SOUTHERN JOURNAL OF SCIENCES

ESTABLISHED IN 1993

Review paper

THE TREND TOWARDS PHENOME-WIDE ASSOCIATION STUDIES (PheWASs) IN COVID-19 RESEARCH

ATÇEKEN, Nazente [1 *]; KOZALAK, Gül [2,3]; ÖZGÜL, Rıza Köksal [4]; SYED, Hamzah[1,5]

[1] Koc University Research Center for Translational Medicine, Koç University, Sariyer, Istanbul, Turkey 34450.

[2] Faculty of Engineering and Natural Sciences (FENS), Sabanci University, Istanbul, Turkey.

[3] Center of Excellence for Functional Surfaces and Interfaces for Nano-Diagnostics (EFSUN), Sabanci University, Istanbul, Turkey.

[4] Department of Pediatric Metabolism, Institute of Child Health, Faculty of Medicine, Hacettepe University, Ankara, Turkey.

[5] Koç University School of Medicine, Istanbul, Turkey.

** Corresponding author*
e-mail: biyolog_nazente@hotmail.com

Received 10 October 2022; received in revised form 28 November 2022; accepted 10 December 2022

ABSTRACT

Background: Coronavirus Disease-2019 (COVID-19) appears in individuals asymptomatically and in various symptomatic forms. Symptomatic diversity can result in diagnosis failures, hospitalization, admission to intensive care, multi-organ failure, and death. The causes and risk factors of the severity of disease symptoms are uncertain. This uncertainty can only be resolved by elucidating the effects of host genes and genetic variations on different phenotypes. **Aim:** This review aimed to emphasize the importance of large-scale genotype-phenotype correlation studies in elucidating the phenotypic diversity in COVID-19 disease. **Methods:** All publications related to Phenome-Wide Association Study (PheWAS) in the PubMed database were searched. PheWAS studies applied to COVID-19 patients have been identified. In addition, studies applied to the genome-wide association study (GWAS)- Electronic health records (EHRs) data and additionally matched to the gene expression data were systematically reviewed. The latest PheWAS methodology and its importance in Large-scale genotype-phenotype correlations are discussed within the context of published COVID-19 studies. **Results**: According to our PubMed search data, there are few PheWAS studies on COVID-19 disease. This review explains the use of PheWAS studies applied to health records and GWAS data, and colocalization studies applied to expression quantitative trait locus (eQTL) analysis to understand the phenotypic variability of COVID-19. **Discussion**; Although there is a very limited number of PheWAS studies on COVID-19 diseases, these studies have obtained important data. At the current stage, there is a need for such studies in COVID-19 research. **Conclusions:** PheWAS is an ideal method for large-scale genotype-phenotype correlation studies that can reveal genetic diversity and phenotypic diversity in the pathophysiology of the disease.

Keywords: *Phenome-wide association study (PheWAS), Genome-wide association study (GWAS), Electronic health records (EHRs), Large-scale genotype-phenotype correlation, COVID-19.*

DOI: 10.48141/SJS.v30.n34.2022.04_ATCEKEN_pgs_32_46.pdf

1. INTRODUCTION

The new coronavirus disease (COVID-19) is caused by the severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2) (WHO, 2022). According to the World Health Organization (WHO), data published as of 25 April 2022 shows 507,501,771 individuals have been affected by COVID-19, of which 6,220,390 individuals have died as a result of the disease (Dashboard, 2022). It has caused a global pandemic and constitutes a public health problem (Mahase, 2020). COVID-19 disease can be seen as symptomatic and asymptomatic. The course and severity of the disease are different from person to person (Guan *et al*., 2020; Kotsev *et al*., 2021).

An important question is the role of host genes in COVID-19 infection. To answer this question, we need to understand the effects of host genes and genetic variations on the different phenotypes of COVID-19. Large-scale genotype-phenotype correlation studies have been demonstrated to do this in several different studies *(Gaziano et al., 2021; Moon et al*., 2021; Verma *et al*., 2021). These studies are most beneficial in the presence of clinical heterogeneity. Clinical heterogeneity investigation is needed when different disease phenotypes are compared between case and control groups. Therefore, the differences in endotype, endophenotype, and symptomatic severity of a disease can be understood. Genomic medicine is committed to elucidating the genetic diversity underlying the phenotypic difference in disease (Manolio *et al*., 2013). Genomic medicine differs from traditional genetics in that it regards the functions-interactions of all variations and genes in the genome (Guttmacher *et al*., 2002). This field deals with the hereditary components of monogenic, polygenic, and infectious diseases. Doing this provides an understanding of the molecular basis of all diseases and enables the development of targeted therapy and personalized treatment strategies. Genomic medicine combines multidisciplinary fields and focuses on diseases by matching genetic characteristics with phenotypic data (Manolio *et al*., 2013; Wei *et al*., 2017). The use of genomic medicine strategies has become essential in COVID-19 research. The basis of these strategies involves genome-wide association study (GWAS) data and Electronic health records (EHRs) containing phenotypic data (Linder *et al*., 2021). EHRs include all individual disease, clinical, and treatment information rather than cohorts specific to a single disease (Linder *et al*., 2021). At the same time, since it is comprised of data from large populations, there is a wealth of data covering a variety of population characteristics (Linder *et al*., 2021).

Phenome-wide association studies (PheWASs) have been used in the evaluation of GWAS results (Denny *et al*., 2010). PheWAS determines the full phenotypic spectrum associated with each genetic trait (Hebbring, 2014). PheWAS was initially developed to reveal the large-scale genotype-phenotype correlations of complex diseases. However, it can be used to evaluate the risk score of genetic variations with disease using EHRs *(Denny et al*., 2016). PheWAS has emerged as a popular high-throughput framework mechanism capable of combining EHR data with GWAS data (Linder *et al*., 2021). In addition, data obtained in PheWAS can be matched with tissue-specific expression Quantitative Trait Locus (eQTL) data to help understand the biological mechanism underlying the genotype-phenotype relationship (Moon *et al*., 2021).

This review describes the PheWAS trend from GWAS, and the methodology of PheWAS. Then, genotype and phenotype data resources for COVID-19 research are mentioned. Next, the importance of PheWAS in cross-genotypic-phenotypic correlation and large-scale genotype-phenotype correlation is discussed. Finally, PheWAS studies on COVID-19 patients using GWAS data and EHR-based data were mentioned. This review aims to give an overview of the current state of COVID-19 research, focusing on genotype-phenotype-related studies. PheWAS is an ideal method for large-scale genotype-phenotype correlation studies. As a result, the genetic diversity in pathophysiology and phenotypic variety of the disease can be revealed.

2. METHODS

The method explains the genetic and phenotypic data used in PheWAS analyses. The methodology of GWAS analysis, which is used as a genetic data source in PheWAS, and its integration into PheWAS analysis are detailed. EHRs, which are also included as the phenotypic data set in the PheWAS analysis, were also examined. It is mentioned how the PheWAS analysis is applied to these data. Finally, the phenotypic diversity in covid-19 patients is mentioned.

PubMed / National Center for Biotechnology Information (https://pubmed.ncbi.nlm.nih.gov/) was used as the source of information about the PheWAS

analyzes applied in our review. PheWAS research articles on COVID-19 disease were selected, and their results were examined. PheWAS research articles and review articles applied to other diseases were excluded.

In addition, 2 terms (A / B) were considered in selected research articles; A) PheWAS analysis articles containing genotypic and phenotypic information B) articles matching phewas analysis with gene expression data.

The research was conducted from February 01 to July 01, 2022. The period searched on the databases was from 2007 to 2022 (includes the date PheWAS analyzes were first performed and the date to date).

3. RESULTS AND DISCUSSION

3.1. Results

To date, all publications related to PheWAS have been scanned in PubMed (Figure 1). These publications selected articles to establish a large-scale genotype-phenotype relationship in COVID-19 patients. Articles were determined according to A and B terms.

Table 1. *Data from PheWAS publications scanned in PubMed and articles included and excluded from our systematic review*

Term	Database	Results	Exclusions
“A”	PubMed	3	777
“B”	PubMed	3	777

Few studies of PheWAS have been conducted on COVID-19 diseases. However, quite comprehensive and important data have been obtained. Increasing the number of these studies is very important. These studies appear to provide full phenotypic data for each GWAS significant variation or genetic locus (Table 2). We have systematically examined and discussed the COVID-19 phenotypes and comorbidities with which these genetic traits are highly correlated.

3.1.1. The GWAS era and the trend toward PheWAS

Investigating the effects of variations that underlie human genetic diversity has been the focus of attention since the completion of the human genome project. Advancing molecular genetic techniques allow for pleiotropic effect studies in disease research. Pleiotropy is when any variation in the genome affects multiple phenotypes (Tyler *et al*., 2016). Pleiotropy investigates the causes of phenotypic differences in diseases from individual to individual and forms the basis of knowledge for personalized medicine applications (Sivakumaran *et al*., 2011). GWAS has been used for large-scale genotypic-phenotypic data sets to illuminate human pleiotropy *(Hindorff et al., 2009; Sivakumaran et al*., 2011). GWAS investigates the pleiotropic effect at the level of variation between a large number of cases and controls. GWAS simultaneously analysis millions of variations across the whole genome (Consortium, 2007; Hindorff *et al*., 2009). GWASs analyze single nucleotide polymorphisms (SNPs) and can display effects as minor as $P<5\times10^{-7}$ (Consortium, 2007). In GWAS, both chip-based microarray and next-generation sequencing techniques are used for analysis. GWASs usually focus on analyzing variations found in the intergenic, intronic, and exonic regions of the human genome (Li *et al*., 2008). Significant variants found are not always causal; however, linkage disequilibrium (LD) can assist in identifying closely correlated variants (Anderson *et al*., 2011).

Population cohorts were defined through the genetic diversity of different ethnic groups (Benjamin *et al*., 2007). Consortium studies were conducted by combining these population cohorts (Consortium, 2007). These consortium studies have contributed to revealing even minor effect variations and genetic differences in populations (Consortium, 2007). The genetic effect size is defined by looking at the allelic frequency of the variations in cases and controls (Bush *et al*., 2012). According to this genetic effect size, the disease-related genetic risk score of the variation is identified (Consortium, 2007; Hindorff *et al*., 2009).

Furthermore, it can explain the importance of ethnic origin, genetic differences, and genetic predispositions in the pathophysiology of the disease. GWAS has also successfully identified genetic risk factors involved in the epidemiology, development, severity, clinical differences, and response to treatment of the disease (Michailidou *et al*., 2015). Thus, a large number of new genetic traits associated with diseases have been detected (Hindorff *et al*., 2009).

GWAS offers researchers a unique opportunity to demonstrate the effect of variations on disease phenotypes. With the rapid increase in GWAS studies, and the data size increasing from Array-based technology to NGS, there was a need to establish a biobank for these samples and for the ease of access for researchers. Existing GWAS data were brought together by the US

National Human Genome Research Institute (NHGRI) in 2008, and the GWAS catalog was created. In 2010, the GWAS catalog website was established in collaboration with the European Bioinformatics Institute (EMBL-EBI) (https://www.ebi.ac.uk/gwas/). This catalog of associations is increasing every year. The catalog details all current SNP-trait relationships to date from common to rare diseases. The data collected in the catalog is combined with other sources to allow statistical project-centric modeling (Buniello *et al*., 2019).

The GWAS era has accelerated human pleiotropy research and led to cross-phenotype associations studies (Tyler *et al*., 2016). The cross-phenotype association approach has attracted great attention in the scientific world (Denny *et al*., 2010). The first PheWAS was performed in 2010 and was utilized to understand genetic pleiotropy in humans (Denny *et al*., 2010). PheWAS can be considered as an inverse method to GWAS. In GWAS studies, the relationship of many genetic traits with a specific phenotype is investigated. However, in PheWAS analyzes, the relationship of a single genetic trait with many clinical phenotypes is studied (Hebbring, 2014). Therefore, PheWAS analysis is complementary to GWAS in disease research (Hebbring, 2014).

PheWAS enables the simultaneous identification of associations between a genetic trait and phenotypic traits, clinical manifestations, and many diseases. At the same time, genetic predispositions underlying disease comorbidities may also emerge (Karaca *et al*., 2020). Furthermore, it can reveal new genotype-phenotypic correlations as thousands of phenotypes can be compared with significant variations of GWAS (Cronin *et al*., 2014; Denny *et al*., 2010; Karaca *et al*., 2020). For example, one study applied PheWAS to FTO gene variants previously reported to be associated with type 2 diabetes and obesity. They used the eMERGE Network (Gottesman *et al*., 2013) and BioVU DNA biobank data (Roden *et al*., 2008). The study found that the FTO gene variant, associated with body mass index (BMI), is also associated with sleep apnea. Furthermore, the variant associated with obesity, non-alcoholic liver disease, fibrocystic breast disease, and gram-positive bacterial association were defined (Cronin *et al*., 2014).

The web tools GRASP (Leslie *et al*., 2014), GeneATLAS (Canela-Xandri *et al*., 2018), and PhenoScanner (Staley *et al*., 2016) are used to perform PheWAS analysis. These websites contain thousands of GWAS study information with millions of SNP-trait information. It allows users to query the full phenotypic spectrum of each SNP with PheWAS analysis.

The genetic component of PheWAS is not restricted by GWAS significant SNPs. Rare variations (MAF < 0.05), mitochondrial variations, copy number variation (CNV), and structural variation (SV) data can also be analyzed (Basile *et al*., 2016; Mitchell *et al*., 2014). On the other hand, apart from genetic data, clinical analysis results, biochemical parameters, environmental measures, and quantitative values can be used in biomarker studies for disease (Liao *et al*., 2017).

3.1.2. Phenome information collection in PheWAS

PheWAS can be used for matching metadata collected from longitudinal studies *(Denny et al*., 2010; Denny *et al*., 2016). Longitudinal studies identify risk factors for a particular disease over multiple time points. It is a research strategy that includes reproducible observations of the same variables for short or long periods (Shadish, 2002). This strategy also categorizes the personal characteristics of the records that exist retrospectively over time or new data to be collected prospectively. The EHR, which contains demographic and clinical characteristics, is longitudinal in nature (Denny *et al*., 2016). For PheWAS, EHRs are utilized as a source of phenome information (Verma *et al*., 2021; Zhou *et al*., 2021).

3.1.3. EHR-based PheWAS

Electronic health records (EHR) with more than 50 years of history have gained popularity in recent years (Gottesman *et al*., 2013; McDonald *et al*., 1977). Many countries have started to organize their data by establishing national health record systems (Linder *et al*., 2021). EHRs have rapid and automated clinical data collection from when patients are recruited. It contains the individual and social characteristics of diseases. EHRs describe the prevalence, course, and outcomes of diseases at the national level while also providing an opportunity to compare and match across international EHRs (Linder *et al*., 2021). EHR data provides the opportunity for electronic phenotyping (e-phenotyping). It represents a more comprehensive e-phenotyping information collection as they automate clinical data collection (Linder *et al*., 2021). The e-phenotyping information collection produces computational large-scale phenotypic big data in terms of disease monitoring, prevention, and development of preventive health strategies and treatment

strategies. As EHR data has accumulated over the years, researchers have highly developed and structured the content of data types. These data types include all observable characteristics of an individual, such as age, gender, BMI, past-existing diseases (hereditary disease, chronic complex disease, infectious disease), drugs used, and allergic conditions (Casey *et al*., 2016; Denny *et al*., 2013; Linder *et al*., 2021).

A procedural medical code system is also implemented within EHRs. This code system called the International Statistical Classification of Diseases, and Related Health Problems (ICD), is a globally accepted system in which medical diagnoses are standardized (Krawczyk *et al*., 2020). The ICD compares and contrasts disease statistics on a global scale. It also procures convenience in identifying the prevalence of diseases, treatment strategies, and taking preventive measures at the international level (Harrison *et al*., 2021). In addition, disease-based phenotypic data algorithms have been developed, including ICD data in the EHR such as PheWAS.

The emergence of EHR and the creation of EHR-linked biobanks allow large-scale genotype-phenotype correlations to be established (Salvatore *et al*., 2021). There are many EHR-linked biobanks, and these biobanks contain multi-omics data as well as genomic data. For example, the UK Biobank (Allen *et al*., 2014), Chinese Kadoorie Biobank (Chen *et al*., 2011), Vanderbilt BioVU (Roden *et al*., 2008), Electronic Medical Records and Genomic Network (eMERGE) (Gottesman *et al*., 2013), which contains data on more than 200 000 individuals worldwide, are among the biobank data sources that can be analyzed in PheWAS. The advancement of machine learning algorithms and the targeting of big data analysis led to the design and development of PheWAS (Gagliano Taliun *et al*., 2020; Salvatore *et al*., 2021). This advancement enables comprehensive correlation analysis based on multivariate regression analysis by integrating multiple datasets (Bush *et al*., 2012). Since PheWAS is a complex big data analysis, it can illuminate the hidden and unknown gene-phenotype relationships associated with any disease (Gagliano *et al*., 2020; Karaca *et al*., 2020; Salvatore *et al*., 2021). It can also scan for the reflection of genetic predispositions leading to population stratification and population-specific disease phenotypes (Bush *et al*., 2012). In addition, combining and comparing different populations data can expose the effect of ancestral-ethnic origin differences on disease phenotypes (Verma *et al*., 2021).

3.1.4. Mapping eQTL information to PheWAS

When PheWAS with phenotype data is applied to GWAS significant variations, the full phenotypic spectrum associated with the SNP is identified. Tissue-specific expression quantitative trait locus (eQTL) data and PheWAS data colocalization analysis can be applied to elucidate the underlying pathophysiological condition in the reflection of genotype to phenotype (Moon *et al*., 2021). eQTL information is available on the Genotype-Tissue Expression (GTEx) portal and is open to researchers (www.gtexportal.org). The GTEx v8 dataset includes whole-genome sequencing (WGS) and RNA-sequencing (RNA-seq) information of 17,382 samples from 838 donors. Gene expression data of cis and trans variations of 52 tissues and two cell lines are also included (www.gtexportal.org). The cis-eQTL value gives the change in expression level relative to the transcriptional start size of genes located close to the LD of the variations (G. Consortium, 2020). In addition, cis-eQTL values can be calculated in multi-tissue and single-tissue (G. Consortium, 2020). In disease studies, the cis-eQTL value in the tissue associated with the disease allows for the interpretation of the variation affect on pathogenesis (G. Consortium, 2020).

3.1.5. Phenotypic diversity in COVID-19 patients

Coronavirus Disease-2019 (COVID-19) is observed in different individuals as asymptomatic and in various symptomatic forms. This diversity can give rise to diagnosis, outpatient treatment, hospitalization, intensive care unit admission, multiple organ failure, and death (Guan *et al*., 2020). The symptomatic variability in COVID-19 is connected with the level of inflammatory response triggered by immune system activation. A systematic immune reaction occurs due to the effector cells involved in the immune response, the release of mediators that mediate inflammation, and their complex interactions. This immune reaction sometimes causes immune hyperactivation or dysregulation, resulting in an abnormal cytokine storm. This uncontrolled immune response results in the development of acute respiratory distress syndrome (ARDS), increased disease severity, multiple organ failure, and even death in COVID-19 infection (Kotsev *et al*., 2021). The basis of this uncontrolled immune response and infection susceptibility, which varies from person to person, is very likely to be based on host genetic diversity.

The causes and risk factors influencing the severity of disease symptoms (mild, moderate, and severe complications) are uncertain. Many studies have been conducted on the severity and mortality rate of COVID-19 disease. In these studies, it has been determined that advanced age, male gender, socioeconomic level, type-2 diabetes, cardiovascular diseases, hypertension, kidney diseases, cancer, obesity, and asthma are connected with severe complications in COVID-19 (Fang *et al*., 2020; Williamson *et al*., 2020). However, these findings are analyzed based on observational and numerical data from limited, regional, hospital-based studies (Williamson *et al*., 2020). It does not reflect the influence of human genotypic structure, genetic predisposition, cross-genotypic-phenotypic correlation, and the importance of population diversity. COVID-19 is not well understood and how host genetic factors contribute to the pathogenesis of disease severity difference and its interaction with its comorbidities (Williamson *et al*., 2020).

3.2. Discussions

Identifying the risk factors that cause COVID-19 severity and symptomatic variation may provide clinical and therapeutic advantages and contribute to developing protective-preventive strategies. Various host genetic traits may likely be risk factors influencing viral susceptibility, immune response, disease progression, and outcomes (Choudhary *et al*., 2021; Debnath *et al*., 2020). GWAS offers the opportunity to identify potential candidate genes associated with severity, development, and symptomatic differences of COVID-19 infection. Since COVID-19 first appeared, numerous GWAS have been performed to identify potential candidate host genetic traits. Variations of *SLC6A20, LZFTL1, CCR9, CXCR6, XCR1, FYCO1* (3p21.31) (Group, 2020), *ABO* (9q34.2) (Wu *et al*., 2020), *HLA* (6p21.33) (Novelli *et al*., 2020), *TMEM189-UBE2V1* (20q13.13) (Wang *et al*., 2020), *ACE2* (Xp22.2) (Hou *et al*., 2020), *TMPRSS2* (21q22.3) (Anastassopoulou *et al*., 2020; Hou *et al*., 2020), *TLR7* (Xp22.2) (Anastassopoulou *et al*., 2020), *ApoE* (19q13.32) (Kuo *et al*., 2020), *IFITM3* (11p15.5) (Thevarajan *et al*., 2020; Zhang *et al*., 2020), *CTSB, CTSL* (8p23.1, 9q21.33) (Lee *et al*., 2020; Yang *et al*., 2021), *PIEZO* (16q24.3) (Cheng *et al*., 2020), *OAS1, OAS2, OAS3* (12q24.13) (Pairo-Castineira *et al*., 2021), *TYK2* (19p13.2) (Pairo-Castineira *et al*., 2021), *DPP9* (19p13.3) (Pairo-Castineira *et al*., 2021), *IFNAR2* (21q22.1) (Pairo-Castineira *et al*., 2021) genes were found to be associated with COVID-19 infection. So far, a large amount of SNP-trait information has been stored in the GWAS catalog (https://www.ebi.ac.uk/gwas/) and at the COVID-19 Host Genetics Initiative (https://www.covid19hg.org/) that can be used in COVID-19 research. However, GWAS has some limitations; i) It can only associate variants with a single phenotype and does not reflect the full phenotypic spectrum *(Kotsev et al*., 2021), and ii) It is insufficient to determine the comorbidities of COVID-19 (Kotsev *et al*., 2021). PheWAS analysis overcomes these limitations by utilizing GWAS data with EHR Data.

Many PheWAS studies have been conducted on COVID-19, matching GWAS and EHR data (Crespi, 2020; Gaziano *et al*., 2021; Lopera *et al*., 2020; Moon *et al*., 2021; Verma *et al*., 2021; Zhou *et al*., 2021) (Table 1). Variations related to COVID-19 severity and severe clinical symptoms in a PheWAS study were associated with many different phenotypes (Verma *et al*., 2021). In the related study: *ABO* locus rs495828 variant was associated with 53 different phenotypes, the most significant of which was detected with venous embolism (Verma *et al*., 2021). The *ABO* locus rs505922 variant was linked with 59 phenotypes and strongly associated with thrombosis count (Verma *et al*., 2021). It has been demonstrated that the *MUC5B* locus rs35705950 variant increases the risk of idiopathic fibrosing alveolitis and is associated with 11 different respiratory features (Verma *et al*., 2021). There was a negative correlation between the risk of CRHR1 gene rs61667602 variant pulmonary fibrosis and *TYK2* locus rs11085727 variant autoimmune conditions (Verma *et al*., 2021). In another study, it was determined that the rs13050728 variant is a risk factor for COVID-19 hospitalization. Through the eQTL analysis, the expression levels of *TPSG1* and *VEGFR2* genes in plasma were lower than in other tissues (Gaziano *et al*., 2021). In addition, the rs4830976 variant was related to COVID-19 hospitalization and caused changes in the expressions of the *ACE2, CA5B, CLTRN, and VEGFD* genes related closely through LD in the eQTL coagulation analysis (Gaziano *et al*., 2021).

PheWAS can be of benefit to elucidating the association of a single disease-related significant locus with intermediate phenotypes (Zhou *et al*., 2021). It is a strategy that can be especially effective for drug and biomarker research (Crespi, 2020). The role of the relevant locus in the pathophysiology of the disease can be revealed using simultaneous eQTL information with PheWASs. For example, to explain the relationship between the 3p21.31 genetic locus

and COVID-19 severity, the expression level changes of the genes in this locus caused by the rs67959919 variation were examined (Zhou *et al*., 2021). This PheWAS and eQTL analysis observed that the rs67959919 variation caused changes in the monocyte count by increasing *CCR1* gene expression and decreasing *CCR2* gene expression. Furthermore, it has been revealed that the same variation plays a role in the level changes in eosinophil and neutrophil counts by decreasing the gene expression of the *CCR3* gene (Zhou *et al*., 2021). In another study, seven different GWAS significant variations (rs657152, rs11385942, rs150892504, rs138763430, rs117665206, rs147149459, and rs151256885) with COVID-19 mortality rates were elucidated concerning drugs (Amlodipine and aspirin) and other clinical phenotypes for medical drug targeting (Crespi, 2020). A locus-targeted PheWAS analysis focused on angiotensin-converting enzyme 2 (*ACE2*) and serine protease *TMPRSS2* genes, which are known to be involved in the virus infecting human cells in COVID-19 infection (Lopera Maya *et al*., 2020). During infection, the ACE2 receptor protein is responsible for cell invasion, while the TMPRSS2 protein takes part in preparing the S protein (Yan *et al*., 2020). Genotype-phenotype correlation analysis was performed between 1273 genetic variations (*ACE2* and *TMPRSS2* genes were in and near regions localized) and 178 quantitative phenotypes in the related study (Lopera Maya *et al*., 2020). In the *ACE2* gene, the variant rs17264937 was highly correlated with Eosinophils, and the rs5980163 variant with triglycerides. rs150965978 is associated with plasma levels of CHIT1 protein, while the variant rs28401567 has been reported to be significantly associated with thrombocytes in the *TMPRSS2* gene (Lopera *et al*., 2020).

According to clinical data, individuals with severe clinical diagnosis in COVID-19 also have a pre-existing disease (Fang *et al*., 2020). The ability of PheWAS to define the cross genotype-phenotype correlation is important in the identification of disease comorbidities and in stating the underlying genetic traits. One study focused on determining the comorbid disease-related phenotypes of 22 variations associated with severe COVID-19 respiratory failure and applied PheWAS and eQTL colocalization analysis to GWAS (Moon *et al*., 2021). Five variations (rs647800, rs11385492, rs12610495, rs3934992, rs134130) were significantly associated with 13 different endocrine, metabolic and immunological phenotypes (Moon *et al*., 2021). Variations of rs647800 and rs11385492 have been reported to be risk factors for monocyte-induced inflammation related to the number of monocytes and the percentage of monocytes found in the blood. The rs647800 variation was associated with thrombin time, and they suggested that it may be effective in developing coagulopathies (comorbidity of severe COVID-19). They revealed that the rs12610495 variation is also associated with fibrotic idiopathic interstitial pneumonias (comorbidity of severe COVID-19). The rs3934992 variation was found to be related to the waist-hip ratio (adjusted for BMI), and it was thought to be a risk factor in obesity (comorbidity of severe COVID-19) (Moon *et al*., 2021).

Through PheWAS analysis, changes in risk factors were also observed according to COVID-19 severity and ethnicity differences. In addition, a correlation was observed between the *LMNA* gene rs581342 variation and neutropenia, *HL-DRA* gene rs9268576 variation, and thyrotoxicosis in parallel with the severity of COVID-19 in Africans (Verma *et al*., 2021).

In some COVID-19 studies, PheWAS was applied only to EHR data (Oetjens *et al*., 2020; Salvatore *et al*., 2021; Song *et al*., 2021). In these studies, COVID-19 positive diagnosis, severity, hospitalization, and mortality were matched with the ICD code information of the individuals. Thousands of phenotypic codes (demographic features, clinical findings, biochemical parameters, age, gender, and all genetic and chronic disease information of individuals) were used in these cohorts consisting of thousands of individuals (Oetjens *et al*., 2020; Salvatore *et al*., 2021; Song *et al*., 2021). In addition, some studies focus on ethnic differences (Salvatore *et al*., 2021). For example, in non-Hispanic Whites, Hematopoietic conditions were associated with ICU admission/death, and mental disorders were associated with death (Salvatore *et al*., 2021). Also, in non-Hispanic Blacks, Circulatory system and genitourinary conditions were associated with ICU admission/death (Salvatore *et al*., 2021). As a result, large-scale phenotypic correlations were detected. However, these studies are based on numerical and observational data. Since they were not matched with genotypic data, the genetic predisposition underlying the detected correlations could not be fully understood.

GWAS provides an excellent source of genotypic data. It allows the comparison of sick-healthy individuals and enables the determination of genetic predispositions according to population differences. However, it is not sufficient on its own as it reflects the genetic spectrum associated with

a single phenotype. Therefore, there is a need for bioinformatics methods to match phenotypic data with GWAS data. COVID-19 has different symptomatic forms and variations in disease severity from person to person. Clarifying the genetic predispositions underlying these phenotypic diversity is important for elucidating the pathogenesis of the disease and for a good prognosis. As a source of phenotypic data, EHRs provide large-scale data of enormous quality. However, phenotypic data alone are not sufficient to elucidate the disease pathogenesis and prognosis. PheWAS analysis has been developed as a bioinformatics method that can be applied to both GWAS data and EHR data. It can also illuminate the full phenotypic spectrum by matching both data. On the other hand, by matching eQTL data with PheWAS data, it contributes to the elucidation of the molecular mechanisms involved in the pathogenesis. As a result of processing these data together with PheWAS analysis, it provides the opportunity to understand the reasons for the phenotypic diversity of COVID-19.

4. CONCLUSIONS

Performing large-scale genotype-phenotype correlation studies are important to determine an individuals risk for the prevention and evaluation of disease severity and symptomatic variations. In addition, conducting studies on different populations and comparing them on a global scale would also be valuable in elucidating genetic predispositions arising from ethnic differences. At this point, PheWASs emerge as an important approach for profiling large-scale genotype-phenotype correlations. The data already available will prepare the foundation for identifying high-risk COVID-19 individuals, developing protective-preventive treatment strategies, and personalized medicine over the standardized-for-everyone approach. This review provides an overview of the PheWAS methodology and its application to COVID-19 studies. We mention the current methodology, data and analytical resources, and COVID-19-associated genetic variant summaries for future PheWASs. To date, many studies have been carried out globally, and a wealth of data is available from GWASs and EHRs. PheWAS is an ideal approach for large-scale genotype-phenotype correlation studies. The quality, number, and impact of such studies will increase in the near future and gain the importance it deserves. Thus, the role of host genetic predispositions and genetic diversity in phenotypic differences of Covid-19 can be revealed.

5. DECLARATIONS

5.1. Study Limitations

The study is limited to the consulted bibliography and the period when the research was conducted (from 2007 to 2022). However, studies on COVID-19 are limited to the years 2019-2022.

5.2. Acknowledgements

We thank Dr. Burak Kadir DURAK for his contribution to the emergence of the review.

5.3. Funding source

The authors funded this research.

5.4. Competing Interests

The authors declare no conflict of interest.

6. REFERENCES:

1. Allen, N. E., Sudlow, C., Peakman, T., & Collins, R. (2014, February 19). UK Biobank Data: Come and Get It. Science Translational Medicine. American Association for the Advancement of Science (AAAS). Retrieved from

http://dx.doi.org/10.1126/scitranslmed.3008601
2. Anastassopoulou, C., Gkizarioti, Z., Patrinos, G. P., & Tsakris, A. (2020, October 22). Human genetic factors associated with susceptibility to SARS-CoV-2 infection and COVID-19 disease severity. Human Genomics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1186/s40246-020-00290-4
3. Anderson, C. A., Soranzo, N., Zeggini, E., & Barrett, J. C. (2011, January 18). Synthetic Associations Are Unlikely to Account for Many Common Disease Genome-Wide Association Signals. (M. L. Freedman, Ed.)PLoS Biology. Public Library of Science (PLoS). Retrieved from http://dx.doi.org/10.1371/journal.pbio.1000580
4. Basile, A. O., Wallace, J. R., Peissig, P., Mccarty, C. A., Brilliant, M., & Ritchie, M. D. (2015, November 18). Knowledge Driven Binning And Phewas Analysis In Marshfield Personalized Medicine Research Project Using Biobin. Biocomputing 2016. Presented at the Proceedings of the Pacific Symposium, World Scientific. Retrieved from http://dx.doi.org/10.1142/9789814749411_0024.
5. Benjamin, E. J., Dupuis, J., Larson, M. G., Lunetta, K. L., Booth, S. L., Govindaraju, D. R., Kathiresan, S., et al. (2007, September). Genome-wide association with select biomarker traits in the Framingham Heart Study. BMC Medical Genetics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1186/1471-2350-8-s1-s11
6. Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., et al. (2018, November 16). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/nar/gky1120
7. Bush, W. S., & Moore, J. H. (2012, December 27). Chapter 11: Genome-Wide Association Studies. (F. Lewitter & M. Kann, Eds.)PLoS Computational Biology. Public Library of Science (PLoS). Retrieved from http://dx.doi.org/10.1371/journal.pcbi.1002822
8. Canela-Xandri, O., Rawlik, K., & Tenesa, A. (2018, October 22). An atlas of genetic associations in UK Biobank. Nature Genetics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41588-018-0248-z
9. Casey, J. A., Schwartz, B. S., Stewart, W. F., & Adler, N. E. (2016, March 18). Using Electronic Health Records for Population Health Research: A Review of Methods and Applications. Annual Review of Public Health. Annual Reviews. Retrieved from http://dx.doi.org/10.1146/annurev-publhealth-032315-021353
10. Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F., Li, L., et al. (2011, September 21). China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. International Journal of Epidemiology. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/ije/dyr120
11. Cheng, C. W., Deivasikamani, V., Ludlow, M. J., De Vecchis, D., Kalli, A. C., Beech, D. J., & Sukumar, P. (2020, June 3). Genetic variants of PIEZO1 associate with COVID-19 fatality. Cold Spring Harbor Laboratory. Retrieved from http://dx.doi.org/10.1101/2020.06.01.20119651
12. Choudhary, S., Sreenivasulu, K., Mitra, P., Misra, S., & Sharma, P. (2021, March 1). Role of Genetic Variants and Gene Expression in the Susceptibility and Severity of COVID-19. Annals of Laboratory Medicine. Annals of Laboratory Medicine. Retrieved from http://dx.doi.org/10.3343/alm.2021.41.2.129
13. Consortium, (2007, June). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/nature05911
14. Aguet, F., Anand, S., Ardlie, K. G., Gabriel, S., Getz, G. A., Graubert, A., et al. (2020, September 11). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. American Association for the Advancement of

Science (AAAS). Retrieved from http://dx.doi.org/10.1126/science.aaz1776
15. Crespi, B. (2020, January 1). Evolutionary medical insights into the SARS-CoV-2 pandemic. Evolution, Medicine, and Public Health. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/emph/eoaa036
16. Cronin, R. M., Field, J. R., Bradford, Y., Shaffer, C. M., Carroll, R. J., Mosley, J. D., Bastarache, L., et al. (2014, August 5). Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. Frontiers in Genetics. Frontiers Media SA. Retrieved from http://dx.doi.org/10.3389/fgene.2014.00250
17. Dashboard, W., (2022). *Retrieved from* https://covid19.who.int/
18. Debnath, M., Banerjee, M., & Berk, M. (2020, June 11). Genetic gateways to COVID-19 infection: Implications for risk, severity, and outcomes. The FASEB Journal. Wiley. Retrieved from http://dx.doi.org/10.1096/fj.202001115R
19. Denny, J. C., Bastarache, L., Ritchie, M. D., Carroll, R. J., Zink, R., Mosley, J. D., Field, J. R., et al. (2013, December). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nature Biotechnology. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/nbt.2749
20. Denny, J. C., Bastarache, L., & Roden, D. M. (2016, August 31). Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. Annual Review of Genomics and Human Genetics. Annual Reviews. Retrieved from http://dx.doi.org/10.1146/annurev-genom-090314-024956
21. Denny, J. C., Ritchie, M. D., Basford, M. A., Pulley, J. M., Bastarache, L., Brown-Gentry, K., Wang, D., et al. (2010, March 24). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/bioinformatics/btq126
22. Fang, X., Li, S., Yu, H., Wang, P., Zhang, Y., Chen, Z., Li, Y., et al. (2020, July 13). Epidemiological, comorbidity factors with severity and prognosis of COVID-19: a systematic review and meta-analysis. Aging. Impact Journals, LLC. Retrieved from http://dx.doi.org/10.18632/aging.103579
23. Gagliano Taliun, S. A., VandeHaar, P., Boughton, A. P., Welch, R. P., Taliun, D., Schmidt, E. M., Zhou, W., et al. (2020, June). Exploring and visualizing large-scale genetic associations by using PheWeb. Nature Genetics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41588-020-0622-5
24. Gaziano, L., Giambartolomei, C., Pereira, A. C., Gaulton, A., Posner, D. C., Swanson, S. A., Ho, Y.-L., et al. (2021, April). Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. Nature Medicine. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41591-021-01310-z
25. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W. A., Li, R., Manolio, T. A., Sanderson, S. C., et al. (2013, October). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genetics in Medicine. Elsevier BV. Retrieved from http://dx.doi.org/10.1038/gim.2013.72
26. (2020, October 15). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. New England Journal of Medicine. Massachusetts Medical Society. Retrieved from http://dx.doi.org/10.1056/NEJMoa2020283
27. Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., et al. (2020, April 30). Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine. Massachusetts Medical Society. Retrieved from http://dx.doi.org/10.1056/NEJMoa2002032
28. Guttmacher, A. E., & Collins, F. S. (2002, November 7). Genomic Medicine — A Primer. (A. E. Guttmacher & F. S. Collins, Eds.)New England Journal of Medicine. Massachusetts Medical Society. Retrieved from http://dx.doi.org/10.1056/NEJMra012240

29. Harrison, J. E., Weber, S., Jakob, R., & Chute, C. G. (2021, November). ICD-11: an international classification of diseases for the twenty-first century. BMC Medical Informatics and Decision Making. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1186/s12911-021-01534-6
30. Hebbring, S. J. (2014, January 9). The challenges, advantages and future of phenome-wide association studies. Immunology. Wiley. Retrieved from http://dx.doi.org/10.1111/imm.12195
31. Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009, June 9). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences. Proceedings of the National Academy of Sciences. Retrieved from http://dx.doi.org/10.1073/pnas.0903103106
32. Hou, Y., Zhao, J., Martin, W., Kallianpur, A., Chung, M. K., Jehi, L., Sharifi, N., et al. (2020, July 15). New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. BMC Medicine. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1186/s12916-020-01673-z
33. Karaca, M., Atceken, N., Karaca, Ş., Civelek, E., Şekerel, B. E., & Polimanti, R. (2020). Phenotypic and Molecular Characterization of Risk Loci Associated With Asthma and Lung Function. Allergy, Asthma & Immunology Research. The Korean Academy of Asthma, Allergy and Clinical Immunology and The Korean Academy of Pediatric Al. Retrieved from http://dx.doi.org/10.4168/aair.2020.12.5.806
34. Kotsev, S. V., Miteva, D., Krayselska, S., Shopova, M., Pishmisheva-Peleva, M., Stanilova, S. A., & Velikova, T. (2021, July 25). Hypotheses and facts for genetic factors related to severe COVID-19. World Journal of Virology. Baishideng Publishing Group Inc. Retrieved from http://dx.doi.org/10.5501/wjv.v10.i4.137
35. Krawczyk, P., & Święcicki, Ł. (2020, February 29). ICD-11 vs. ICD-10 – a review of updates and novelties introduced in the latest version of the WHO International Classification of Diseases. Psychiatria Polska. Komitet Redakcyjno - Wydawniczy Polskiego Towarzystwa Psychiatrycznego. Retrieved from http://dx.doi.org/10.12740/PP/103876
36. Kuo, C.-L., Pilling, L. C., Atkins, J. L., Kuchel, G. A., & Melzer, D. (2020, June 8). ApoE e2 and aging-related outcomes in 379,000 UK Biobank participants. Aging. Impact Journals, LLC. Retrieved from http://dx.doi.org/10.18632/aging.103405
37. Kuo, C.-L., Pilling, L. C., Atkins, J. L., Masoli, J. A. H., Delgado, J., Kuchel, G. A., & Melzer, D. (2020, May 26). APOE e4 Genotype Predicts Severe COVID-19 in the UK Biobank Community Cohort. The Journals of Gerontology: Series A. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/gerona/glaa131
38. Lee, I.-H., Lee, J.-W., & Kong, S. W. (2020, November). A survey of genetic variants in SARS-CoV-2 interacting domains of ACE2, TMPRSS2 and TLR3/7/8 across populations. Infection, Genetics and Evolution. Elsevier BV. Retrieved from http://dx.doi.org/10.1016/j.meegid.2020.104507
39. Leslie, R., O'Donnell, C. J., & Johnson, A. D. (2014, June 15). GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/bioinformatics/btu273
40. Li, M., Li, C., & Guan, W. (2008, February 6). Evaluation of coverage variation of SNP chips for genome-wide association studies. European Journal of Human Genetics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/sj.ejhg.5202007
41. Liao, K. P., Sparks, J. A., Hejblum, B. P., Kuo, I., Cui, J., Lahey, L. J., Cagan, A., et al. (2017, March 29). Phenome-Wide Association Study of Autoantibodies to Citrullinated and Noncitrullinated Epitopes in Rheumatoid Arthritis. Arthritis & Rheumatology. Wiley. Retrieved from http://dx.doi.org/10.1002/art.39974
42. Linder, J. E., Bastarache, L., Hughey, J. J., & Peterson, J. F. (2021, August 31). The Role of Electronic Health Records in Advancing Genomic Medicine. Annual Review of Genomics and Human Genetics. Annual Reviews. Retrieved from

http://dx.doi.org/10.1146/annurev-genom-121120-125204
43. Lopera Maya, E. A., van der Graaf, A., Lanting, P., van der Geest, M., Fu, J., Swertz, M., Franke, L., et al. (2020, June 8). Lack of Association Between Genetic Variants at ACE2 and TMPRSS2 Genes Involved in SARS-CoV-2 Infection and Human Quantitative Phenotypes. Frontiers in Genetics. Frontiers Media SA. Retrieved from http://dx.doi.org/10.3389/fgene.2020.00613
44. Mahase, E. (2020, March 12). Covid-19: WHO declares pandemic because of “alarming levels” of spread, severity, and inaction. BMJ. BMJ. Retrieved from http://dx.doi.org/10.1136/bmj.m1036
45. Manolio, T. A., Chisholm, R. L., Ozenberger, B., Roden, D. M., Williams, M. S., Wilson, R., Bick, D., et al. (2013, April). Implementing genomic medicine in the clinic: the future is here. Genetics in Medicine. Elsevier BV. Retrieved from http://dx.doi.org/10.1038/gim.2012.157
46. McDonald, C. J., Murray, R., Jeris, D., Bhargava, B., Seeger, J., & Blevins, L. (1977, March). A computer-based record and clinical monitoring system for ambulatory care. American Journal of Public Health. American Public Health Association. Retrieved from http://dx.doi.org/10.2105/ajph.67.3.240
47. Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M. J., et al. (2015, March 9). Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nature Genetics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/ng.3242
48. Mitchell, S. L., Hall, J. B., Goodloe, R. J., Boston, J., Farber-Eger, E., Pendergrass, S. A., Bush, W. S., et al. (2014, April 15). Investigating the relationship between mitochondrial genetic variation and cardiovascular-related traits to develop a framework for mitochondrial phenome-wide association studies. BioData Mining. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1186/1756-0381-7-6
49. Moon, C. Y., Schilder, B. M., Raj, T., & Huang, K. (2021, June). Phenome-wide and expression quantitative trait locus associations of coronavirus disease 2019 genetic risk loci. iScience. Elsevier BV. Retrieved from http://dx.doi.org/10.1016/j.isci.2021.102550
50. Novelli, A., Andreani, M., Biancolella, M., Liberatoscioli, L., Passarelli, C., Colona, V. L., Rogliani, P., et al. (2020, September 3). <scp>HLA</scp> allele frequencies and susceptibility to <scp>COVID</scp> -19 in a group of 99 Italian patients. HLA. Wiley. Retrieved from http://dx.doi.org/10.1111/tan.14047
51. Oetjens, M. T., Luo, J. Z., Chang, A., Leader, J. B., Hartzel, D. N., Moore, B. S., Strande, N. T., et al. (2020, November 12). Electronic health record analysis identifies kidney disease as the leading risk factor for hospitalization in confirmed COVID-19 patients. (H. Mischak, Ed.)PLOS ONE. Public Library of Science (PLoS). Retrieved from http://dx.doi.org/10.1371/journal.pone.0242182
52. Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., et al. (2020, December 11). Genetic mechanisms of critical illness in COVID-19. Nature. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41586-020-03065-y
53. Roden, D., Pulley, J., Basford, M., Bernard, G., Clayton, E., Balser, J., & Masys, D. (2008, May 21). Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. Clinical Pharmacology & Therapeutics. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/clpt.2008.89
54. Salvatore, M., Gu, T., Mack, J. A., Prabhu Sankar, S., Patil, S., Valley, T. S., Singh, K., et al. (2021, March 25). A Phenome-Wide Association Study (PheWAS) of COVID-19 Outcomes by Race Using the Electronic Health Records Data in Michigan Medicine. Journal of Clinical Medicine. MDPI AG. Retrieved from http://dx.doi.org/10.3390/jcm10071351
55. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference (2nd ed.): Boston, Massachusetts: Houghton Mifflin
56. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J. G., Zgaga, L., Manolio, T., Rudan, I., et al.

(2011, November). Abundant Pleiotropy in Human Complex Diseases and Traits. The American Journal of Human Genetics. Elsevier BV. Retrieved from http://dx.doi.org/10.1016/j.ajhg.2011.10.004

57. Song, R. J., Ho, Y.-L., Schubert, P., Park, Y., Posner, D., Lord, E. M., Costa, L., et al. (2021, May 13). Phenome-wide association of 1809 phenotypes and COVID-19 disease progression in the Veterans Health Administration Million Veteran Program. (V. C. Marconi, Ed.)PLOS ONE. Public Library of Science (PLoS). Retrieved from http://dx.doi.org/10.1371/journal.pone.0251651
58. Staley, J. R., Blackshaw, J., Kamat, M. A., Ellis, S., Surendran, P., Sun, B. B., Paul, D. S., et al. (2016, June 17). PhenoScanner: a database of human genotype–phenotype associations. Bioinformatics. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/bioinformatics/btw373
59. Thevarajan, I., Nguyen, T. H. O., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C. E., Jia, X., et al. (2020, March 16). Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. Nature Medicine. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41591-020-0819-2
60. Tyler, A. L., Crawford, D. C., & Pendergrass, S. A. (2015, July 28). The detection and characterization of pleiotropy: discovery, progress, and promise. Briefings in Bioinformatics. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/bib/bbv050
61. Verma, A., Tsao, N., Thomann, L., Ho, Y.-L., Iyengar, S., Luoh, S.-W., Carr, R., et al. (2021, May 23). A Phenome-Wide Association Study of genes associated with COVID-19 severity reveals shared genetics with complex diseases in the Million Veteran Program. Cold Spring Harbor Laboratory. Retrieved from http://dx.doi.org/10.1101/2021.05.18.21257396
62. Wang, F., Huang, S., Gao, R., Zhou, Y., Lai, C., Li, Z., Xian, W., et al. (2020, November 10). Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. Cell Discovery. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41421-020-00231-4
63. Wei, W.-Q., Bastarache, L. A., Carroll, R. J., Marlo, J. E., Osterman, T. J., Gamazon, E. R., Cox, N. J., et al. (2017, July 7). Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. (A. Rzhetsky, Ed.)PLOS ONE. Public Library of Science (PLoS). Retrieved from http://dx.doi.org/10.1371/journal.pone.0175508
64. WHO, (2022). *Retrieved from* https://www.who.int/emergencies/diseases/novel-coronavirus-2019
65. Williamson, E. J., Walker, A. J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C. E., Curtis, H. J., et al. (2020, July 8). Factors associated with COVID-19-related death using OpenSAFELY. Nature. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1038/s41586-020-2521-4
66. Wu, Y., Feng, Z., Li, P., & Yu, Q. (2020, October). Relationship between ABO blood group distribution and clinical characteristics in patients with COVID-19. Clinica Chimica Acta. Elsevier BV. Retrieved from http://dx.doi.org/10.1016/j.cca.2020.06.026
67. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020, March 27). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. American Association for the Advancement of Science (AAAS). Retrieved from http://dx.doi.org/10.1126/science.abb2762
68. Yang, J., Chen, T., & Zhou, Y. (2021, January 4). Mediators of SARS-CoV-2 entry are preferentially enriched in cardiomyocytes. Hereditas. Springer Science and Business Media LLC. Retrieved from http://dx.doi.org/10.1186/s41065-020-00168-4
69. Zhang, Y., Qin, L., Zhao, Y., Zhang, P., Xu, B., Li, K., Liang, L., et al. (2020, April 29). Interferon-Induced Transmembrane Protein 3 Genetic Variant rs12252-C

Associated With Disease Severity in Coronavirus Disease 2019. The Journal of Infectious Diseases. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/infdis/jiaa224

70. Zhou, J., Sun, Y., Huang, W., & Ye, K. (2021, February 2). Altered Blood Cell Traits Underlie a Major Genetic Locus of Severe COVID-19. (A. B. Newman, Ed.)The Journals of Gerontology: Series A. Oxford University Press (OUP). Retrieved from http://dx.doi.org/10.1093/gerona/glab035

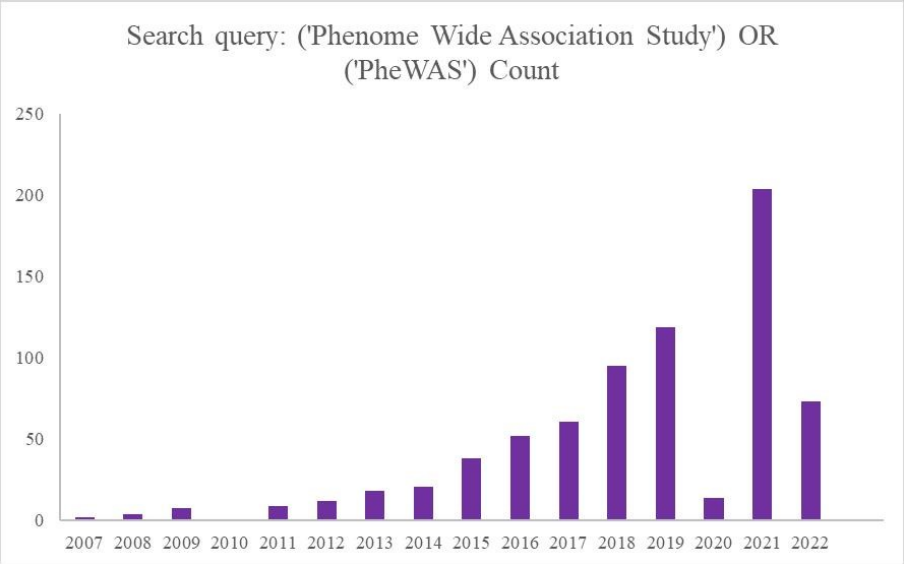

Figure 1. *There are 780 PheWAS publications published in Pubmed so far*

Table 2. *The data of PheWAS analyses carried out to determine the large-scale genotype-phenotype correlation using GWAS and EHR data in COVID-19 research are given in the table. In the PheWAS analyzes performed, there are those with the most significant genotype-phenotype correlation*

Term	rs number	Lokus/Gene	Other PheWAS phenotypes	Covid-19 phenotypes and severity	References
A	rs495828	*ABO* lokus	Venous embolism	Critical illness and hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs505922	*ABO* lokus	Thrombosis	Critical illness and hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs35705950	*MUC5B* locus	İdiopathic fibrosing alveolitis, respiratory features	Critical illness and hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs61667602	*CRHR1*	reduced risk of pulmonary fibrosis, Post-inflammatory pulmonary fibrosis, Idiopathic fibrosing alveolitis, Other alveolar and parietoalveolar pneumonopathy	Hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs11085727	*TYK2* locus	Psoriasis, Cutaneous lupus erythematosus, Lupus (localized and systemic), Psoriatic arthropathy, reduced risk for autoimmune conditions,	Critical illness and hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs9501257	*HLA-DPB1*	Rosacea	Critical illness of Covid-19	(Verma *et al*., 2021)
A	rs9268576	*HLA-DRA*	Rheumatoid arthritis and other inflammatory polyarthropathies	Hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs111837807	*CCHCR1*	Sarcoidosis, Vitiligo	Critical illness and hospitalization of Covid-19	(Verma *et al*., 2021)
A	rs9896243	*NSF*	Post-inflammatory pulmonary fibrosis	Critical illness of Covid-19	(Verma *et al*., 2021)
B	rs13050728	*TPSG, VEGFR2*		Covid-19 of hospitalization	(Gaziano *et al*., 2021)
B	rs4830976	*ACE2, CA5B, CLTRN, VEGFD*		Covid-19 of hospitalization	(Gaziano *et al*., 2021)
B	rs67959919	*CCR, CCR2*	Monocytes count	Severe Covid-19	(Zhou *et al*., 2021)
B	rs67959919	*CCR3*	Eosinophil count, neutrophil count	Severe Covid-19	(Zhou et al., 2021)
A	rs657152	*ABO* blood group	Clotting time, (PEF) peak expiratory flow, HB concentration, monocyte count, Amlodipine, and aspirin	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs11385942	*LZTFL1, CCR9*	Monocyte, neutrophil, granulocyte, eosinophil and macrophage traits, lymphocyte count, antithrombotic agents, hypertension, Type 2 diabetes, blood clot, DVT (deep vein thrombosis), allergic and atopic diseases, and BMI	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs150892504	*EVAP2*	Platelet count and BMI,	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs138763430	*BRF2*	Lymphocyte count and FEV1/FVC ratio (forced expiratory volume/forced vital capacity), Amlodipine	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs117665206	*TMEM181*	FEV1, PEF, Amlodipine	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs147149459	*ALOXE3*	FVC, PEF, and FEV1	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs151256885	*ALOXE3* (intronic)	Blood clot, eosinophil percentage, DVT, allergic and atopic diseases, Amlodipine and aspirin	Covid-19 mortality	(Crespi *et al*., 2020)
A	rs17264937	*ACE2*	Eosinophils	Covid-19 infection	(Lopera *et al*., 2020)
A	rs5980163	*ACE2*	Triglycerides	Covid-19 infection	(Lopera *et al*., 2020)
A	rs150965978	*TMPRSS2*	Plasma levels of CHIT1 protein	Covid-19 infection	(Lopera *et al*., 2020)
A	rs28401567	*TMPRSS2*	Thrombocytes	Covid-19 infection	(Lopera *et al*., 2020)
B	rs647800		Thyroid-stimulating hormone, Hematocrit, Monocyte count, Hemoglobin concentration, Red blot cell count, Activated partial thromboplastin time, Total kolestrol, Legs-leg fat ratio (male),	severe COVID-19 with respiratory failure	(Moon *et al*., 2021)
B	rs11385942		Monocyte percentage of White cells, Monocyte count	severe COVID-19 with respiratory failure	(Moon *et al*., 2021)
B	rs3934992		Waist-hip ratio (adjucted for BMI)	severe COVID-19 with respiratory failure	(Moon *et al*., 2021)
B	rs134130		Celebellar vermal lobules VI VII	severe COVID-19 with respiratory failure	(Moon *et al*., 2021)
B	rs12610495	*DPP9*	Fibrotic idiopathic interstitial pnemonias	severe COVID-19 with respiratory failure	(Moon *et al*., 2021)