MalaCards <https://www.malacards.org/> (accessed on 15 June 2021). elite genes with their corresponding diseases were kindly provided by MalaCards team (the MalaCards dump for GeneCards Suite Version v5.2). Elite genes are defined by gene-disease scores ≥ 25 , computed as a weighted sum of individual scores derived from eight sources of information: OMIM, ClinVar, Orphanet, SwissProt-Humsavar, GeneTests, DISEASES, Novoseek, and GeneCards. The list of primary immunodeficiency (PID) genes was taken from the IUIS report [13]. A gene was considered to be related to a specific

system disease (immune, respiratory or nervous) according to the classification in any of the resources used.

2.2. Transcriptomic Data in Healthy Tissues and Tissue- and Blood Cell-Specific Gene Enrichment Analysis

We used data for tissue expression from the site ‘Downloadable data’ from Human Protein Atlas (HPA) homepage (<https://www.proteinatlas.org/about/download> (accessed on 20 August 2021)). The following data sets were downloaded and analyzed: HPA (43 tissues, $n = 43$) and GTEx ($n = 34$) sets for tissue gene data, and HPA ($n = 19$) and Monaco ($n = 30$) sets for blood cell gene data. All data are based on The Human Protein Atlas version 21.0 and Ensembl version 103.38. HPA sets included data on 19,670 genes, while GTEx and Monaco [15] sets included data on 18,816 genes. All GTEx and Monaco genes were also present in the HPA series. The TissueEnrich R package [16] was used to carry out tissue- and blood cell-specific gene enrichment analysis. Tissue enriched, group enriched, and tissue enhanced genes identified with TissueEnrich were grouped together under the name ‘tissue-specific genes’. The teEnrichment function (with default parameters) was applied to calculate the enrichment of tissue-specific genes in the groups of immune, respiratory, and nervous system disease genes [16].

2.3. Gene Occurrence in Pathology Transcriptomics from ENRICHR Libraries for GEO Studies

In this section, we determined how often each gene was detected as DE in studies samples for a specific tissue. Enrichr libraries Disease Perturbations from GEO down studies and Disease Perturbations from GEO up studies were downloaded from <https://maayanlab.cloud/Enrichr/#libraries> (accessed on 25 August 2021). After exclusion of non-human samples, there were 568 samples collected by crowding [17] from 339 studies. Gene names were checked with HGNC Multi-symbol Checker (http://www.genenames.org/cgi-bin/symbol_checker (accessed on 26 August 2021)) and revised so that one HGNC term combined all synonyms and corresponded to one stable ENSG ID (Ensembl Gene Identifier) (Ensembl version 106). Since stable ENSG ID are kept the same throughout Ensembl releases, ENSG IDs were further used to match genes from HPA and GEO studies. Tissues on which the studies were conducted were checked manually using the original descriptions of the studies on the GEO website. Furthermore, tissues were grouped by types to access the number of tissue-specific DE findings.

2.4. GO Enrichment Analysis

Gene Ontology enrichment analysis was used to functionally characterize DE genes. Enriched biological processes were considered in the sets of DE genes in blood/bone marrow, lung, cerebral cortex, skeletal muscle, and skin. Each sample was analyzed separately with two tools, Enrichr: GO Biological Process 2021 (<https://maayanlab.cloud/Enrichr/> (accessed on 25 August 2021)) and STRING: GO Biological Process (<https://string-db.org/> (accessed on 18 September 2021)). For gene set enrichment analyses, the false discovery rate (FDR) threshold was set at 0.05. We collected data on all enriched GO terms in all samples within tissue type and calculated the frequency of the occurrence of each GO term for the tissue. The REVIGO tool was used for data visualization and reducing redundancy of GO terms [18].

2.5. Mapping HPO Annotations to Genes

To link the results of transcriptomics studies with the variability of phenotypic abnormalities encountered in human genetic diseases, we used standardized terms for phenotypic abnormalities in hereditary diseases, which were obtained from Human Phenotype Ontology (HPO) [19]. The list of HPO gene-phenotype associations in all HPO diseases was downloaded from (https://ci.monarchinitiative.org/view/hpo/job/hpo.annotations/lastSuccessfulBuild/artifact/rare-diseases/util/annotation/genes_to_phenotype.txt (accessed on 19 July 2021)). Each HPO phenotype was mapped to a HPO category available

through search and downloads of phenotypic terms relating to individual diseases. Categorization of the term did not depend on a specific disease. From the total of 27 categories in HPO, the categories “Inheritance” and “Clinical modifier” were excluded.

2.6. Identification of Haploinsufficient Genes

Among genes associated with genetic diseases, there are a relatively large fraction of haploinsufficient genes, which encode proteins required in large or highly titrated amounts. Haploinsufficiency can be defined as prediction of sensitivity to decreased gene dosage. The comparative analysis of haploinsufficient and haplosufficient genes could further clarify the connection between gene expression and variability of phenotypic manifestations of a disease. Haploinsufficiency was analyzed with DECIPHER, which provides haploinsufficiency scores calculated based on predicted haploinsufficiency probabilities (<https://www.deciphergenomics.org/about/downloads/data> (accessed on 18 October 2021)). Scores in the range 0–10% correspond to the highest probability of a gene to be haploinsufficient.

2.7. Statistical Analysis

To analyze differences between continuous variables in compared data sets, we used the non-parametric Mann–Whitney U test. It is used to test the null hypothesis that two compared samples were pooled from the same sample. FDR-corrected p values < 0.05 were considered significant.

3. Results

3.1. Disease Genes

From a total of 9972 genes with reliable proof of being associated with a disease, 9963 genes were extracted from OMIM, ORPHANET, DDG2P, DisGeNet and/or MalaCards (Figure 1A), and nine more genes were found in the IUIS report (Supplementary Tables S1 and S2). Classifications which allowed the identification of genes associated with rare and/or non-rare diseases were available in ORPHANET, DisGeNet, and/or MalaCards. The largest number of genes associated with non-rare diseases was presented in DisGeNet (Figure 1B, Supplementary Table S1).

3.2. Transcriptomic Profile of Disease- and Non-Disease Genes across Healthy Human Tissues and Blood Cells

We examined the transcriptomic profile in healthy tissues with the use of two tissue (HPA and GTEx) and two blood cell (HPA and Monaco) data sets. Correlation analysis revealed high Spearman correlation coefficients for all gene expression levels in the same tissues ($n = 25$, ρ 0.89–0.95) and in the same blood cells ($n = 14$, ρ 0.84–0.92). Identical tissues and blood cells are listed in Figure 2A. Expression profiles in TPM (Transcripts Per Million) were analyzed in specific system disease genes, other disease genes and non-disease genes (Supplementary Figures S1–S3, Supplementary Table S3). A comparison of expression levels of genes involved in the development of immune system diseases and genes associated with non-immune diseases showed that the former had higher levels of RNA expression in all immune related tissues and some other tissues (38/77 in two tissue sets) (Supplementary Table S3). Similar results were found in blood cells (43/49 in two blood cell sets). Lower expression levels were detected in four tissues (nervous and muscle tissues) whereas no differences were registered in 40 tissues, mainly urogenital, endocrine, and nervous (Supplementary Figure S1). Respiratory disease genes compared to non-respiratory disease genes had higher expression levels in eight tissues (8/77) with maximal differences detected in lung tissue (Supplementary Figure S2). Nervous system disease genes were expressed at higher levels than non-nervous system disease genes in 15 tissues (15/77) including nervous tissues (Supplementary Figure S3). In blood cell series, respiratory, and nervous system disease genes mainly did not differ by the level of expression from other disease genes. In all four tissue/blood cell sets, higher

expression levels were typically observed for disease genes compared to non-disease genes (Supplementary Figures S1–S3).

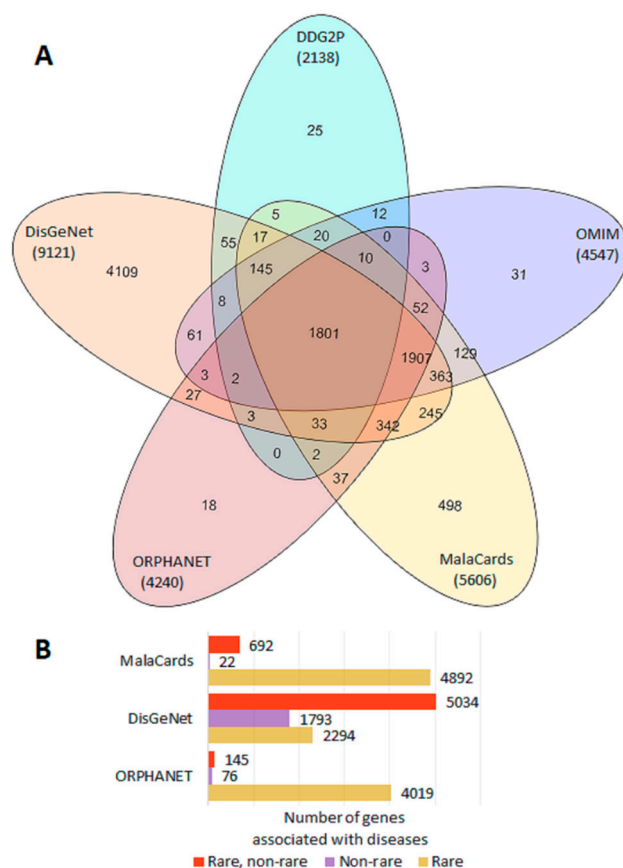


Figure 1. Representation of disease genes in gene–phenotype databases. (A) Venn diagram of gene sharing by OMIM, ORPHANET, DDG2P, DisGeNet, and MalaCards. (B) Number of genes associated with rare and non-rare diseases in ORPHANET, DisGeNet, and MalaCards.

Since gene-phenotype databases differ by sources for data collection and disease classifications, the consistency of transcriptomics results was assessed with a sensitivity analysis for two tissue (HPA and GTEx) data sets. The sequential exclusion of each of the five gene-phenotype databases did not have an essential impact on the results. Compared to other disease genes and non-disease genes, immune disease genes had higher expression levels in immune tissues, respiratory disease genes were expressed at higher levels in lungs, and nervous system disease genes were expressed at higher levels in nervous tissues. Some exceptions were mainly found for certain nervous system related tissues from the GTEx data set (Supplementary Table S4).

3.3. Tissue- and Blood Cell-Specific Gene Enrichment

TissueEnrich was used for tissue- and blood cell-specific gene enrichment analysis. First, we assessed the numbers of tissue-enriched genes and tissue similarity measured by the Jaccard index. The numbers of genes enriched in the same tissues (HPA and GTEx) or blood cells (HPA and Monaco) and Jaccard indices strongly differed (Figure 2A). The Jaccard index varied in the range of 0.15–0.63 for tissues and 0.11–0.50 for blood cells, which showed the diversity of transcriptomics sets. Next, we focused on tissue- and blood cell-specific gene enrichment analysis of immune, respiratory, and nervous system disease genes. Despite a rather strong overlap between these genes (Figure 2B), the gene set enrichment analysis showed characteristic enrichment of immune, respiratory, and nervous system disease genes in corresponding tissues from HPA and GTEx (Figure 2C). Immune disease genes were mainly enriched in immune related tissues (grouped together

from tonsil to NK cells [20]), liver and lung; respiratory disease genes were primarily enriched in lung and also in several immune and muscle tissues; and nervous system disease genes were enriched in brain and muscle tissues. Immune disease genes were overrepresented in almost all blood cells in HPA and Monaco sets with the most significant results obtained for PBMCs. Compared to immune disease genes, respiratory and nervous system disease genes were poorly enriched in blood cells (Figure 2C). Thus, despite the diversity of transcriptomics sets, the results of the gene set enrichment analysis were concordant within both tissue- and blood cell transcriptomes. These results, in particular, supported the literature data on higher expression levels of disease genes in pathologically linked tissues compared to other tissues.

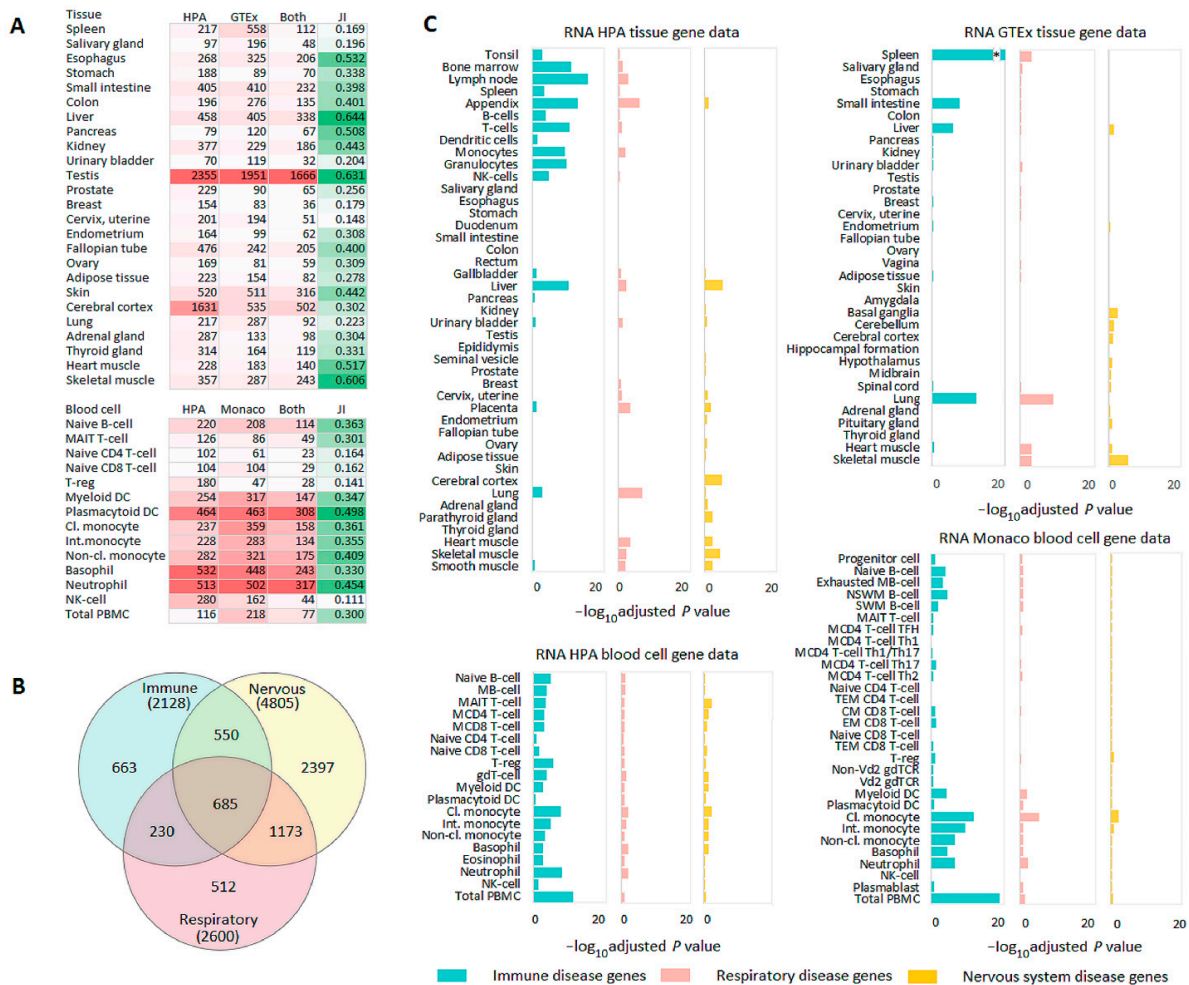


Figure 2. Tissue- and blood cell-specific gene enrichment analysis. (A) Heatmap with the number of genes (colored in red) enriched in the same tissues (HPA and GTEX) or blood cells (HPA and Monaco). Enriched genes include genes relating to all categories of enrichment in a corresponding tissue (tissue enriched, group enriched, and tissue enhanced). Jaccard index (JI) (colored in green) shows tissue or blood cell similarity determined by gene enrichment analysis with TissueEnrich. (B) Venn diagram showing an overlap among immune, respiratory, and nervous system disease genes. (C) TissueEnrich results for immune, respiratory, and nervous system disease genes. Abbreviations: Cl., Classical (monocyte); CM, Central memory (CD8 T-cell); EM, Effector memory (CD8 T-cell); Int., Intermediate (monocyte); M, Memory (MB-cell, MCD4 T-cell, MCD8 T-cell); Non-cl., Non-classical (monocyte); NSWM, Non-switched memory (B-cell); SWM, Switched memory (B-cell); TEM, Terminal effector memory (CD4, CD8 cells); * $-\log_{10}$ adjusted p value = 43.22.

3.4. Gene Occurrence in DE Studies and Their Expression Profiles in Normal Tissues

GEO disease perturbations samples from the Enrichr libraries included data on 18,920 and 18,457 down- and up-regulated genes, respectively. Taking into account strong gene expression correlation among closely related tissues [21], we grouped all tissues, on which GEO studies were conducted by tissue types (Supplementary Table S5). After that, we selected the disease genes (Supplementary Table S1) in all GEO samples (9021/9972, 90.5% and approximately half of all genes included in the Enrichr libraries) (Supplementary Table S6) and demonstrated that they had a much larger total number of GEO associations and gene-unique tissue associations compared to non-disease genes (p value $< 1 \times 10^{-300}$) (Supplementary Figure S4). Then, we calculated the number of disease genes associated with phenotypes presented in GEO studies ($n = 4848$) and the number of gene-phenotype associations supported by expression changes ($n = 955$) (Supplementary Table S7). Out of 955 genes, 100 genes were both down and up-regulated. The percentage of samples with the associations (from the total number of samples per phenotype taking into account the analyzed tissue) was 24.4% and 22.4% for down- and up-regulated genes, respectively. Thus, the disease genes were actively involved in a variety of pathological processes, often unrelated to the diseases, to which they predispose.

Next, we analyzed the pools of DE genes by assessing the expression of each gene in normal tissues and the number of gene records in the libraries of GEO studies. Figure 3A shows the distribution of genes by their presence/absence among DE genes in GEO samples on the relevant tissues (blood/bone marrow, lung, brain) as a function of their expression level in normal tissues (total PBMC from HPA and Monaco blood cells, lung, and cerebral cortex from HPA and GTEx tissues). The higher the gene expression was in a tissue (in normal state), the higher the probability was that it registered as changing expression between cases and controls. A similar phenomenon was recorded for tissue-specific genes (Figure 3B). Summarized TPMs for each gene across all tissues in HPA and GTEx were correlated with the number of gene-unique tissue associations in GEO samples (Figure 3C, Supplementary Table S6). Although tissue sets differed greatly, there was a moderate correlation (ρ 0.39–0.44) between the sum of TPMs and the number of tissues, in which the expression of the gene changed.

3.5. GO Enrichment Analysis

To uncover biological processes (BP) which are characteristic for DE genes in GEO studies performed on the selected tissues, we considered each sample separately with the use of GO tools. Some BP were typical for many samples in the same or even different tissues. Top 20 frequent terms in Enrichr GO Biological Process 2021 are presented in Figure 4A (blood/bone marrow, lung, and brain) and Supplementary Figure S5A (skeletal muscle and skin). The most frequent term for the genes down-regulated in blood/bone marrow, lung, and brain, namely 'neutrophil-mediated immunity', was represented in 68.2%, 68.4% and 42.31% of samples, respectively; for up-regulated genes, this term occurred in 70.9%, 36.1% and 64.0% of samples. Noteworthy, down- and up-regulated genes were frequently enriched in the same BP (Figure 4B, Supplementary Figure S5B). The results from STRING GO enrichment supported these observations (Supplementary Tables S8 and S9). The concordant results from the two GO enrichment tools showed that immune response, translation, and tissue-specific processes were most often disturbed in all considered tissues. The Enrichr results for the subsets of neoplastic diseases (Supplementary Table S5) fit these conclusions (Supplementary Table S10).

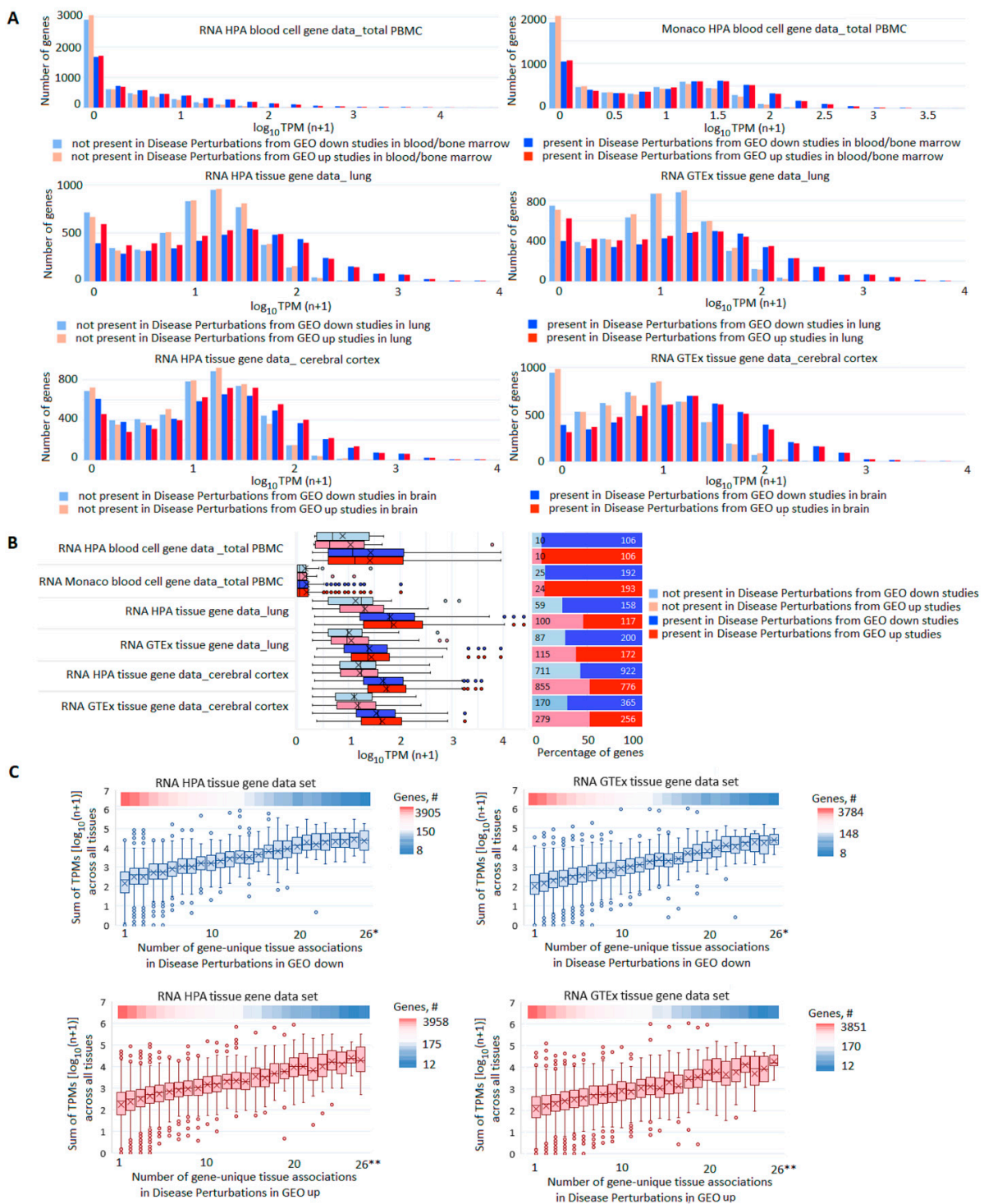


Figure 3. Comparison of gene expression profiles in norm and pathology. **(A)** Distribution of genes by their presence/absence in GEO studies’ libraries in dependence on their expression level in corresponding normal tissues. For all pairs of compared sets, the differences in expression levels between genes present or absent in GEO libraries were highly significant (p value $< 1 \times 10^{-300}$). **(B)** Expression levels (left panel) and the number (right panel) of tissue-specific genes present/absent among DE genes in GEO libraries. **(C)** Box plots reflecting correlations between the sum of TPMs across all tissues in HPA or GTEx data sets and the number of GEO tissues in which the expression of the gene changed. The upper heatmap band on each panel reflects the number of genes at each point; * 26–28; ** 26–30.

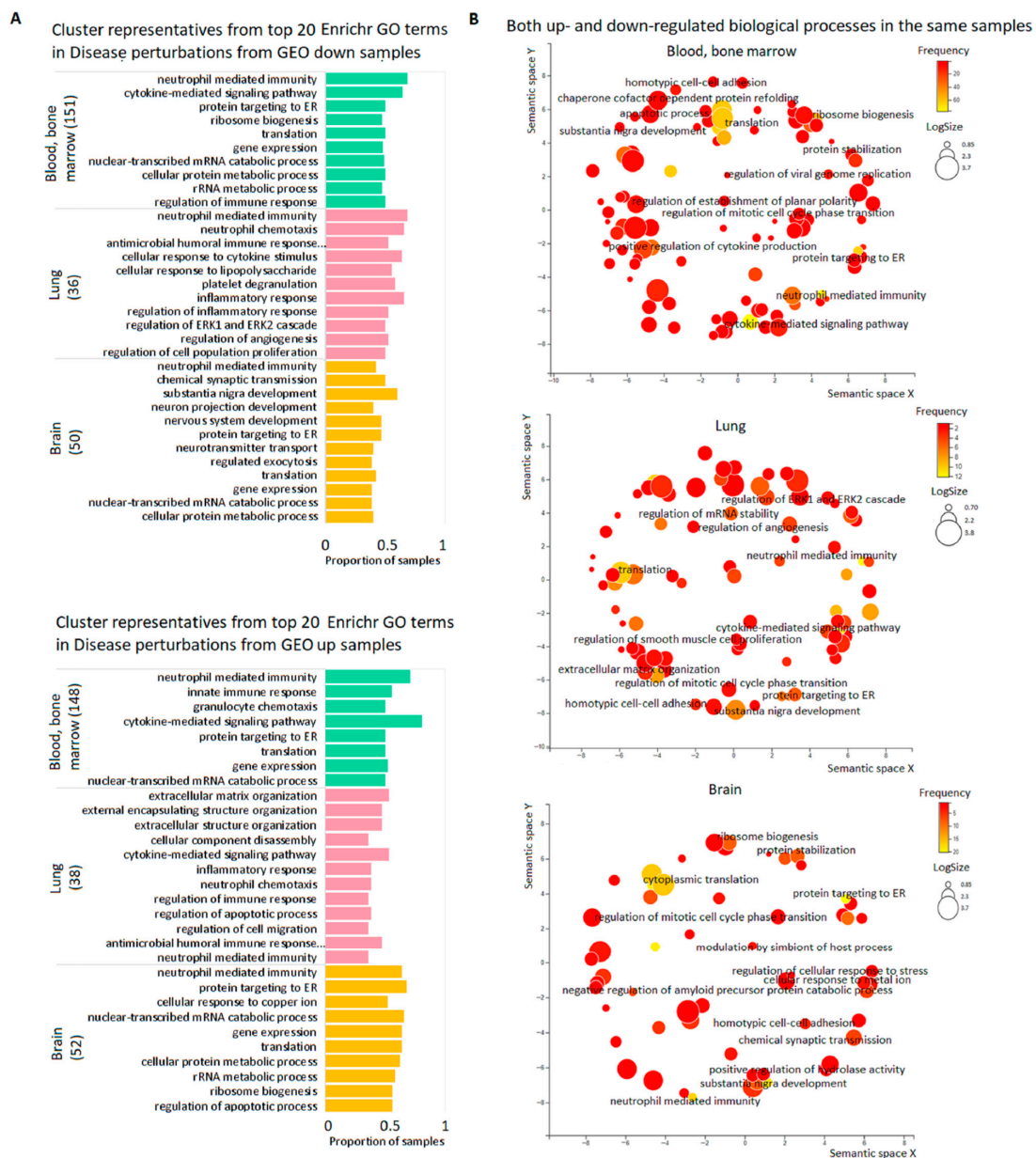


Figure 4. Top biological processes dysregulated in the pools of tissue samples. **(A)** Cluster representatives for the most frequent BP dysregulated in blood/bone marrow, lung, and brain. Cluster representatives for GO terms (terms remaining after reducing redundancy) were obtained using the REVIGO tool. The number of samples for each tissue is indicated in brackets. **(B)** REVIGO scatterplots for the same BP down- and up-regulated in the same samples. GO terms are plotted according to log size on the x -axis and frequency of occurrence (yellow is larger, red is smaller) in the samples on the y -axis. The size of the circles is proportional to the underlying frequency of the GO term in a reference database—the EBI GOA database. Functionally similar GO terms are located close to each other.

3.6. Gene-Based Analysis of Associated HPO Terms and Expression in Norm and Pathology

From a total of 4531 HPO genes, 4526 (99.9%) were present in our list of disease genes. The set of phenotypic terms included 8347 unique items, which were assigned to 25 categories corresponding to various body systems or types of manifestation (Supplementary Table S11). For our set of 4526 genes, we compared tissue-specific and non-specific genes (total PBMC, lung, and cerebral cortex) by the number of terms relating to phenotypic categories “Blood and blood-forming tissues”, “Respiratory”, and “Nervous system”, and the number of GEO

samples, in which expression changes in the relevant tissues were registered (Figure 5A). A similar analysis was performed for gene sets having associations within phenotypic categories “Musculature” and “Integument” (Supplementary Figure S6). In each series of comparisons, we found highly significant differences. At the same time, when considering the pools of HPO terms/categories and genes enriched/enhanced in any tissue, it turned out that tissue-specific genes had fewer associations compared to tissue non-specific genes (Figure 5B). We next analyzed a subset of haploinsufficient genes ($n = 841$) defined as having haploinsufficiency (HIS) scores $\leq 10\%$ (Supplementary Table S11). HIS genes compared to other disease genes from HPO had (i) higher TPM sums across all tissues in HPA and GTEx, (ii) higher number of associations with HPO terms and categories, and (iii) higher number of GEO samples and tissues, in which expression changes were recorded (Figure 5C). Thus, the level of gene expression in healthy tissues correlated with the number of associations found in DE studies, as well as with the number of phenotypic terms and categories.

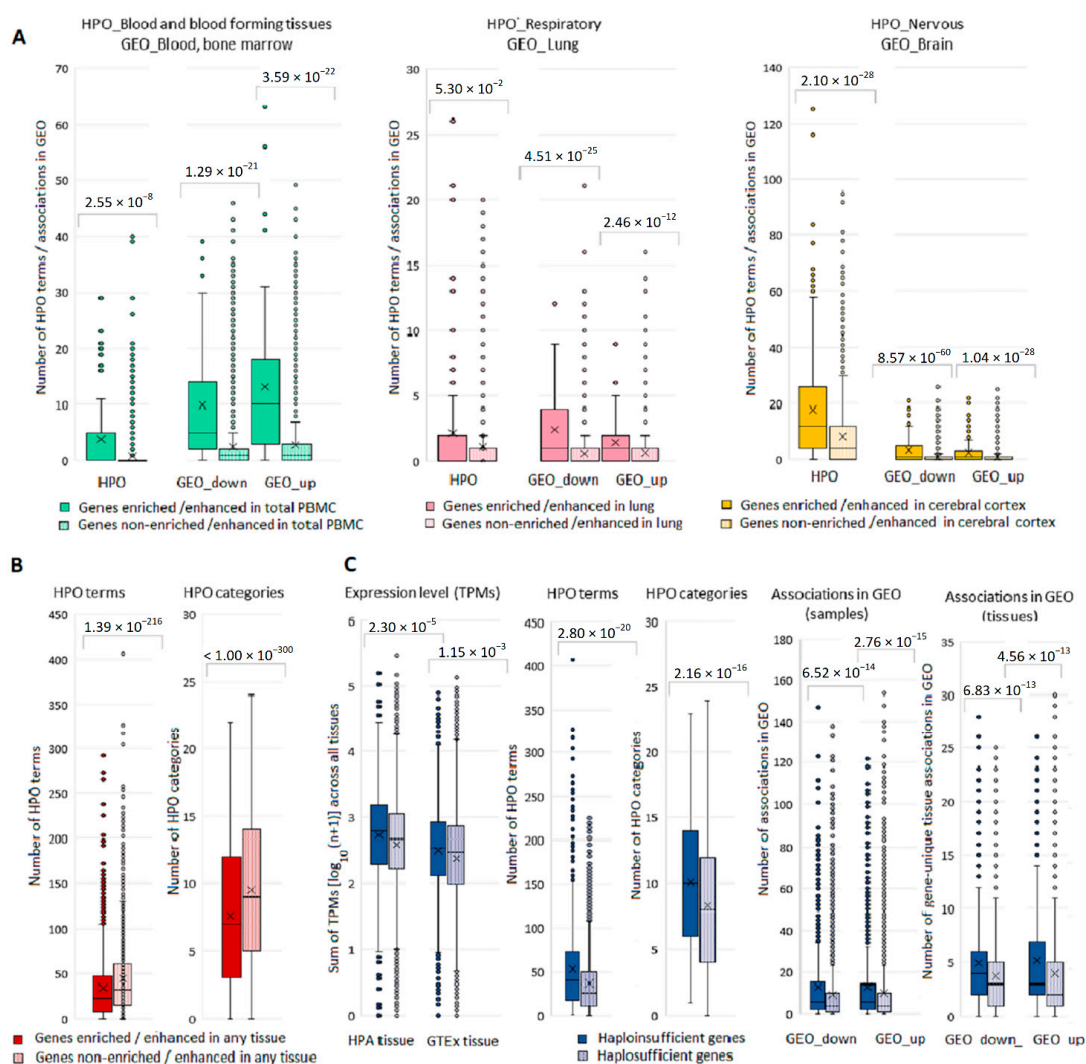


Figure 5. Box-plots depicting differences in the number of phenotypic associations and associations gene expression—disease in GEO for tissue specific and haploinsufficient genes. (A) Tissue-specific and non-specific genes with the number of terms within relevant phenotypic categories and the number of GEO samples, in which expression changes in relevant tissues were registered. (B) Tissue-specific and non-specific genes (in any tissue) with the number of terms and categories from the entire pool of HPO terms/categories. (C) Haploinsufficient and haplosufficient genes and their TPM sums across all tissues in HPA and GTEx, number of associations with HPO terms and categories and number of GEO samples and tissues, in which expression changes were recorded.

4. Discussion

In this study, we considered gene expression profiles in healthy tissues and the frequency of gene occurrence as DE in pathology studies and compared these data with a diversity of phenotypic manifestations within a corresponding organ system. We created a set of disease genes using six manually curated gene-phenotype databases. To our knowledge, this is the largest dataset of this kind. With this dataset, we (i) validated the previous findings [6] of higher expression levels of disease genes in pathologically linked tissues than in other tissues, and (ii) demonstrated that disease genes also had higher expression levels than non-disease genes in almost all tissues and blood cells. At the whole genome level, we also showed that (iii) genes with high expression levels in the tissue (in norm) were more likely to be registered in GEO pathology studies in the relevant tissue. Based on the results of GO enrichment analyses for the sets of DE genes in GEO studies, we concluded that (iv) immune response, translation, and tissue-specific processes were the processes most often disturbed in considered tissues and (v) moreover, these down- and up-regulated blocks of genes were frequently found in the same samples, which suggests their compensatory involvement in the same biological processes. Using the subset of disease genes presented in the HPO database, we also established that (vi) the level of gene expression in one/several (for tissue-specific genes) or many (for haploinsufficient genes) healthy tissues correlated with the number of phenotypic terms, i. e. the variety of phenotypic manifestations, within the relevant top-level categories corresponding to organ systems.

Disease genes are often overexpressed in normal tissues where gene defects lead to the development of pathology [6,22]. In confirmation of this phenomenon, we demonstrated higher expression levels for the immune, respiratory, and nervous system disease genes in the corresponding tissues. We also observed that disease genes from our set had higher expression levels than non-disease genes in almost all tissues and blood cells. The disease gene set included highly expressed genes specific for different tissues. Averaging of expression levels in the groups of disease and non-disease genes could partially explain this finding. Beyond that, independently of tissue specificity, genes with relatively high expression levels contribute to the maintenance of tissue regulatory framework and high expression levels may be essential for the functional activity of the tissue [6,23].

In case-control studies comparing transcript levels between healthy and sick individuals, gene expression changes are likely to be a consequence of a disease, rather than its cause or a simple correlation [24]. This hypothesis is consistent with the results of our study, which demonstrated that the disease genes changed their expression not only and not so much in diseases to which they predispose, but also in other pathologies. A relatively small number of gene-phenotype associations supported by expression changes, may be additionally explained by the multifactorial and polygenic nature of the majority of GEO diseases. The disease genes, as well as highly expressed genes in general, were registered as DE genes more often and in more tissues than other genes. Given these results, we supposed that pathology hits highly expressed genes as the main determinants of normal tissue structure and function. Highly expressed genes evolve slowly engaging in multiple interactions in gene-gene networks [25–27]. Moreover, highly expressed genes are enriched for biologically important and/or haploinsufficient genes exhibiting higher levels of connectivity compared to other genes [25,28]. High network connectivity was also shown for disease genes [29]. Hence, within tissue interactomes, highly expressed/disease genes might crucially influence dynamic gene expression through their interplay with other genes. The observation of a large number of disease-affected tissues for highly expressed genes is in accordance with the fact that gene expression level correlates with the expression breadth (number of tissues, in which the gene is expressed) [30]. Taken together, these data contribute to understanding of how and why highly expressed/disease genes are frequently dysregulated in pathology studies.

Both down- and up-regulated genes were broadly overrepresented in biological processes relating to immune response, translational control and ribosome biogenesis and

tissue-specific perturbations. Our modules did not completely coincide with those identified in the study of Crow et al. [4]. The above-mentioned study was designed to generate a general view from the expression studies, thus all conditions, of which only 25% were determined as diseases, were analyzed together. We did not find overrepresentation of sex and extracellular matrix related genes among preferably DE genes, however, our tissue-based approach allowed us to identify a module that was unique for each tissue. The translational module from our study might be considered interconnected with the modules associated with cell cycle and response to stimulus from the study of Crow et al. [4]. The only consistent result for both studies was the block of DE immune genes.

Due to the direct link between blood and immunity, blood was widely used to study immune-mediated diseases, however, blood is also a useful resource to characterize an individual's current health status. Peripheral blood contains a wide variety of cells which protect the body from exogenous and endogenous pathogens, thus responding to any alteration of physiological condition of the body and producing messengers between the immune system and other tissues [31]. Multiple diseases, aging, and obesity largely share blood transcriptome profile associated with inflammation and immunity [32]. In our study, immune response genes were also expressed differently in immune and non-immune diseases in tissues other than blood. An expression profile of a damaged tissue reflects tissue infiltration by immune cells implicated in disease development and resolving. Ribosome biogenesis and protein translation was another important block of biological processes often affected in different diseases. These processes are essential for cell growth, proliferation, and response to stress; therefore, their dysregulation affects cellular function thus contributing to cancer and diseases other than cancer [33,34]. As for tissue-specific processes, disease-related expression changes in affected tissues are associated with suppressing and subsequent restoration or compensation of tissue functions. On a more general level, we can speculate that an opposite direction of biological processes may occur as a consequence of the body's efforts to maintain tissue homeostasis in response to the impact of the stimulus [10,35].

A gene-based comparison of phenotypic diversity in hereditary diseases with the number of expression associations in multifactorial diseases showed their correlation. Phenotypic manifestations of rare diseases are directly related to the gene that is the root cause of the disease, while DE in common diseases reflects the consequences of the disease rather than its prerequisite. However, in any case these results indicate phenotypic impact of the gene at the organ/tissue level. Tissue-specific genes showed organ and tissue-specific effects, while highly and widely expressed haploinsufficient genes exhibited an extended phenotypic impact. The diversity of phenotypic manifestations and DE reflects certain gene properties and can be considered as an indicator of the biological significance of a gene [12].

Among the limitations of the study, it is necessary to note the list of disease genes, which contains most well-characterized genes. Many of them are highly expressed and have broad tissue specificity. GEO datasets may be biased towards better detection of highly expressed genes, as a transcriptome analysis provides more reliable quantitative measurements for highly expressed genes. These limitations reflect a problem which is clearly visible when summarizing the results and is less noticeable in the framework of individual studies aimed at defining DE genes and enriched GO biological processes. Our analysis contributes to the field confirming the need for a nuanced interpretation of DE findings in pathology studies. Our other limitations are related to different methods for the sample and library preparations in GEO studies and the use of the restricted set of GEO samples extracted via a crowdsourcing project. However, this approach also has some strengths, which partially compensate for the shortcomings, since all GEO samples were manually validated for quality [17]. Besides, splitting a series into samples increases the diversity of prioritized genes beneficially influencing the search and interpretation of commonalities and differences in gene-phenotype relationships [36].

5. Conclusions

Our results support the conclusion on predictable DE for a specific fraction of genes [4]. These genes are generally highly expressed and are enriched for disease genes. The frequent overlaps between down- and up-regulated BP found in our study are in line with the statement about the predictability of DE of genes belonging to certain functional blocks. These are primarily genes implicated in immune response, translation, and tissue-specific processes. The expression level in healthy tissues correlates with both the number of DE findings and diversity of phenotypic associations in pathology studies. Establishing a link between expression and phenotype contributes to the deeper and more nuanced understanding of the mode of development and progression of human diseases.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13071168/s1>. Supplementary Figure S1: Expression profiles of immune disease related genes, other disease genes and non-disease genes. Genes with TPM < 1 were excluded. FDR corrected p values are marked with asterisks: * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.00005$. Summary statistics are given in Supplementary Table S3; Supplementary Figure S2: Expression profiles of respiratory disease related genes, other disease genes and non-disease genes. Genes with TPM < 1 were excluded. FDR corrected p values are marked with asterisks: * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.00005$. Summary statistics are given in Supplementary Table S3; Supplementary Figure S3: Expression profiles of nervous system disease related genes, other disease genes and non-disease genes. Genes with TPM < 1 were excluded. FDR corrected p values are marked with asterisks: * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$, **** $p < 0.00005$. Summary statistics are given in Supplementary Table S3; Supplementary Figure S4: Box-plots depicting differences in the number of associations in GEO samples for disease genes compared to non-disease genes; Supplementary Figure S5: The results of GO enrichment analysis. (A) Cluster representatives for the most frequent BP dysregulated in skeletal muscle and skin. Cluster representatives for GO terms (terms remaining after reducing redundancy) were obtained using the REVIGO tool. The number of samples for each tissue is indicated in brackets. (B) REVIGO scatterplots for the same BP down-and up-regulated in the same samples. GO terms are plotted according to log size on the x -axis and frequency of occurrence (yellow is larger, red is smaller) in the samples on the y -axis. The size of the circles is proportional to the underlying frequency of the GO term in a reference database—the EBI GOA database. Functionally similar GO terms are located close to each other; Supplementary Figure S6: Box-plots depicting differences in the number of phenotypic associations within categories Musculature and Integument and gene expression-disease associations in the relevant tissues in GEO for tissue-specific and tissue non-specific genes; Supplementary Table S1: The list of disease genes from OMIM, ORPHANET, DDG2P, IUIS, DisGeNet and MalaCards. The total number of diseases per gene in each database is indicated. The numbers of immune, nervous, and respiratory diseases are specified in accordance with the database classifications. Examples of respiratory, nervous, and ‘other’ diseases include no more than three diseases per gene. An example of a complete list of phenotypes for a given gene is shown for genes associated with immune diseases in Supplementary Table S2. The same disease can belong to different classifications. Columns “Rare-1, Non-rare-2” are present in ORPHANET, DisGeNet and MalaCards blocks. In ORPHANET ‘2’ indicates genes that are associated with only non-rare phenotypes in Europe; in DisGeNet ‘2’ indicates genes that are not associated with OMIM/ORPHANET phenotypes and are not related to Congenital, Hereditary, and Neonatal Diseases and Abnormalities according MeSH classifications; MalaCards ‘2’ indicates genes that are not associated with fetal, genetic or rare phenotypes according to MalaCards global classifications; Supplementary Table S2: The list of immune disease genes with the full list of specific phenotypes identified as related to the immune system in accordance with database classifications; Supplementary Table S3: Summary statistics for a comparison of expression levels of (i) genes involved in the development of immune, respiratory and nervous system diseases, (ii) genes associated with non-immune, non-respiratory and non-nervous system diseases and (iii) non-disease genes. Genes with TPM < 1 were excluded. The highlighted in bold data show that genes involved in the development of immune system diseases had higher levels of RNA expression in all immune related tissues and some other tissues (38/77 in two tissue sets) compared to genes associated with non-immune diseases; Supplementary Table S4: Sensitivity analysis for a comparison of expression levels of (i) genes involved in the development of immune, respiratory and

nervous system diseases, (ii) genes associated with non-immune, non-respiratory and non-nervous system diseases and (iii) non-disease genes. Genes with TPM < 1 were excluded. Data are shown for immune, respiratory and nervous system related tissues. The highlighted in bold data show results for disease genes in pathologically linked tissues. Adjusted *p* values (*P_{adj}*) were calculated taking into account all tissues (HPA, *n* = 43; GTEx, *n* = 34) and the number of samples compared. Supplementary Table S5: The list of GEO samples and tissues considered in the study; Supplementary Table S6: The list of differentially expressed (DE) genes in GEO samples with summary statistics data. Enriched genes include genes relating to all categories of enrichment in a corresponding tissue (tissue enriched, group enriched, tissue enhanced); Supplementary Table S7: The list of disease genes linked to the diseases from GEO studies within the Enrichr libraries. The table includes data for genes associated with the same diseases in the gene-phenotype databases considered in the paper and in GEO studies; Supplementary Table S8: The results of GO enrichment analysis with STRING. Dysregulated biological processes were analyzed separately for each sample. The differences in the number of samples between down- and up-regulated genes are due to non-significant enrichment results for some samples; Supplementary Table S9: The results of GO enrichment analysis via STRING processed with REVIGO: cluster representatives for the same BP down- and up-regulated in the same samples; Supplementary Table S10: The results of GO enrichment analysis with Enrichr for the subsets of neoplastic diseases. Disease classification is given in Supplementary Table S5. Dysregulated biological processes were analyzed separately for each sample; Supplementary Table S11: The list of 4526 HPO genes considered in the study. The table includes data on the number of gene—unique phenotypic term associations within 25 categories, haploinsufficiency scores, data on the results from the TissueEnrich analysis and the number of gene expression-disease associations in the Enrichr libraries for GEO disease studies.

Author Contributions: L.E.S. conceived the idea. All authors contributed to the study design. S.K.A. and A.N.K. supervised the project. D.S.K., D.A.S. and L.E.S. analyzed data and prepared figures. S.K.A. and A.N.K. helped to interpret the data. D.S.K. and L.E.S. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the state assignments of the Ministry of Education and Science of Russia (No. 0112-2019-0002 “Genetic technologies in biology, medicine, agriculture and environmental management activities” and No. FGWS-2022-0001 “Multiple organ dysfunction and insufficiency in critical conditions: leading mechanisms of development, new methods of diagnosis and treatment”).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Material.

Acknowledgments: We would like to thank Simon Fishilevich for providing the MalaCards source data.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Rodriguez-Esteban, R.; Jiang, X. Differential gene expression in disease: A comparison between high-throughput studies and the literature. *BMC Med. Genom.* **2017**, *10*, 59. [[CrossRef](#)] [[PubMed](#)]
2. Sonawane, A.R.; Weiss, S.T.; Glass, K.; Sharma, A. Network Medicine in the Age of Biomedical Big Data. *Front. Genet.* **2019**, *10*, 294. [[CrossRef](#)] [[PubMed](#)]
3. Hekselman, I.; Yeger-Lotem, E. Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nat. Rev. Genet.* **2020**, *21*, 137–150. [[CrossRef](#)] [[PubMed](#)]
4. Crow, M.; Lim, N.; Ballouz, S.; Pavlidis, P.; Gillis, J. Predictability of human differential gene expression. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 6491–6500. [[CrossRef](#)]
5. Sigalova, O.M.; Shaeiri, A.; Forneris, M.; Furlong, E.E.; Zaugg, J.B. Predictive features of gene expression variation reveal mechanistic link with differential expression. *Mol. Syst. Biol.* **2020**, *16*, e9539. [[CrossRef](#)]
6. Lage, K.; Hansen, N.T.; Karlberg, E.O.; Eklund, A.C.; Roque, F.S.; Donahoe, P.K.; Szallasi, Z.; Jensen, T.S.; Brunak, S. A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20870–20875. [[CrossRef](#)]

7. Silva, A.S.; Wood, S.H.; van Dam, S.; Berres, S.; McArdle, A.; de Magalhães, J.P. Gathering insights on disease etiology from gene expression profiles of healthy tissues. *Bioinformatics* **2011**, *27*, 3300–3305. [[CrossRef](#)]
8. Lagreid, A.; Hvidsten, T.R.; Midelfart, H.; Komorowski, J.; Sandvik, A.K. Predicting gene ontology biological process from temporal gene expression patterns. *Genome Res.* **2003**, *13*, 965–979. [[CrossRef](#)]
9. Xu, G.; Liu, C.; Liang, T.; Zhang, Z.; Jiang, J.; Chen, J.; Xue, J.; Zeng, H.; Lu, Z.; Zhan, X. Gene expression profile and bioinformatics analysis revealed key molecular characteristics of chordoma-before and after TNF- α treatment. *Medicine* **2020**, *99*, e18790. [[CrossRef](#)]
10. Saxena, V.; Orgill, D.; Kohane, I. Absolute enrichment: Gene set enrichment analysis for homeostatic systems. *Nucleic Acids Res.* **2006**, *34*, e151. [[CrossRef](#)]
11. Onur, T.S.; Laitman, A.; Zhao, H.; Keyho, R.; Kim, H.; Wang, J.; Mair, M.; Wang, H.; Li, L.; Perez, A.; et al. Downregulation of glial genes involved in synaptic function mitigates Huntington's disease pathogenesis. *eLife* **2021**, *10*, e64564. [[CrossRef](#)] [[PubMed](#)]
12. Vogt, I.; Prinz, J.; Worf, K.; Campillos, M. Systematic analysis of gene properties influencing organ system phenotypes in mammalian perturbations. *Bioinformatics* **2014**, *30*, 3093–3100. [[CrossRef](#)] [[PubMed](#)]
13. Tangye, S.G.; Al-Herz, W.; Bousfiha, A.; Chatila, T.; Cunningham-Rundles, C.; Etzioni, A.; Franco, J.L.; Holland, S.M.; Klein, C.; Morio, T.; et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. *J. Clin. Immunol.* **2020**, *40*, 24–64. [[CrossRef](#)] [[PubMed](#)]
14. Sattler, S. The Role of the Immune System beyond the Fight against Infection. *Adv. Exp. Med. Biol.* **2017**, *1003*, 3–14. [[CrossRef](#)] [[PubMed](#)]
15. Monaco, G.; Lee, B.; Xu, W.; Mustafah, S.; Hwang, Y.Y.; Carré, C.; Burdin, N.; Visan, L.; Ceccarelli, M.; Poidinger, M.; et al. RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* **2019**, *26*, 1627–1640.e7. [[CrossRef](#)]
16. Jain, A.; Tuteja, G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **2019**, *35*, 1966–1967. [[CrossRef](#)]
17. Wang, Z.; Monteiro, C.; Jagodnik, K.; Fernandez, N.F.; Gundersen, G.W.; Rouillard, A.D.; Jenkins, S.L.; Feldmann, A.S.; Hu, K.S.; McDermott, M.G.; et al. Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd. *Nat. Commun.* **2016**, *7*, 12846. [[CrossRef](#)]
18. Supek, F.; Bošnjak, M.; Škunca, N.; Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **2011**, *6*, e21800. [[CrossRef](#)]
19. Köhler, S.; Carmody, L.; Vasilevsky, N.; Jacobsen, J.O.B.; Danis, D.; Gouridine, J.P.; Gargano, M.; Harris, N.L.; Matentzoglou, N.; McMurry, J.A.; et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* **2019**, *47*, D1018–D1027. [[CrossRef](#)]
20. Uhlen, M.; Karlsson, M.J.; Zhong, W.; Tebani, A.; Pou, C.; Mikes, J.; Lakshmikanth, T.; Forsström, B.; Edfors, F.; Odeberg, J.; et al. A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. *Science* **2019**, *366*, eaax9198. [[CrossRef](#)]
21. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **2017**, *550*, 204–213. [[CrossRef](#)] [[PubMed](#)]
22. Goh, K.I.; Cusick, M.E.; Valle, D.; Childs, B.; Vidal, M.; Barabási, A.L. The human disease network. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8685–8690. [[CrossRef](#)] [[PubMed](#)]
23. Sonawane, A.R.; Platig, J.; Fagny, M.; Chen, C.Y.; Paulson, J.N.; Lopes-Ramos, C.M.; DeMeo, D.L.; Quackenbush, J.; Glass, K.; Kuijjer, M.L.; et al. Understanding tissue-specific gene regulation. *Cell Rep.* **2017**, *21*, 1077–1088. [[CrossRef](#)] [[PubMed](#)]
24. Porcu, E.; Sadler, M.C.; Lepik, K.; Auwerx, C.; Wood, A.R.; Weihs, A.; Sleiman, M.S.B.; Ribeiro, D.M.; Bandinelli, S.; Tanaka, T.; et al. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat. Commun.* **2021**, *12*, 5647. [[CrossRef](#)]
25. Drummond, D.A.; Bloom, J.D.; Adami, C.; Wilke, C.O.; Arnold, F.H. Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14338–14343. [[CrossRef](#)]
26. Gout, J.F.; Kahn, D.; Duret, L.; Paramecium Post-Genomics Consortium. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet.* **2010**, *6*, e1000944. [[CrossRef](#)]
27. Zhang, W.; Landback, P.; Gschwend, A.R.; Shen, B.; Long, M. New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* **2015**, *16*, 202. [[CrossRef](#)]
28. Norris, M.; Lovell, S.; Delneri, D. Characterization and prediction of haploinsufficiency using systems-level gene properties in yeast. *G3 (Bethesda)* **2013**, *3*, 1965–1977. [[CrossRef](#)]
29. Kim, S.S.; Dai, C.; Hormozdiari, F.; van de Geijn, B.; Gazal, S.; Park, Y.; O'Connor, L.; Amariuta, T.; Loh, P.R.; Finucane, H.; et al. Genes with High Network Connectivity Are Enriched for Disease Heritability. *Am. J. Hum. Genet.* **2019**, *104*, 896–913. [[CrossRef](#)]
30. Park, J.; Xu, K.; Park, T.; Yi, S.V. What are the determinants of gene expression levels and breadths in the human genome? *Hum. Mol. Genet.* **2012**, *21*, 46–56. [[CrossRef](#)]
31. Sakai, Y.; Nasti, A.; Takeshita, Y.; Okumura, M.; Kitajima, S.; Honda, M.; Wada, T.; Nakamura, S.; Takamura, T.; Tamura, T.; et al. Eight-year longitudinal study of whole blood gene expression profiles in individuals undergoing long-term medical follow-up. *Sci. Rep.* **2021**, *11*, 16564. [[CrossRef](#)] [[PubMed](#)]
32. Schmidt, M.; Hopp, L.; Arakelyan, A.; Kirsten, H.; Engel, C.; Wirkner, K.; Krohn, K.; Burkhardt, R.; Thiery, J.; Loeffler, M.; et al. The Human Blood Transcriptome in a Large Population Cohort and Its Relation to Aging and Health. *Front. Big Data* **2020**, *3*, 548873. [[CrossRef](#)] [[PubMed](#)]

33. Tahmasebi, S.; Khoutorsky, A.; Mathews, M.B.; Sonenberg, N. Translation deregulation in human disease. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 791–807. [[CrossRef](#)] [[PubMed](#)]
34. Turi, Z.; Lacey, M.; Mistrik, M.; Moudry, P. Impaired ribosome biogenesis: Mechanisms and relevance to cancer and aging. *Aging* **2019**, *11*, 2512–2540. [[CrossRef](#)]
35. Armingol, E.; Officer, A.; Harismendy, O.; Lewis, N.E. Deciphering cell-cell interactions and communication from gene expression. *Nat. Rev. Genet.* **2021**, *22*, 71–88. [[CrossRef](#)]
36. Fine, R.S.; Pers, T.H.; Amariuta, T.; Raychaudhuri, S.; Hirschhorn, J.N. Benchmark: An Unbiased, Association-Data-Driven Strategy to Evaluate Gene Prioritization Algorithms. *Am. J. Hum. Genet.* **2019**, *104*, 1025–1039. [[CrossRef](#)]