The FASEB Journal

FASEB
Federation of American Societies for Experimental Biology

RESEARCH ARTICLE OPEN ACCESS

Genomic Locus Tagged by Lung Function GWAS SNV rs12477314 (2q37.3) Acts as a Regulatory Region for a Systemic Inflammatory Phenotype

Godwin M. Grech[1] | Katherine A. Fawcett[2] | Robert J. Hall[3] | Godfrey Grech[4] | Roger Ellul-Micallef[1] | Ian P. Hall[3] | Anthony G. Fenech[1]

[1]Department of Clinical Pharmacology and Therapeutics, Faculty of Medicine & Surgery, University of Malta, Msida, Malta | [2]Department of Population Health Sciences, University of Leicester, Leicester, UK | [3]Division of Respiratory Medicine and NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, UK | [4]Department of Pathology, Faculty of Medicine & Surgery, University of Malta, Msida, Malta

Correspondence: Godwin M. Grech (ggre0010@um.edu.mt)

Received: 2 January 2025 | **Revised:** 1 May 2025 | **Accepted:** 23 May 2025

Funding: Ian P. Hall holds an NIHR Senior Investigator award and is supported by a Wellcome Trust Discovery Award. This study was in part funded by the Malta Government Scholarship Scheme (MGSS).

Keywords: biological pathway analysis | COPD | CRISPR | DNA methylation | enhancers | RNA-seq | SNV

ABSTRACT

SNV rs12477314 (C>T; 1000G MAF = 0.14), which maps to an intergenic region on 2q37.3, is a genome-wide significant association signal for pulmonary function in genome-wide association study meta-analyses. Bioinformatic analysis revealed that the intergenic region in proximity to the sentinel SNV is enriched for histone methylation markers suggestive of active enhancer regions modifiable by DNA methylation. The aim of this study was to investigate the functionality of putative enhancer/s and their potential interaction with CpG islands in the genomic region tagged by rs12477314 and their relevance to lung disease, in particular COPD. Two independent CRISPR/Cas9n-targeted deletions of the putative enhancer/s were performed in an airway epithelial cell line (NCI-H460). Deletion clones were subjected to RNA-Seq, and differential expression gene (DEG) datasets were generated using the *Cufflinks* version 2.2.1 pipeline (p-FDR <0.05). Biological pathway analysis was performed using Qiagen's Ingenuity Pathway Analysis. Associations with the blood proteome were explored in UK Biobank. Our results suggest that the deleted regions are co-acting enhancers regulating overlapping gene expression patterns. The DEG datasets from the two genomic deletions are enriched for similar canonical pathways, which may contribute to a pro-inflammatory phenotype. Pathway-based regulatory effects analysis of the two DEG datasets resulted in identifying potential downstream biological processes. There was overlap between the pathways identified in protein association datasets and the DEG datasets. Our results suggest that the genomic region tagged by SNV rs12477314 constitutes a regulatory region responsible for regulating biological pathways conducive to a systemic inflammatory phenotype.

1 | Introduction

In the past 15 years, numerous large-scale GWAS have been conducted focusing both on lung function (mainly spirometric) measures, as well as COPD [1–8]. These studies have identified a large number of genomic regions showing association with lung function traits and disease. However, identifying the mechanisms underlying these observed associations remains a challenge [9].

The locus tagged by rs12477314 (downstream of *HDAC4*) was one of the first identified as showing association with lung

function measures (FEV_1/FVC). The association was discovered in a GWAS and large-scale follow-up study led by the SpiroMeta and CHARGE consortia [6]. SNV rs12477314 (C>G,T) is a relatively common variant present across multiple ethnicities, with a minor allele frequency (MAF) of 0.14 (T) in the 1000Genomes Phase 3 combined population. SNV rs12477314 is located in an intergenic region on 2q37.3 flanked by two oppositely transcribed protein-coding genes, *HDAC4* and *TWIST2*, and a lincRNA: *FLJ43879*. The T allele of rs12477314 was associated with an increased FEV_1/FVC ratio. The effect size of rs12477314 explained a modest proportion of the additive genetic variance in FEV_1/FVC. Soler Artigas et al. reported consistent directions of estimated effects on lung function between adults and children at 7–9 years of age for rs12477314, indicating a possible effect on lung function development. The effect on lung function of rs12477314 was similarly apparent in both never smokers and ever smokers.

In a recent multi-ancestry genome-wide association analyses focusing on genes and pathways influencing lung function and COPD, SNV rs12477314 as well as other SNVs in high LD with rs12477314 were all reported to have significant posterior probabilities, suggesting either a shared causal variant across traits or more than one distinct causal variant underlying the association signals ([10]). However, SNV rs12477314 was reported to have the highest posterior probability when compared to the other SNVs within its LD block [10].

Our initial bioinformatic analysis revealed that the intergenic region in proximity to SNV rs12477314 is enriched for histone methylation markers suggestive of active enhancer regions with potential for regulation by DNA methylation. In this study, we therefore investigated the possibility that epigenetic regulation at the genomic region tagged by SNV rs12477314 could be driving changes in gene expression that might explain the observed association with lung function.

2 | Methods

2.1 | Experimental and Methodology Overview

A bioinformatics review of genomic region 2q37.3, in proximity to sentinel SNV rs12477314, was performed with the aim of gaining insights into the mechanisms underlying the GWAS signal associating sentinel SNV to lung function measures. Functionality of putative enhancer/s identified from the bioinformatics review and their potential interaction with CpG islands in the intergenic regions encompassing the LD block of sentinel SNV rs12477314 were investigated through targeted CRISPR Cas9n deletion/s of putative enhancer regions, followed by RNA-Seq of generated CRISPR deletion clones as compared to wild type (WT) cells, and generation of differentially expressed genes' datasets and resultant DEG data analysis.

2.2 | Bioinformatic Review of Genomic Region 2q37.3, in Proximity to Sentinel SNV

Web-based resource LDlink https://ldlink.nci.nih.gov [11] was used to investigate linkage disequilibrium ($r^2 \geq 0.5$) between SNV rs12477314 and other known SNVs in the genomic region in proximity to the sentinel SNV. Methylation quantitative trait loci (meQTLs) of sentinel SNV and of SNVs within its LD block ($r^2 \geq 0.5$) were explored using the Genevar 3.3 database (http://www.sanger.ac.uk/resources/software/genevar). Expression quantitative trait loci (eQTLs) of sentinel SNV and of SNVs within its LD block ($r^2 \geq 0.5$) were also explored. The genotype-tissue expression (GTEx v8) portal (https://gtexportal.org) was queried for eQTLs in lung tissue, and also other tissue datasets available in the database. ChIP data sets originating from various cell types published in the Integrative Genomic Viewer (IGV)—Broad Institute (http://software.broadinstitute.org/software/igv/; [12]) were analyzed for chromatin predictive markers of enhancer regions in the intergenic region around sentinel SNV, ranging from around 40 Kbp upstream to around 40 Kbp downstream of rs12477314 (Chr 2:239877148; GRCH37.13) and encompassing the LD block $r^2 \geq 0.5$. H3K4me1 and H3K27ac were used as the markers for gene enhancer regions. Markers for promoter regions/activation, insulator regions, and repressors—H3K4me3, CTCF, and H3K27me3, respectively, were also analyzed.

2.3 | PheWAS Analysis and Protein Association Analysis

Deep-pheWAS [13] was used to test the association of rs12477314 with 1899 phenotypes in UK Biobank, including clinically curated composite phenotypes for health outcomes, continuous traits from primary care data, disease progression, and drug response. Binary outcomes with fewer than 50 cases and continuous traits in fewer than 100 individuals were excluded. An additive genetic model was tested using either linear regression (for continuous traits) or logistic regression (for binary outcomes), adjusting for age, sex, genotyping array, and the first 10 ancestry-based principal components (PCs), with a false discovery rate of 1%.

Protein levels were measured from UK Biobank participants' blood samples ($N = 54\,219$) using the Olink platform. Association between rs12477314 C>T and protein levels were tested assuming an additive genetic model. All proteins showing association (p-value < 0.05) were then subjected to biological pathway analysis and upstream regulator analysis using Qiagen's Ingenuity Pathway Analysis (IPA). For canonical pathway analysis, the $-\log$ (p-value) > 1.3 (p-value < 0.05) was taken as the threshold, the Z-score > 2 was defined as the threshold of significant activation, whilst Z-score < -2 was defined as the threshold of significant inhibition. For upstream regulators, the p-value of overlap < 0.05 was set as the threshold.

2.4 | Selection of Genomic Regions Targeted for CRISPR Deletion

On the basis of the bioinformatic data, in particular chromatin signatures for enhancer regions, as well as eQTLs and CpG DNA methylation markers identified as quantitative trait loci of SNV rs12477314 (Chr 2:239877148; GRCH37.13), and its proxy SNVs, selected genomic regions were targeted for deletion. The CRISPR/Cas9 nuclease technique was used for this purpose.

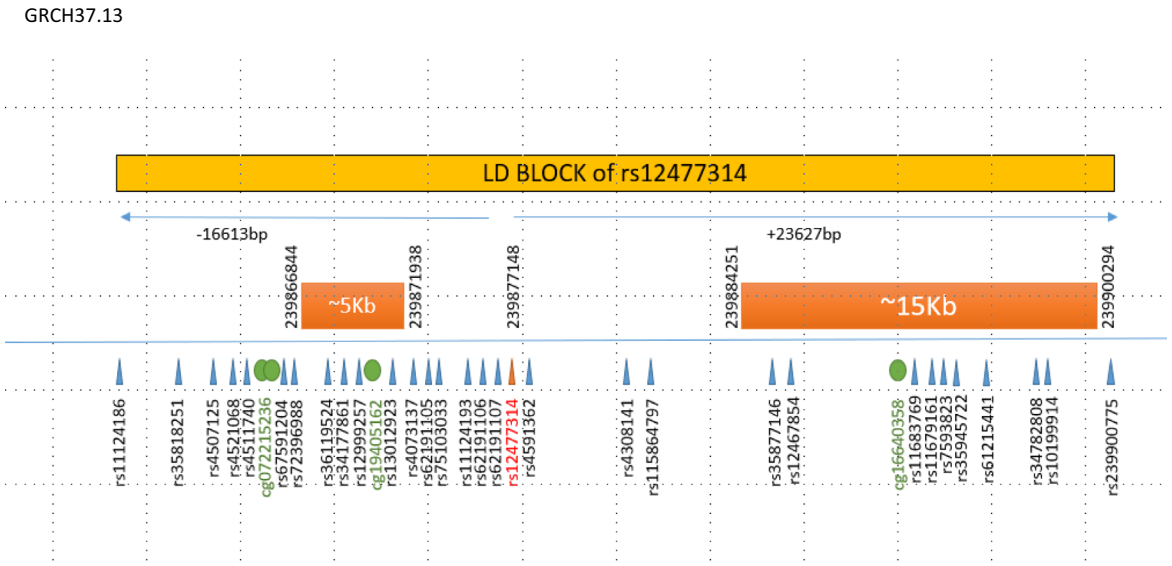

FIGURE 1 | A diagrammatic presentation of the SNVs constituting the LD block ($r^2 \geq 0.5$) of rs12477314, superimposed with the two targeted CRISPR deletions shown as orange boxes: 'Deletion A' consisting of a circa 5 Kb CRISPR deletion upstream of rs12477314 and 'Deletion B' consisting of a circa 15 Kb CRISPR deletion downstream of rs12477314. The diagram shows that 'Deletion A' encompasses a CpG DNA methylation marker (cg19405162) and four proxy SNVs of rs12477314: rs36119524, rs34177861, rs12999257 and rs13012923, whereas 'Deletion B' encompasses a CpG DNA methylation marker (cg16640358) and nine proxy SNVs of rs12477314: rs35877146, rs12467854, rs11683769, rs11679161, rs7593823, rs35945722, rs61215441, rs34782808, and rs10199914.

Two regions were targeted for CRISPR deletion: (i) Deletion 'A' consisting of a circa 5 Kb region upstream of rs12477314 (Chr 2:239866844–239871938; GRCH37.13) and (ii) Deletion 'B' consisting of a circa 15 Kb region downstream of rs12477314 (Chr2:239884251–239900294; GRCH37.13). Figure 1 is a diagrammatic presentation of the LD block ($r^2 \geq 0.5$) of rs12477314, superimposed with the two targeted CRISPR deletions, and annotating the presence of CpG DNA methylation markers. Table 1 summarizes the genomic features within the two targeted CRISPR deletions, including SNVs in LD with sentinel SNV, meQTLs (cg methylation markers), eQTLs in lung tissue, and H3K4me1 chromatin signatures indicative of enhancer activity, which are associated or potentially associated with the lung function GWAS signal as tagged by sentinel SNV rs12477314. The selected deletion regions also took into consideration CRISPR guide RNA design and the related predicted efficiency score.

2.5 | CRISPR Plasmid Construction

CRISPR guide RNA was designed using the tool at crispr.mit.edu. Guide RNA sequences meeting the 85% efficiency quality threshold were selected from the generated nuclease analysis report. The guide RNA sequences were synthesized as two complementary single-stranded DNA oligonucleotides, which were subsequently annealed and inserted into the CRISPR/Cas9 nuclease plasmid as per the manufacturer's protocol. We used two different CRISPR/Cas9 nuclease constructs each targeting one of the two CRISPR/Cas9 mediated genome deletions: (i) an all-in-one vector construction system supplied by Addgene (#42230) and based on the pX330 vector originally developed by the Feng Zhang laboratory; (ii) CRISPR/Cas9 GeneArt (GenScript) CRISPR nuclease vector. Guide RNA sequences of 'Deletion A': GCATGCTCGTGTGCTAGCTC 5′ of targeted deletion cutting at location Chr2: 239866844 (GRCH37.13), and GATCCAGGTTTCTGCGTCCG 3′ of targeted deletion cutting at location Chr2:239871938 (GRCH37.13). Guide RNA sequences of the 'Deletion B': GCGTATACTGCTCGGGTGAT 5′ of targeted deletion cutting at location Chr2:239884251 (GRCH37.13), and AACGTTGCCCTAGTACCCAT 3′ of targeted deletion cutting at locationChr2: 239900294 (GRCH37.13).

A scrambled control plasmid expressing the non-specific gRNA (5′-GCACTACCAGAGCTAACTCA-3′) was applied as a negative control. The guide RNA sequences were synthesized as two complementary single stranded DNA oligonucleotides, which were subsequently annealed and inserted into CRISPR/Cas9 nuclease plasmid. This non-specific gRNA sequence has been cited in various publications concerning gene targeting using CRISPR/Cas9 [14–16].

2.6 | Cell Culturing

ATCC NCI-H460 cell line HTB-177, a human lung epithelial large cell carcinoma cell line, was used as the cell model. The NCI-H460 cell line harbors the LD block of sentinel SNV rs12477314, as revealed by genotyping by Sanger sequencing of sentinel SNV and several SNVs (rs34177861, rs12999257, and rs13012923) within its LD block ($r^2 > 0.5$). NCI-H460 cells were cultured in RPMI-1640 (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS). Cells were kept in a humidified 10% CO_2 atmosphere at 37°C. Cells were transfected with CRISPR construct plasmid/s and plasmid vector pBabe-puro (Addgene #1764) encoding a puromycin resistance gene, using NeuroMag (OzBiosciences) and magnetofection technology.

TABLE 1 | Genomic features within the two targeted CRISPR deletions.

Targeted CRISPR deletion	Deleted genomic feature	meQTL SNVs	eQTL/Tissue	Other genomic associations/relevance	Association with lung function GWAS [6]
Deletion A	cg19405162	rs12477314			Sentinel SNV rs12477314
	cg19405162	rs4511740		rs4511740 is also eQTL for *TWIST2* in lung tissue	LD $r^2 = 0.8367$, with rs12477314
	rs34177861		*TWIST2*/Lung		LD $r^2 = 0.9461$, with rs12477314
	rs36119524				LD $r^2 = 0.9461$, with rs12477314
	rs12999257				LD $r^2 = 0.8046$, with rs12477314
	rs13012923				LD $r^2 = 0.8718$, with rs12477314
	H3K4me1 chromatin signatures			Putative enhancer activity	
Deletion B	cg16640358	rs4511740		rs4511740 is also eQTL for *TWIST2* in lung tissue	LD $r^2 = 0.8367$, with rs12477314
	rs35877146				LD $r^2 = 0.7890$, with rs12477314
	rs12467854				LD $r^2 = 0.7379$, with rs12477314
	rs11683769				LD $r^2 = 0.7438$, with rs12477314
	rs11679161				LD $r^2 = 0.7438$, with rs12477314
	rs7593823				LD $r^2 = 0.7438$, with rs12477314
	rs35945722				LD $r^2 = 0.7656$, with rs12477314
	rs61215441				LD $r^2 = 0.4840$, with rs12477314
	rs34782808				LD $r^2 = 0.5165$, with rs12477314
	H3K4me1 chromatin signatures			Putative enhancer activity	

Note: Table summarizes genomic features including SNVs in high LD of $r^2 \geq 0.5$ with sentinel SNV rs12477314, meQTLs (cg methylation markers), eQTLs in lung tissue, and H3K4me1 chromatin signatures indicative of enhancer activity, which are associated or potentially associated with the lung function GWAS signal [6] as tagged by sentinel SNV rs12477314. Bioinformatic data sourced from publicly available resources: LDlink https://ldlink.nci.nih.gov [11] for Linkage disequilibrium data; Genevar database (https://www.sanger.ac.uk/tool/genevar), and mQTLdb database (http://www.mqtldb.org/) for analysis of meQTLs; GTEx database v8 (https://gtexportal.org/) for analysis of eQTLs; and IGV (https//software.broadinstitute.org/software/igv/) for annotation of enhancer chromatin signatures.

2.7 | Isolation of CRISPR Clones

CRISPR positive clones were enriched for using puromycin selection (1 μg/mL of culture medium) and isolated by serial dilution in 96-well plates. CRISPR clones were first screened for targeted deletion using PCR and gel electrophoresis and then confirmed with Sanger sequencing following clone expansion. Primers for screening and sequencing of CRISPR clones were designed using the NCBI Primer BLAST online tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast). For the targeted 'Deletion A', screening primers were designed bordering the target deletion region (forward primer 5′-GGCAATGAGTGGCAAGTTGG-3′; reverse primer 5′-AGCAAACACTGATGGAGCCT-3′). This single set of primers captured both the CRISPR deletion product and the non-deleted product. For the targeted 'Deletion B', and considering the significant size of the deletion, two sets of screening

primers were designed: one set consisted of primers bordering the target deletion region, capturing the CRISPR deletion product (forward primer 5′-GCACGTTGGTAGGAATGTAAAC-3′; reverse primer 5′-TTAGAAGATATGCGGGTGGC-3′), and the second set of primers designed with primers within the targeted deletion sequence, capturing the non-deleted product (forward primer 5′-AGAGAACGCCCTGACAATGG-3′; reverse primer 5′-CTGGGGCCCGTCACATAAAT-3′). Sequencing primers were designed bordering the target deletion region: 'Deletion A' (forward primer 5′-AGGCCTTTCCCCCTAAGAAC-3′; reverse primer 5′-GATGGTGTCCTCGCCCTTTC-3′) with a theoretical amplicon size of 837 bp for the deleted product, whereas 'Deletion B' (forward primer 5′-ATCAGCCCAAATGCCCATC-3′; reverse primer 5′-ATACCATCACACACCCACC-3′) with a theoretical amplicon size of 601 bp for the deleted product.

Three bi-allelic clones were isolated for 'Deletion A'. For all three bi-allelic clones, sequencing confirmed a deletion of 5094 bp, in the region between 238945148 and 238950242 on Chr:2, GRCh38.p12, thus corresponding to the CRISPR target deletion sites as designed. One monoallelic clone was isolated for 'Deletion B'. Sequencing confirmed a deletion of 15 639 bp, in the region between 238962985 and 238978624 on Chr:2, GRCh38.p12. The latter implied a deletion offset of +430 bp (insertion relative to predicted deletion) at the 5′end, and −26 bp (deletion relative to predicted deletion) at the 3′end.

2.8 | RNA-Seq Experiments

H460 cells were transfected at passage 5 with the CRISPR/Cas9n plasmids targeting the desired genome deletions, as well as the pBabe-puro plasmid vector for eventual puromycin-mediated CRISPR clone enrichment. Transfected cells underwent a further four subculturing passages prior to RNA extraction. In parallel to the above procedure, H460 cells were transfected with the CRISPR/Cas9 scrambled control plasmid and pBabe-puro. Similarly, wild type/non-treated H460 cells were cultured as controls in parallel with the CRISPR/Cas9-transfected cells and subcultured over the same time period as for the CRISPR/Cas9-transfected cells prior to RNA extraction. RNA samples originating from the below-listed cell samples were prepared for RNA-Seq and differential gene expression (DEG) analysis: (i) 3 biological replicates of CRISPR 'Deletion A' H460 clones; (ii) 2 technical replicates of CRISPR 'Deletion B' H460 clones; (iii) 3 biological replicates of scrambled CRISPR H460 cells (Negative control); (iv) 3 biological replicates of wild type non-transfected H460 cells.

2.9 | RNA-Seq Bioinformatic Workflow

RNA-Seq of CRISPR deletion clones was compared to wild type cells through generation of differential expressed genes' datasets and resultant DEG data analysis. We tested the null hypothesis that the logarithmic fold change between treatment and control for gene expression is zero, i.e., that the gene is not at all affected by the treatment (in our case effect of targeted CRISPR deletion). RNA was extracted from CRISPR clones using QIAzol. Quantitation of RNA and evaluation of integrity was determined by electrophoretic assays using the Agilent 2100 Bioanalyzer.

RNA-Seq was performed at the RNA-Seq service provider, BGI Genomics, China. RNA-Seq protocol consisted of four standard steps: (i) quality checks and filtering of the raw reads to obtain the 'clean reads' (ii) alignment of raw reads to the reference genome, (iii) assembly of gene expression from aligned reads, and (iv) generation of differential gene expression from assembled gene expression. Clean reads were achieved by filtering out the low-quality reads, reads with adaptors, and reads with unknown bases (N bases more than 5%), using SOAPnuke v1.5.2. Clean reads were stored in the FASTQ format and genome-mapped using HISAT v2.0.4 (http://www.ccb.jhu.edu/software/hisat). The .bam files contained data that was aligned to human genome assembly GRCh37-hg19. Differential gene expression analysis was carried out using *cufflinks* tools version 2.2.1, freely available from http://cole-trapnell-lab.github.io/cufflinks/. Assembled transcripts from each .bam file were generated using *cufflinks*. This generated the transcript assembly file "transcripts.gtf" for each .bam file, which listed the FPKM values for genes for each sample analyzed. The following gene expression data files were generated: three biological replicates H460-CRISPR1, H460-CRISPR2, and H460-CRISPR3 for 'Deletion A'; two technical replicates H460-CRISPR4 and H460-CRISPR5 for 'Deletion B'; three replicates H460-WT1, H460-WT2, and H460-WT3 for the wild type H460 cell; and three replicates H460-WTSCR1, H460-WTSCR2, and H460-WTSCR3 for the wild type scrambled H460 cell. All transcripts.gtf files were merged into one single transcriptome assembly. This step also included the incorporation of a genomic annotation file for GRCh37-hg19, which was downloaded from the University of California Santa Cruz Genomic Institute (UCSC) at URL http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips/genes/hg19.refGene.gtf.gz. The .gz file was expanded to hg19.refGene.gtf before being used. Merging was carried out using *cuffmerge*. The *cuffmerge*-generated "merged.gtf" files were then used to carry out differential expression analysis on all samples. Differential gene expression files were subsequently generated using *cuffdiff* on the "merged.gtf" file, multiple times, with each run specifying which samples were to be compared against each other for differential gene expression analysis. This procedure generated multiple files, including a tab-delimited "gene_exp.diff" file, which contained the respective differential gene expression data, as well as the following fields for each differentially expressed gene: test id, gene id, gene locus, sample_1, sample_2, status, value_1, value_2, $\log_2$ (fold change), test stat, *p*-value, *q*-value, significant. We applied the default false discovery rate (FDR) for *Cuffdiff* of 5%. *Cuffdiff* reports/output included the gene- and transcript-related attributes, i.e., common name and location, gene and transcript expression level changes, and main statistics of fold change (in $\log_2$ scale), *p*-value, and *q*-value (FDR adjusted *p*-value*; p*-FDR). Application of the FDR adjusted *p*-value generated the list of differentially expressed genes passing multiple-test adjustment controlled for false discovery rate. Thresholds of *q*-value (*p*-FDR) < 0.05 and of $\log_2$ fold change of ± 1 (fold change of ± 2) were sequentially applied in order to determine the statistically relevant differentially expressed genes. Significant differentially expressed genes for each analysis were ranked on the basis of $\log_2$ fold change.

The bioinformatic workflow was applied in two stages. In the first stage, we generated DEG datasets for each of the five CRISPR deletion clones using the *cufflinks* pipeline by

comparing each CRISPR deletion clone with the H460 wild type cells samples set made up of three biological samples replicates. This approach allowed us to analyze DEGs $\log_2$ fold change across the individual deletion clones for each of the two CRISPR deletions. In the second stage, we generated DEG datasets for the CRISPR deletion sample set as a whole, using the *cufflinks* pipeline. This approach involved gene expression data comparison between two individual sample sets. For example, H460-CRISPR 'Deletion A' dataset made up of gene expression data of H460-CRISPR1, H460-CRISPR2, and H460-CRISPR3 was compared as one dataset, with the H460-Wild Type sample set made up of gene expression data of H460-WT1, H460-WT2, and H460-WT3 to generate DEG dataset 'H460-WT vs CRISPR Deletion A'. Five DEG datasets were consequently generated: H460-WT vs. H460-WTSCR; H460-WT vs. CRISPR Deletion A; H460-WTSCR vs. CRISPR Deletion A; H460-WT vs. CRISPR Deletion B; H460-WTSCR vs. CRISPR Deletion B.

Heatmap analysis was applied to the DEG datasets to evaluate sample data clustering and differential gene expression variation across the individual samples. Fold changes ($\log_2$ fold changes) and *p*-values (negative log10-transformed *p*-values) were combined to provide a graphical summary of DEGs in the form of a 'volcano plot'. Volcano plots were generated using VolcaNoseR web based tool (https://huygens.science.uva.nl/VolcaNoseR).

2.10 | Analysis of DEG Datasets

Differentially expressed gene analysis was performed on the generated DEG datasets. In order to isolate only those genes which were differentially expressed as a result of the CRISPR deletions, we compared gene expression across the different cell populations. Differential gene expression which occurred between untreated cells, and those transfected with scrambled guide-RNA construct was omitted from analysis. Furthermore, in order to add robustness to our analysis, we only considered those genes that were differentially expressed both when comparing CRISPR deleted to untreated cells, as well as when comparing CRISPR deleted to CRISPR-scrambled transfected cells. The resultant subset of genes denoted as 'region of interest' genes in the Venn diagram (Figure 2) were generated by inputting the respective lists of differentially expressed genes meeting the differential expression cutoff criteria ($\log_2$ fold change <-1 or $>+1$; *q*-value (*p*-FDR) <0.05) into the tool Venny 2.1.0 available at https://bioinfogp.cnb.csic.es. This procedure was followed independently for the 'Deletion A' and 'Deletion B' datasets. The Venn Diagram 'region of interest' as displayed in the 'gray zone' represents DEGs that are differentially expressed as a consequence of the CRISPR deletions only, and limits the inclusion of DEGs as a consequence of transfection procedure itself and potential resultant off-target effects on differential gene expression. This approach applies an additional level of stringency as it controls for effects other than those due to the CRISPR deletion itself. Once the genes within the region of interest were identified, the respective tab separated .diff files were imported into the Galaxy genomics toolset available at

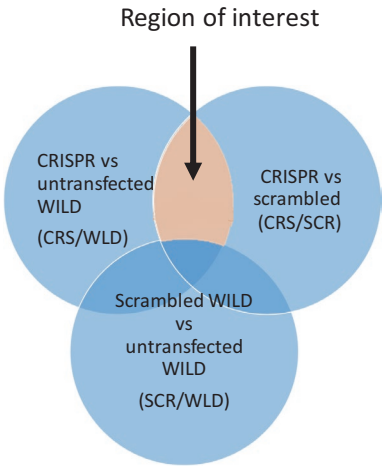

FIGURE 2 | Venn diagram analysis. Venn diagram approach to identify only those differentially expressed genes which showed consistently altered expression when compared to untreated/untransfected cells and to cells transfected with CRISPR plasmids carrying a scrambled guide RNA sequence.

https://usegalaxy.org/ and filtered using appropriate Query functions, in order to only isolate the region of interest subset of genes, with their respective fold change and statistical data.

The resultant DEG datasets were subjected to biological pathway and network analysis using Qiagen's Ingenuity Pathway Analysis (IPA) from Ingenuity Systems, Qiagen China Co. Ltd. IPA core analysis was run on the basis of the following data from our DEG datasets: gene identifier, $\log_2$ fold change and corresponding *p*-FDR value. We defined DEGs as $\log_2$ fold change <-1 or $>+1$; *q*-value (*p*-FDR) <0.05. The following filters were applied in the IPA core analysis: species (*Homo Sapiens*), organ/tissue (lung), cut off values for DEGs $\log_2$ fold change (<-1 or $>+1$, as a minimum), and corresponding *p*-FDR (<0.05). IPA core analysis was run on the DEGs displayed in the 'region of interest' for 'Deletion A' Venn Diagram analysis and 'Deletion B' Venn Diagram analysis. IPA core analysis does not take individual DEG expression levels into account, and is based mainly on the assumption that transcriptionally altered genes have been determined using a suitable cut off applied to the measured expression change, but taking in consideration the direction of fold change, i.e., whether individual DEGs are upregulated or downregulated in the DEG dataset [17]. The DEGs in the 'region of interest' originate from two overlapping DEG datasets (Wild Type vs. CRISPR, and Wild Type Scrambled vs. CRISPR) with their respective fold change values and statistical data; however, in consideration of IPA's analytics, IPA core analysis on the 'region of interest' DEGs with statistical data originating from the two DEG datasets for each CRISPR deletion resulted in practically the same results. The IPA core analysis presented were thus referred to as 'Deletion A' DEG dataset and 'Deletion B' DEG dataset.

For canonical pathway analysis, the $-\log$ (*p*-value) >1.3 (*p*-value <0.05) was taken as the threshold, the *Z*-score >2 was defined as the threshold of significant activation, whilst the *Z*-score <-2 was defined as the threshold of significant inhibition. For regulator effects and downstream biological function

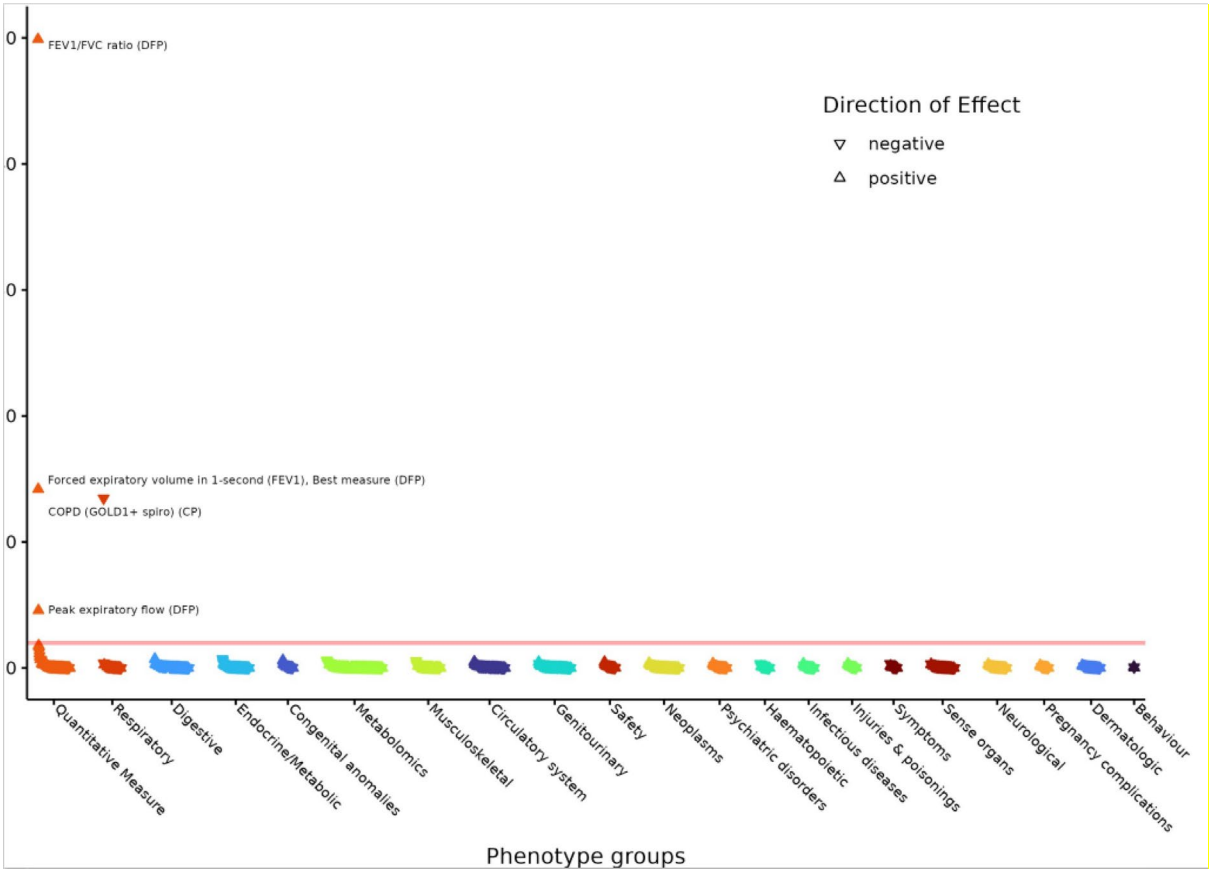

FIGURE 3 | Phenome-wide association study of rs12477314 in UK Biobank participants. Phenotypes are color-coded by category and the direction of the triangle indicates direction of effect with respect to the minor T allele. The red line shows the FDR threshold of 1%. Statistically associated phenotypes are labeled.

analysis, consistency scores were calculated, where a high consistency score indicates accurate results for the analysis. For upstream regulators, the *p*-value of overlap < 0.05 was set as the threshold. The algorithm used for calculating the Z scores and *p* values of overlap has been described by Kraemer et al. [17].

3 | Results

3.1 | Deep-PheWAS Analysis

Deep-PheWAS analysis plots (Figure 3) show the traits/phenotypes resulting from the Phenome-wide association study of rs12477314 in UK Biobank participants. PheWAS analysis confirms the association of the minor T allele with respiratory phenotypes, mainly increased FEV_1/FVC ratio, increased FEV_1, increased peak expiratory flow, and decreased risk of COPD. None of the non-respiratory phenotype groups reached statistical significance in this analysis.

3.2 | Protein Association Analysis and Related Canonical Pathways

Association of SNV rs12477314 with 2823 proteins measured in the UK Biobank blood samples ($N = 54\,219$) using the OLINK platform resulted in one association reaching FDR significance. This was advanced glycosylation end product receptor (AGER) (*p*-value of 8×10^{-6}, Beta value -0.01596). At a nominal significance threshold of *p*-value < 0.05, SNV rs12477314 associated with 203 proteins. These proteins were analyzed in IPA to identify enriched canonical pathways and upstream regulators.

The top 10 enriched canonical pathways are 'TNF signaling', 'Hepatic fibrosis/Hepatic stellate cell activation', 'ISGylation signaling pathway', 'Necroptosis signaling pathway', 'N-acetylglucosamine degradation II', 'Irritable bowel syndrome signaling pathway', 'Cachexia signaling pathway', 'CGAS-STING signaling pathway', 'Role of hypercytokinemia/hyperchemokinemia in the pathogenesis of influenza', and 'Role of PKR in interferon induction and antiviral response'. Table 2 lists the top 10 canonical pathways and corresponding proteins including the $-\log$ (*p*-value) and the *Z*-score. The top 10 enriched canonical pathways show significant *p*-values, ranging from 0.0001 for 'TNF signaling' to 0.001 for 'Role of PKR in interferon induction and antiviral response'. The highest *Z*-score magnitude was obtained for the 'Cachexia signaling pathway' (negative *Z*-score: -3.162). Upstream regulator analysis was also run on IPA. The most significant upstream regulator based on *p*-value was the transcription regulator PFDN5 (*p*-value of 8.27×10^{-6}). The largest negative *Z*-score, indicating significant inhibition, was reported for TNF (*Z*-score: -2.426), whereas the largest positive *Z*-score, indicating

TABLE 2 | Top 10 canonical pathways and corresponding overlapping proteins.

Ingenuity canonical pathways	−log (*p*-value)	Ratio	*Z*-score	Proteins
TNF signaling	4.14	0.0877	−2.236	BAG4, OTUD7B, TAB2, TAX1BP1, XIAP
Hepatic fibrosis/hepatic stellate cell activation	3.94	0.0417	NA	BAX, COL28A1, COL3A1, IL1A, MYH9, PDGFC, SMAD2, TLR4
ISGylation signaling pathway	3.72	0.055	−1.633	EIF2AK2, NFAT5, STAT2, TLR4, TRIM25, XIAP
Necroptosis signaling pathway	3.6	0.0432	−2.646	EIF2AK2, PPIF, STAT2, TAB2, TIMM8A, TLR4, TNIP1
N-Acetylglucosamine degradation II	3.47	0.5	NA	AMDHD2, NAGK
Irritable bowel syndrome signaling pathway	3.4	0.031	−2.333	ATG16L1, CDH1, IL17D, IL17RB, IL1A, IL22, TAB2, TLR4, TNIP1
Cachexia signaling pathway	3.27	0.0272	−3.162	EIF2AK2, EIF2AK3, IL17D, IL1A, INHBC, PRKAG3, PRKAR2A, SMAD2, STAT2, TLR4
CGAS-STING signaling pathway	3.18	0.0435	−2.449	ATG16L1, ATP6V1G1, BAX, EIF2AK3, IL17D, IL1A
Role of hypercytokinemia/ hyperchemokinemia in the pathogenesis of influenza	3.16	0.0543	−2	EIF2AK2, IFIT3, IL1A, STAT2, TLR4
Role of PKR in interferon induction and antiviral response	3.16	0.0432	−2.236	BAX, EIF2AK2, PDGFC, STAT2, TAB2, TLR4

Note: Table includes the *p*-value of the correlation, the ratio of overlap of proteins in the dataset with a number of proteins making up the canonical pathway as in Ingenuity's knowledge database and *Z*-score where applicable. Data generated with IPA core analysis based on proteins meeting the statistical significance threshold of $p < 0.05$.

significant activation was reported for miR-532-3p (*Z*-score: 3.194).

3.3 | Differentially Expressed Gene Analysis

Volcano plots (Figure 4) show the spread of DEGs for the CRISPR deletions when compared to wild type or wild type scrambled cells. A significant number of DEGs (approximately 20% for 'Deletion A', and 28% for 'Deletion B') cluster at the top corners of the volcano plot, i.e., they achieve statistical significance with respect to the applied thresholds. The volcano plot representing the DEG dataset for wild type vs. wild type scrambled comparison showed very few significant changes, suggesting a minimal impact of the transfection procedure on differential gene expression.

Venn diagram analysis identified 1003 DEGs in the 'region of interest' for CRISPR 'Deletion A' and 1451 DEGs in the 'region of interest' for 'Deletion B'. 'Deletion A' DEGs consisted of 672 downregulated genes and 331 upregulated genes. Similarly, 'Deletion B' DEGs showed a higher proportion of downregulated genes: 847 downregulated genes and 604 upregulated genes (Figure 5A). Comparison of the 'region of interest' DEGs from the two CRISPR deletions revealed a significant DEG overlap between the two CRISPR deletions: 338 common differentially expressed genes (Figure 5B). Overall, 302 DEGs (89%) showed the same direction of fold change in both the CRISPR 'Deletion A' and CRISPR 'Deletion B' DEG datasets. The potentially largest contributors to biological function emerging as a consequence of CRISPR deletions, were ranked by magnitude of DEG fold change. Table 3 lists the top 20 common differentially expressed genes in CRISPR 'Deletion A' and CRISPR 'Deletion B'. DEG data shows that the listed top 20 common DEGs show a remarkable similarity in relative magnitude of fold change.

The 'region of interest' DEGs for the two separate CRISPR deletions were analyzed for intra-sample differential expression variation (Figure 6). Both The CRISPR 'Deletion A' and the CRISPR 'Deletion B' heatmaps shows in general consistency in both direction and magnitude of fold change across the individual deletion clone samples.

3.4 | Biological Pathway and Network Analysis of DEGs

An integrated view of gene expression patterns for both differential expression gene datasets originating from the two separate CRISPR deletions was obtained by running Ingenuity Pathway (Ingenuity Systems; Qiagen China Co. Ltd.) core analysis, overlaying DEG data to the Ingenuity knowledge database to generate canonical pathways, information on potential upstream regulators, and downstream networks leading to biological functions and diseases.

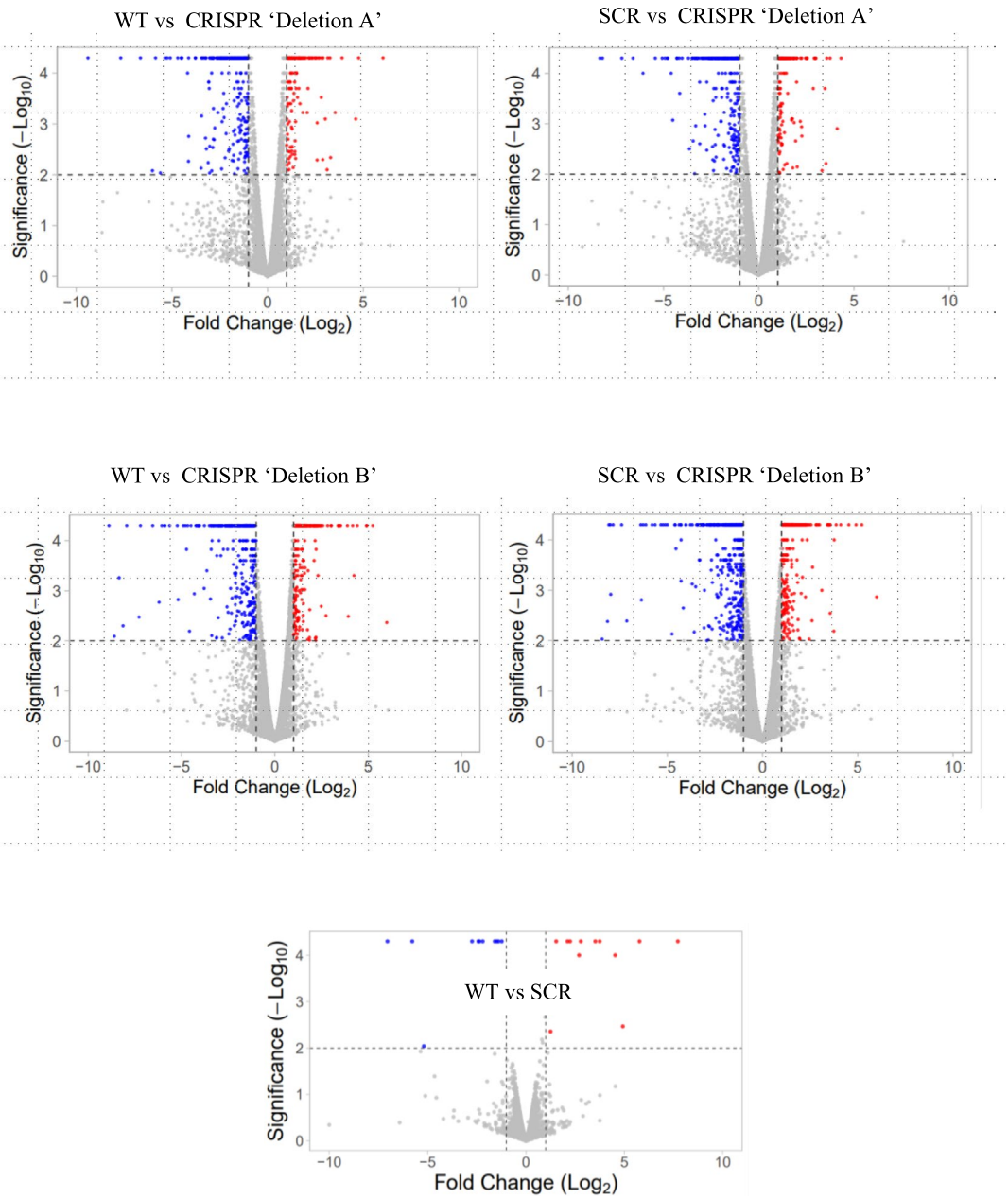

FIGURE 4 | Volcano plots of the DEG datasets. Diagram shows volcano plots of the five DEG datasets: H460-Wild Type vs. CRISPR 'Deletion A' (*WT* vs *'Deletion A'*), H460-Wild Type Scrambled vs. CRISPR 'Deletion A' (SCR vs *CRISPR 'Deletion A'*), H460-Wild Type vs. CRISPR 'Deletion B' (*WT* vs *'Deletion B'*), H460-Wild Type Scrambled vs. CRISPR 'Deletion B' (SCR vs *CRISPR 'Deletion B'*), and H460-Wild Type vs. H460-WT Scrambled (*WT* vs *SCR*). Plot displays the –log10 *p* value vs. log2-fold change in expression. The *p* value cutoff (equivalent to *p*FDR < 0.05 in our datasets) is noted with the horizontal dashed line, whereas the $\log_2$ (fold change) cutoff of < −1 and > 1 is noted with the vertical dashed line. DEGs meeting the significance thresholds of *q* value (*p*FDR) < 0.05 and $\log_2$ (fold change) of < −1 and > 1 are displayed as colored dots: Blue dots represent downregulated DEGs, whereas red dots represent upregulated DEGs. Volcano plots generated with VolcaNoseR web based tool (https://huygens.science.uva.nl/VolcaNoseR).

Figure 7 lists (starting from the most significant) the canonical pathways that were enriched for in both 'Deletion A' and 'Deletion B' DEG datasets. For 'Deletion A', the top six enriched canonical pathways are 'Hepatic fibrosis/Hepatic stellate cell activation', 'Axonal guidance signaling', 'Atherosclerosis signaling', 'Role of macrophages, fibroblasts and endothelial cells in Rheumatoid arthritis', 'Airway pathology in COPD', and 'Human embryonic stem cell pluripotency'. Table 4A lists the top six canonical pathways and corresponding DEGs including the −log (*p*-value) and the *Z*-score. The top six enriched canonical pathways show highly significant *p*-values, ranging from 0.000001 for 'Hepatic fibrosis/Hepatic stellate cell activation' to 0.00089 for 'Human embryonic stem cell pluripotency'; 'Airway pathology in COPD' was enriched for with a *p*-value of 0.00085. The highest *Z*-score magnitude for 'Deletion A' was obtained for the PPAR signaling pathway as indicated by the color saturation in bar graph (negative *Z*-score: −1.897). For 'Deletion B', the top six enriched canonical

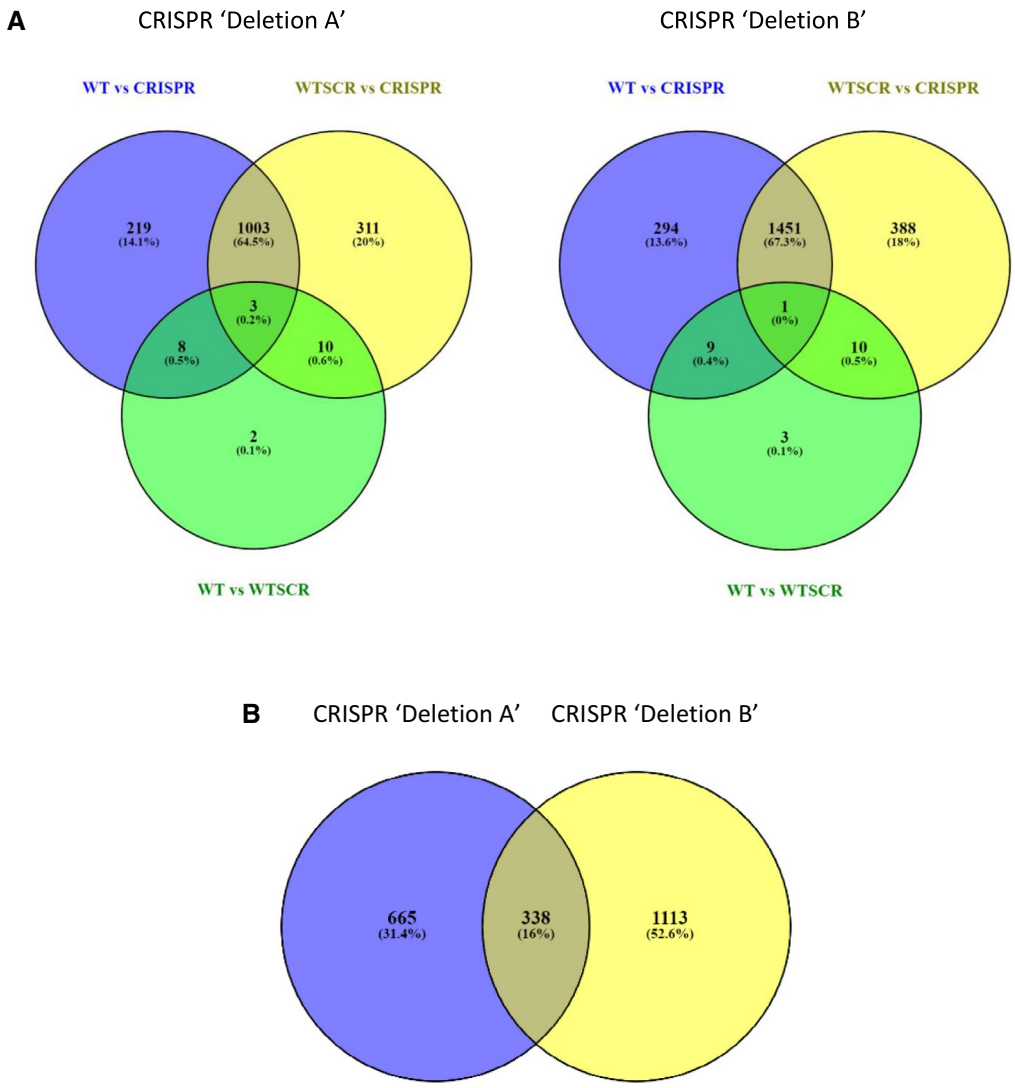

FIGURE 5 | Venn diagram presentation of DEGs. (A) The 'gray zone' represents DEGs that are differentially expressed as a consequence of the CRISPR deletions. Data shown for 'Deletion A' incorporating DEG datasets H460-Wild Type vs. H460-WTScrambled, H460-Wild Type vs. CRISPR 'Deletion A', H460-Wild Type Scrambled vs. CRISPR 'Deletion A'; and 'Deletion B' incorporating DEG datasets H460-Wild Type vs. H460-WTScrambled, H460-Wild Type vs. CRISPR 'Deletion B', H460-Wild Type Scrambled vs. 'Deletion B'. (B) Venn diagram presentation compares DEGs between the 'region of interest' DEGs of the two CRISPR deletion experiments and shows that three hundred and thirty eight genes overlap between the two datasets, i.e., are commonly differentially expressed. DEGs meet the statistical significance criteria: q value (pFDR) threshold of <0.05 and $\log_2$ (fold change) threshold of <-1 and >1. Venn diagram generated using Venny2.1 (http://bioinfogp.cnb.csic.es/ttols/venny/index.html).

pathways are 'Atherosclerosis signaling', 'Hepatic fibrosis/Hepatic stellate cell activation', 'Th1 and Th2 activation pathway', 'Th1 pathway', 'Axonal guidance signaling', and 'Airway pathology in COPD'. Table 4B lists the top six canonical pathways and corresponding DEGs including the −log (p-value) and the Z-score. Highly significant p-values were also obtained for the top six enriched canonical pathways ranging from 0.00016 for 'Atherosclerosis signaling' to 0.004 'Airway pathology in COPD'. The highest Z-score magnitude for 'Deletion B' was obtained for the 'oxidative phosphorylation' pathway (negative Z-score: −2.828).

Upstream regulator analysis was also run on IPA for both 'Deletion A' and 'Deletion B' DEG datasets. The goal of this analysis is to identify molecules upstream of the genes in the datasets that potentially explain the observed gene expression changes. We prioritized resulting upstream regulators that are known transcriptional regulators and that are themselves differentially expressed in our DEG datasets. For 'Deletion A', the most significant upstream regulators/transcription factors based on p-value were *JUN* (p-value of 3.3×10^{-8}), *EGR1* (p-value of 2.0×10^{-7}), and *SMAD3* (p-value of 3.0×10^{-4}). As reflected by a positive significant Z-score, *EGR1* and *SMAD3* were activated in the 'Deletion A' DEG dataset. The transcriptional repressor *PER2* located on 2q37.3 and the closest DEG to the deleted regions was also identified as a transcriptional regulator (p-value of 0.01) for the 'Deletion A' DEG dataset. For 'Deletion B', the most significant upstream regulators/transcription factors based on p-value were *SMARCA4* (p-value of 5.1×10^{-10}), *NFκB1* (p-value of 2.5×10^{-6}), *PRDM1* (p-value of 8.1×10^{-5}), and *EPAS1* (p-value of 1.8×10^{-5}). Z-score values indicated *PRDM1* in an inhibited state, whereas *EPAS1* is activated.

TABLE 3 | Top 20 common differentially regulated genes in 'Deletion A' and 'Deletion B' datasets, ranked by $\log_2$ (fold change) magnitude.

		CRISPR Deletion A		CRISPR Deletion B	
Gene	**Locus**	**$\log_2$ (fold change)**	***q* Value**	**$\log_2$ (fold change)**	***q* Value**
LUM	chr12:91439108–91505274	−9.38415	0.00077	−8.88461	0.0009
PCSK2	chr20:17206751–17465222	−4.54477	0.00077	−9.80508	0.0020
SHH	chr7:155592590–155605157	−5.1648	0.00077	−7.79277	0.0009
CDH18	chr5:19470867–20575822	−7.26571	0.00077	−5.08624	0.0009
LRATD1	chr2:14772840–14792040	−5.28939	0.00077	−6.64472	0.0195
MIR331, MIR3685	chr12:95701557–95704389	−4.69583	0.00077	−7.19665	0.0009
FMOD	chr1:203309755–203320250	−3.44904	0.00077	−7.27492	0.0278
GALNT17	chr7:70596322–71178584	−4.25044	0.00077	−6.22291	0.0009
BMP2	chr20:6748332–6760893	−5.67	0.00077	−4.49277	0.0009
STS	chrX:6966960–7272851	−4.23516	0.00077	−5.37764	0.0009
MYOM1	chr18:3066804–3219966	−3.34288	0.00077	−6.17861	0.0057
CELF4	chr18:34823007–35145812	−5.01455	0.00077	−4.18206	0.0009
CAMK1D	chr10:12391545–12877545	−2.80766	0.00077	−6.20127	0.0165
PAX7	chr1:18957339–19075360	4.76259	0.00077	4.23791	0.0062
RNF150	chr4:141780960–142054623	−4.25622	0.00077	−4.66029	0.0041
NPY1R	chr4:164245113–164253748	−4.19671	0.00077	−4.39221	0.0062
CPNE4	chr3:131252403–131758450	−3.24008	0.00077	−4.85496	0.0161
EGR1	chr5:137801167–137804992	2.48365	0.00077	5.58241	0.0009
TMEM132E	chr17:32906503–32966496	−4.42234	0.00077	−3.58186	0.0009
GABBR2	chr9:101050377–101471473	−3.50807	0.00077	−4.43562	0.0009

Note: DEG statistics ($\log_2$ fold change and *q* value) as obtained in CRISPR 'Deletion A' DEG dataset (Wild Type vs. CRISPR 'Deletion A') and CRISPR 'Deletion B' DEG dataset (Wild Type vs. CRISPR 'Deletion B'). The listed genes meet the statistical significance criteria of $\log_2$ (fold change) threshold of < -1 and > 1, and *q* value (*p*FDR) threshold of < 0.05. $\log_2$ (fold change) values prefixed with a minus sign indicates gene downregulation, as compared to no sign which indicates gene upregulation.

The DEG dataset was also evaluated using Ingenuity Regulator effects analysis, integrating the upstream regulator results with downstream effects including biological functions, diseases, and phenotypic or functional outcomes, with the aim of identifying those biological processes and functions that are likely to be causally affected by up- and downregulated genes, as well as predict whether these same processes are increased or decreased. A consistency score is applied to generated networks based on consistency with the predicted state of the regulator, recorded direction of DEG, and the expected impact on disease or biological function. Figure 8 shows the downstream regulators as identified in the DEG datasets for 'Deletion A' and 'Deletion B' and the resultant corresponding expected downstream effects. For the 'Deletion A' DEG dataset, analysis returned the following downstream biological functions/diseases: chemotaxis, damage of bone, concentration of prostaglandins, and differentiation of bone cells. For 'Deletion B' DEG dataset analysis, the following were the downstream biological functions/diseases: accumulation of neutrophils, recruitment of cells, activation of granulocytes, metabolism of eicosanoid, synthesis of fatty acid, proliferation of bone cells, and growth of bone tissue.

4 | Discussion and Conclusions

The main aim of the studies presented here was to gain insight into possible mechanisms underlying the observed association signal between variants at the rs12477314 locus and lung function. Figure 9 summarizes the hypothetical mechanisms based on the findings of this study.

Phe-WAS analysis applied to UK Biobank participants confirmed the association of the minor T allele with respiratory phenotypes, including increased FEV_1/FVC ratio, and indicating a decreased risk of COPD. Protein analysis of rs12477314 in the UK biobank blood samples resulted in a significant association (negative correlation) with AGER. AGER is a pro-inflammatory pattern recognition receptor that has been implicated in numerous inflammatory diseases, including COPD and interstitial lung disease, mainly pulmonary fibrosis [18]. In COPD, enhanced expression of AGER and its ligands contributes to the chronic inflammatory response in emphysematous COPD [19, 20]. In contrast, AGER expression is downregulated in pulmonary fibrosis [21]. In order to gain further mechanistic insight in the association of rs12477314 with proteins meeting

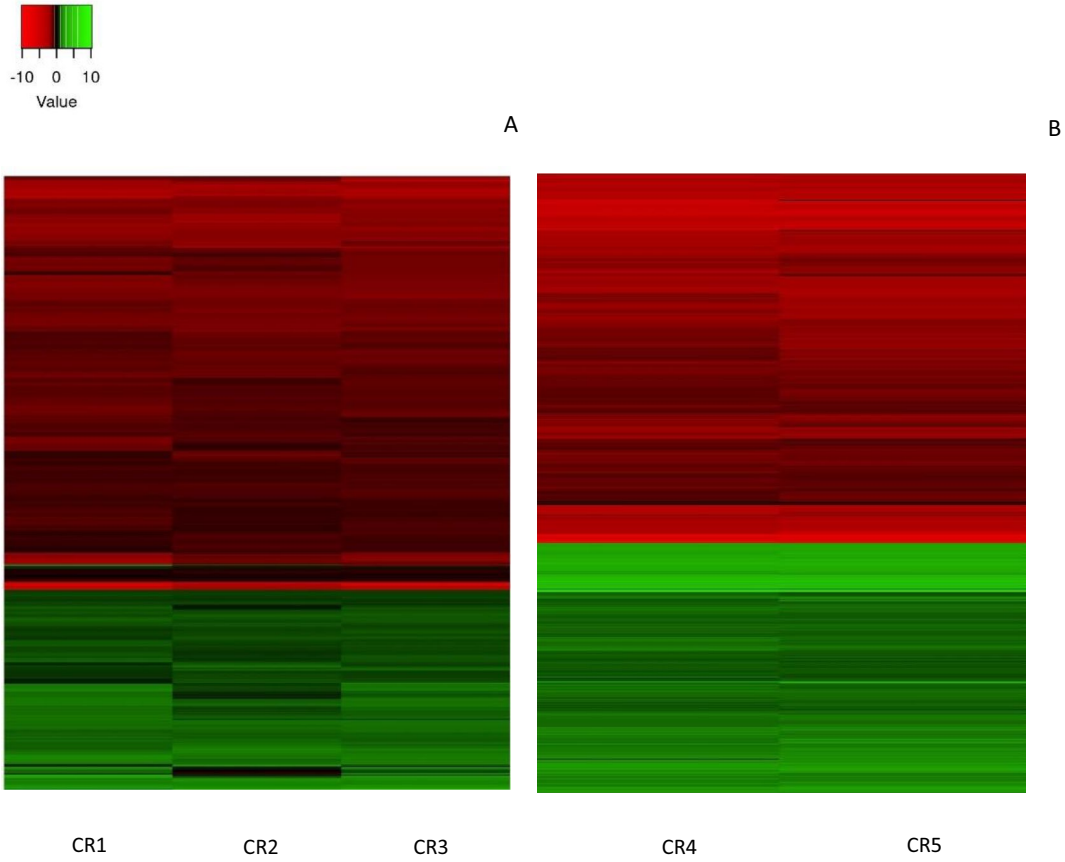

FIGURE 6 | CRISPR 'Deletion A' and 'Deletion B' clones DEGs Heatmaps. Diagram shows the two heatmaps labeled A and B, for CRISPR 'Deletion A' clones and CRISPR 'Deletion B' clones, respectively. Heatmaps represent the $\log_2$-fold change magnitude and direction of detected DEGs (on the *y*-axis) for each of the individual deletion clones, labeled 'CR1'–'CR5', representing H460-CRISPR1—H460-CRISPR5, respectively on the *x*-axis. The color scale ranges from intense red for downregulated DEGs with $\log_2$-fold change < -10, to black for $\log_2$-fold change of 0 (no differential expression), to intense green for upregulated DEGs with $\log_2$-fold change > 10. DEGs generated using *Cuffdiff*, by comparing each CRISPR deletion clone sample with the H460 Wild Type SCR dataset consisting of combined three biological replicates. Heatmaps were generated with a Heatmapper web-based tool (https://www.heatmapper.ca).

a nominal significance threshold of p-value < 0.05, biological pathway analysis was performed. Three of the top enriched canonical pathways, i.e., 'TNF signaling', 'Hepatic fibrosis/Hepatic stellate cell activation', and 'Necroptosis signaling pathway' are inflammatory pathways. Both TNF signaling and the necroptosis signaling pathway have been associated with the progression of COPD [22, 23]. Both the latter mentioned signaling pathways were significantly inhibited in our biological pathway analysis. This appears to support the association of the sentinel SNV with decreased risk of COPD as emerging from our Phe-WAS analysis.

After using a bioinformatic approach to identify the most likely regulatory regions underlying the association, we used a targeted CRISPR approach in an airway epithelial cell system to explore differential gene expression patterns caused by deletions of the key regions. The level of perturbation in terms of differential expression seen after CRISPR deletion of putative regulatory regions indicates that our deleted regions harbor active enhancers. Interestingly, the DEG datasets derived from the two genomic deletions show a remarkable similarity in terms of both identity of differentially expressed genes and also in their direction of fold change. This suggests that the two putative enhancer regions are co-acting enhancers regulating similar or overlapping gene expression patterns.

Interestingly, the expression levels of the two closest genes to this region, HDAC4 and TWIST2, were not altered significantly in cells where these putative enhancers had been deleted. This is perhaps surprising given that one of the genetic variants in this region, rs34177861, is a known eQTL for TWIST2, albeit in lung tissue rather than epithelial cells per se. However, we did identify a large number of other DEGs, indicating this region is actively involved in gene regulation. The observation that there was marked overlap between the DEGs seen following both CRISPR-targeted deletions and the lack of identification of DEGs in the control experiments is strong evidence that the observed changes in gene expression are due to altered regulation driven by these two regulatory regions rather than off-target effects of the CRISPR approach.

Biological pathway analysis of the gene expression changes caused by the two deletions also shows significant similarity. The top six enriched canonical pathways for both deletions show four canonical pathways in common: 'Atherosclerosis signaling', 'Hepatic fibrosis/Hepatic stellate cell activation', 'Axonal

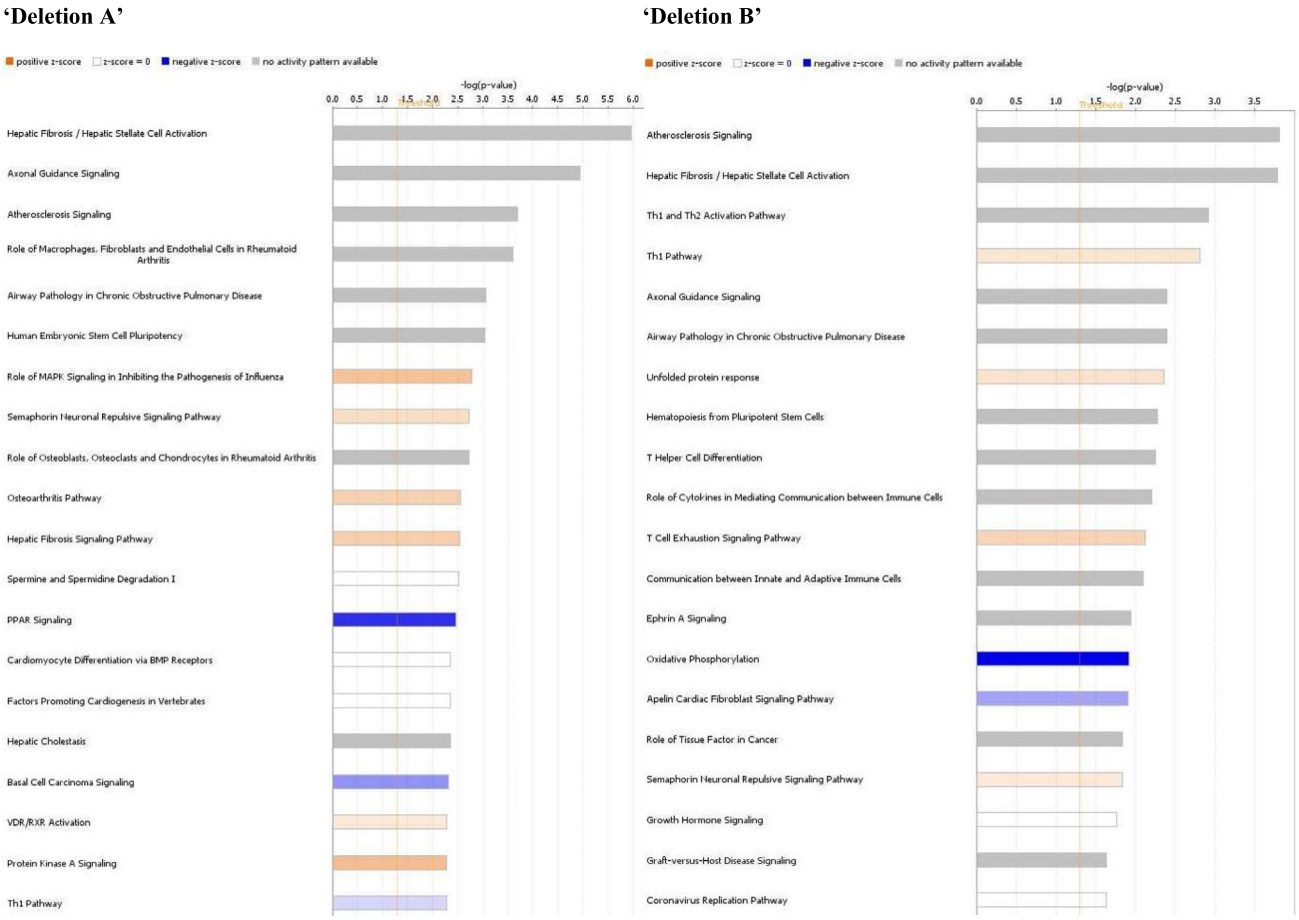

FIGURE 7 | Canonical pathways for the 'Deletion A' and 'Deletion B' DEG datasets. The bar chart represents significance of gene enrichment based on the *p*-value. The vertical line denotes the significance threshold of −log (*p*-value) 1.3 equivalent to *p*-value < 0.05. The bar chart color intensity shows predicted directionality of change; blue scale for negative *Z*-score (deactivation) and red scale for positive *Z*-score (activation). Gray color bar charts indicate no activity pattern available in Ingenuity knowledge database. Data generated with IPA core analysis based on DEGs with statistical significance threshold of $\log_2$ fold change < −1 or > +1 and corresponding *q* value (*p*FDR) < 0.05.

guidance signaling', and 'Airway pathology in COPD'. All the top canonical pathways are potentially involved in inflammatory and immunological responses. Hepatic fibrosis, atherosclerosis, as well as other top canonical pathways including rheumatoid arthritis, and Th1 and Th2 activation pathways have been associated to a systemic inflammatory phenotype [24–26]. In the case of 'Axonal guidance signaling', neural guidance proteins have been reported to play a role in the immune system [27]. The 'hepatic fibrosis/hepatic stellate cell activation' pathway was also enriched for in our protein association analysis. Thus, enrichment for this pathway following the two deletions can be considered as further evidence of the involvement of the hepatic fibrosis pathway.

Our DEG analysis returned 'Airway pathology in COPD' as one of the top canonical pathways for both deletions. The following DEGs overlapped with 'Airway pathology in COPD': *CXCL2, CXCL8, IL6, CXCL3, IL1A, IL1B, L1F, FGF2, IL18, CCL20, LIF, TGFB1, MMP1*. A key role in airway inflammation has been described in the literature for the latter mentioned DEGs. *CXCL2* expression correlates with airway neutrophilic recruitment and inflammation [28]. CXCL2 binds with CXCR2, which is in turn markedly upregulated in airway epithelial cells of COPD patients, leading to neutrophil infiltration and airway inflammation [29]. Similar to CXCL2, CXCL8 and CXCL3 are involved in airway neutrophilic recruitment in COPD through the CXCLs-CXCR2 axis [29–31]. *CXCL8* increase is also reported to be closely associated with simultaneous increase of IL6 through oxidant mechanisms that include activation of transcription factor NF-κB [31]. As for IL1A, IL1B, TGF β1, FGF2, and IL18, mounting evidence also indicates a central role in airway inflammation, with increase in transcription correlating with COPD pathogenesis [32–35]. Of significance to our study, the majority of mentioned DEGs (i.e., *CXCL2, CXCL8, IL6, CXCL3, IL1A, IL1B, L1F, FGF2, LIF, and TGFB1*) exhibited an increase in transcription (fold change range of 4 to 17) across both deletions, thus correlating with a heightened inflammatory response.

The identification of the 'atherosclerosis signaling' pathway in our DEG analysis is also of interest. Atherosclerosis is a chronic and progressive inflammatory disease of the intimal layer of arteries, arising from an imbalance in lipid metabolism and maladaptive immune response [36–38]. The following DEGs overlapped with 'atherosclerosis signaling': *CLU*, *CXCL8*, *F3*, *IL18*, *IL1α*, *IL1β*, *IL6*, *MMP1*, *PDGFA*, *PDGFC*, *PLA2G4A*, *PLA2G7*, *PLA2R1, TGFB1, TNFRSF14, and CD40*. The role of *IL1α, IL1β, IL6, CXCL8, and IL18* in the sustained inflammatory process within the atherosclerotic plaque is well documented

TABLE 4 | Top six canonical pathways and corresponding overlapping DEGs for (A) 'Deletion A' and (B) 'Deletion B' DEG datasets.

Deletion A Ingenuity canonical pathways	−log (*p*-value)	Ratio	*Z*-score	DEGs
Hepatic fibrosis/hepatic stellate cell activation	5.95	0.17	NA	CD14, COL22A1, COL4A1, COL4A2, COL5A1, CXCL3, CXCL8, ECE1, EDNRB, FGF2, FLT4, IGFBP3, IGFBP4, IL10RA, IL1A, IL1B, IL6, MMP1, PDGFA, PDGFC, PDGFRB, SMAD3, TGFA
Axonal guidance signaling	4.94	0.118	NA	ADAM19, ADAMTS9, ARHGEF6, BMP2, BMP5, EFNA1, EPHA4, EPHB1, FYN, FZD3, ITGB4, LIMK2, MMP1, NFATC3, NRP1, NRP2, NTN1, NTNG2, PDGFA, PDGFC, PIK3R1, PLXND1, PPP3CA, PRKCB, PRKD3, ROBO1, SDCBP, SEMA3A, SEMA3D, SEMA3E, SEMA5A, SEMA6A, SHH, TUBB2B, WIPF1, WNT5A
Atherosclerosis signaling	3.69	0.173	NA	CLU, CXCL8, F3, IL18, IL1A, IL1B, IL6, MMP1, PDGFA, PDGFC, PLA2G4A, PLA2G7, PLA2R1
Role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis	3.59	0.118	NA	C5AR1, CAMK2B, CREB5, CXCL8, DKK3, FGF2, FZD3, IL18, IL1A, IL1B, IL6, JUN, MMP1, MYC, MYD88, NFATC3, PDGFA, PDGFC, PIK3R1, PPP3CA, PRKCB, PRKD3, TCF4, TCF7L2, WNT5A
Airway pathology in chronic obstructive pulmonary disease	3.07	0.179	NA	CXCL2, CCL20, CXCL3, CXCL8, FGF2, IL18, IL1A, IL1B, IL6, LIF, MMP1
Human embryonic stem cell pluripotency	3.05	0.149	NA	BMP2, BMP5, FGF2, FZD3, PDGFA, PDGFC, PDGFRB, PIK3R1, SMAD3, SMAD6, TCF4, TCF7L2, WNT5A

Deletion B Ingenuity canonical pathways	−log (*p*-value)	Ratio	*Z*-score	DEGs
Atherosclerosis signaling	3.8	0.16	NA	CD40, CLU, CXCL8, IL18, IL1A, IL6, ITGB2, PDGFA, PLA2G7, PLA2R1, TGFB1, TNFRSF14
Hepatic fibrosis/hepatic stellate cell activation	3.79	0.126	NA	CCN2, CD40, COL22A1, COL4A4, COL5A1, COL6A1, COL8A1, CXCL8, EDN1, FGF2, FLT4, IL10RA, IL1A, IL6, PDGFA, TGFB1, TGFBR1
Th1 and Th2 activation pathway	2.92	0.123	NA	CD274, CD40, HLA-B, IL10RA, IL18, IL6, ITGB2, JAG1, NFATC4, NFIL3, PIK3R1, TGFB1, TGFBR1
Th1 pathway	2.82	0.141	0.378	CD274, CD40, HLA-B, IL10RA, IL18, IL6, ITGB2, NFATC4, NFIL3, PIK3R1
Axonal guidance signaling	2.39	0.082	NA	ADAM19, ADAM8, ADAMTS9, ARHGEF6, BMP2, EFNA1, EPHA2, EPHB1, FYN, ITGB2, NFATC4, NRP1, NRP2, NTN1, PDGFA, PIK3R1, PLCD4, PLXNB1, PRKAR2B, PRKCB, SEMA3A, SEMA3C, SEMA5A, SHH, TUBB4A
Airway pathology in chronic obstructive pulmonary disease	2.39	0.143	NA	CXCL2, CCL20, CXCL8, FGF2, IL18, IL1A, IL6, LIF, TGFB1

Note: Table includes the *p*-value of the correlation, the ratio of overlap of DEGs in dataset with number of genes making up the canonical pathway as in Ingenuity's knowledge database, and *Z*-score where applicable. Data generated with IPA core analysis based on DEGs with statistical significance threshold of $\log_2$ fold change < -1 or $> +1$ and corresponding *q* value (*p*FDR) < 0.05.

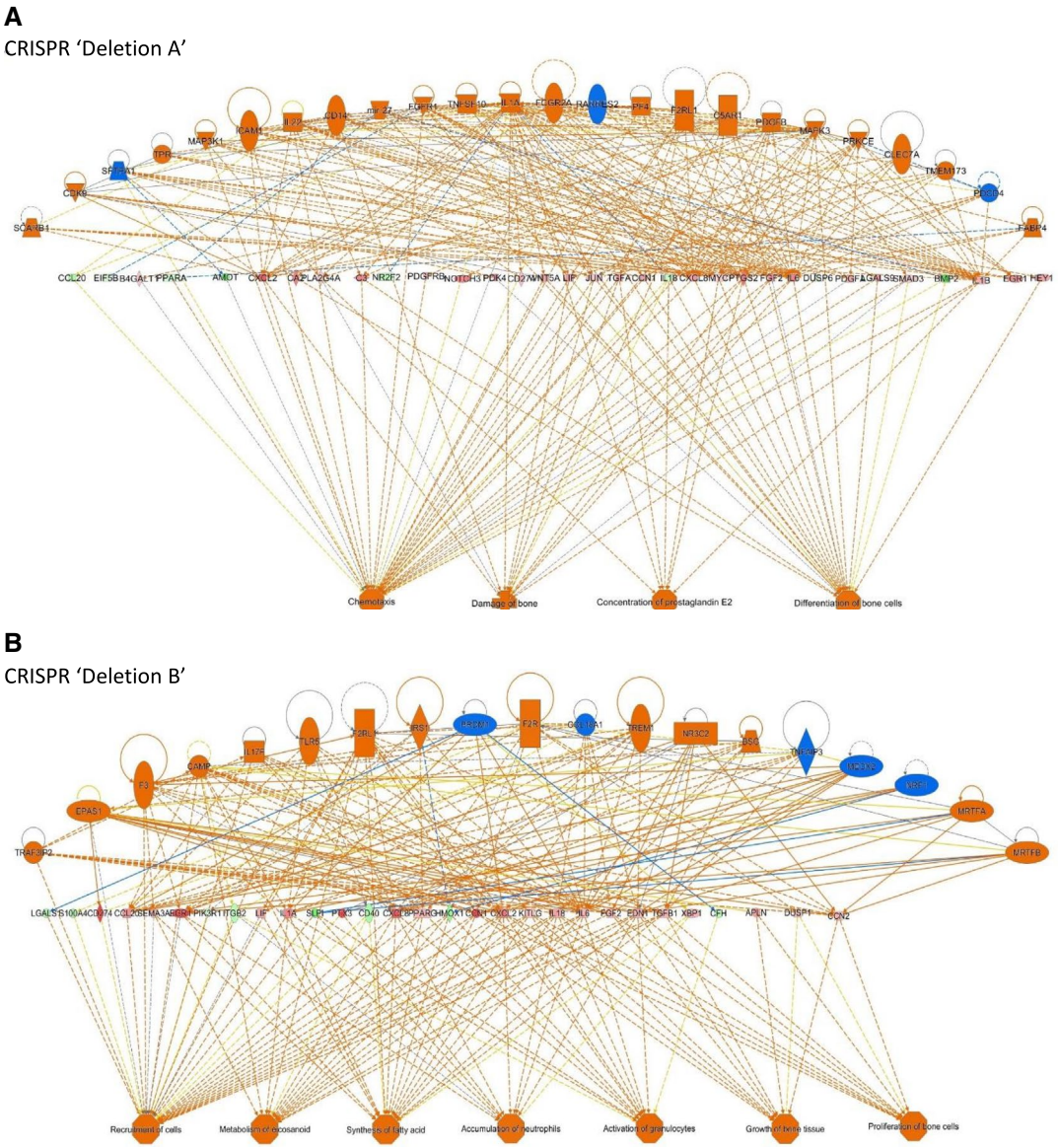

FIGURE 8 | Downstream regulators and biological processes/diseases of (A) 'Deletion A' and (B) 'Deletion B' DEG datasets. Analysis integrates upstream regulators with downstream regulators and predicts the expected biological functions/diseases. Upstream regulators: Blue for 'inhibited' and orange for 'activated'. Downstream regulators: Red scale for 'Upregulated' and green scale for 'Downregulated'. The downstream effect represented by the connecting lines: Orange for 'Leads to activation', blue for 'Leads to Inhibition', and yellow for 'inconsistent findings with downstream node prediction/expression. Data generated with IPA core analysis based on DEGs with statistical significance threshold of $\log_2$ fold change < -1 or $> +1$ and corresponding q value (pFDR) < 0.05.

in the literature [39–46]. Of note is that for all mentioned pro-inflammatory genes in our atherosclerosis dataset, except for IL18, the direction of fold change correlates with a heightened inflammatory response typical of the atherosclerosis process. Similarly, for the blood coagulation genes *F3, PDGFA*, and *PDGF*, which are typically highly expressed in the developing atherosclerotic plaque [47–49], the direction of fold change as recorded in our RNA-Seq analysis correlates with a pro-atherogenic phenotype.

Our results emerging from canonical pathway analysis suggest that genomic region tagged by SNV rs12477314 constitutes a regulatory region responsible for regulating biological pathways conducive to a systemic inflammatory phenotype. Regulator effects analysis of our DEG datasets also yielded results which aligned with this pro-inflammatory phenotype hypothesis. For 'Deletion A', the cluster of networks with the highest consistency score included 'concentration of prostaglandin E2' and 'chemotaxis'. Chemotaxis of leukocytes and lymphocytes is a main biological process contributing to inflammatory disease including COPD, atherosclerosis and rheumatoid arthritis [50–52]. 'Deletion B' showed a remarkable similarity in downstream biological processes with the highest consistency score networks including 'accumulation of neutrophils', 'recruitment of cells', 'activation of granulocytes' and 'metabolism of eicosanoid', all of which are synonymous with a pro-inflammatory phenotype.

The airway epithelium plays a key role in regulating airway responses in both health and disease. Our targeted deletion of the enhancer regions using CRISPR/Cas9 attempts to

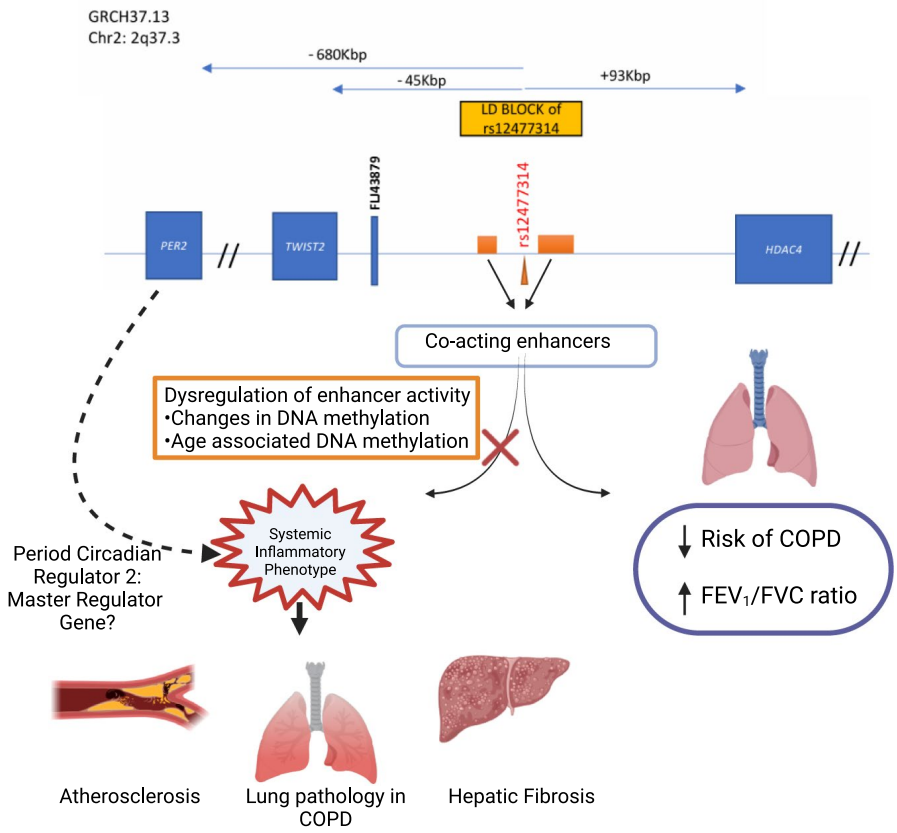

FIGURE 9 | Hypothetical mechanisms underlying the observed association signal between variants at the rs12477314 locus and lung function. Created in BioRender. https://BioRender.com.

experimentally simulate the dysregulation of the enhancer regions under investigation. In vivo, organismal aging and consequent increased risk of multiple age-related diseases have been reported to strongly correlate with changes in dynamic DNA methylation [53, 54]. Global hypomethylation and region-specific hypermethylation are two main general mechanisms resulting in age-dependent changes in DNA methylation [55]. In turn, age-dependent DNA methylation can also be modified by extrinsic risk factors such as smoking and obesity [56]. Interestingly, the most significant canonical pathways emerging from our study, i.e., related to COPD, atherosclerosis, hepatic fibrosis, and rheumatoid arthritis, are all considered common diseases of aging [57–60]. The specific role of age-associated DNA methylation changes in the pathological process of COPD and in abnormal tissue repair in COPD is, however, to date unclear due to a lack of solid data [58]. The term inflammaging has been coined to describe chronic and systemic low-grade inflammation that develops with age and correlates with biological aging and worsening of age-related diseases. This process has been described by the age-induced activation of the innate immune system resulting in increased levels of pro-inflammatory cytokines, with concomitant decline of the adaptive immune system, leading to the development of low-grade systemic inflammation [61].

The extent of perturbation in gene expression as observed in our DEG datasets indicates that the majority of gene dysregulation is likely due to indirect secondary effects of the deletion. In line with the cis-acting tendency of enhancers, chromosome two resulted as the most dysregulated chromosome in terms of the number of DEGs. As discussed above, the genes immediately proximal to the deleted regions, i.e., *HDAC4* (downstream) and *FLJ43879* and *TWIST2* (upstream) were not dysregulated. The closest DEG to the putative enhancers is *PER2* located on 2q37.3 at circa 680 Kb upstream of 'Deletion A'. *PER2* is transcriptionally upregulated in both our DEG datasets with fold changes of 4.9 for 'Deletion A' and 2.8 for 'Deletion B'. *PER2* is a transcriptional repressor that forms a core component of the circadian clock and whose dysregulation has been implicated in various human pathologies including cancer, cardiovascular disease, and COPD and lung function decline [62–66]. We postulate *PER2* as a potential master regulator that could explain at least in part the differential gene expression reported in our study.

In conclusion, our study suggests that the observed association seen in GWAS of lung function/COPD for the region tagged by rs12477314 may be due to altered gene expression patterns as a consequence of regulation by 2 enhancer regions at this locus.

Author Contributions

Anthony G. Fenech conceived the study. Anthony G. Fenech, Ian P. Hall, and Roger Ellul-Micallef designed and directed research. Godwin M. Grech performed the research. Godwin M. Grech, Ian P. Hall, and Anthony G. Fenech analyzed and interpreted the data. Katherine A. Fawcett, Robert J. Hall, and Godfrey Grech contributed data analysis or analysis tools. Godwin M. Grech wrote the first draft of the paper. Anthony G. Fenech, Ian P. Hall, and Roger Ellul-Micallef edited and reviewed the paper. All authors read and approved the paper.

Acknowledgments

Ian P. Hall holds an NIHR Senior Investigator award and is supported by a Wellcome Trust Discovery Award. This study was in part funded by the Malta Government Scholarship Scheme (MGSS).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

RNAseq datasets that support the findings of this study are available in the Supporting Information of this article. The proteomic association datasets can be accessed through UK Biobank.

References

1. D. B. Hancock, M. Eijgelsheim, J. B. Wilk, et al., "Meta-Analyses of Genome-Wide Association Studies Identify Multiple Novel Loci Associated With Pulmonary Function," *Nature Genetics* 42, no. 1 (2010): 45–52.

2. B. D. Hobbs, K. de Jong, M. Lamontagne, et al., "Genetic Loci Associated With Chronic Obstructive Pulmonary Disease Overlap With Loci for Lung Function and Pulmonary Fibrosis," *Nature Genetics* 49, no. 3 (2017): 426–432.

3. E. Repapi, I. Sayers, L. V. Wain, et al., "Genome-Wide Association Study Identifies Five Loci Associated With Lung Function," *Nature Genetics* 42, no. 1 (2010): 36–42.

4. P. Sakornsakolpat, D. Prokopenko, M. Lamontagne, et al., "Genetic Landscape of Chronic Obstructive Pulmonary Disease Identifies Heterogeneous Cell-Type and Phenotype Associations," *Nature Genetics* 51, no. 3 (2019): 494–505.

5. N. Shrine, A. L. Guyatt, A. M. Erzurumluoglu, et al., "New Genetic Signals for Lung Function Highlight Pathways and Chronic Obstructive Pulmonary Disease Associations Across Multiple Ancestries," *Nature Genetics* 51, no. 3 (2019): 481–493.

6. M. Soler Artigas, D. W. Loth, L. V. Wain, et al., "Genome-Wide Association and Large-Scale Follow Up Identifies 16 New Loci Influencing Lung Function," *Nature Genetics* 2011, no. 43 (2011): 1082–1090.

7. M. Soler Artigas, L. V. Wain, S. Miller, et al., "Sixteen New Lung Function Signals Identified Through 1000 Genomes Project Reference Panel Imputation," *Nature Communications* 6 (2015): 8658.

8. L. V. Wain, N. Shrine, S. Miller, et al., "Novel Insights Into the Genetics of Smoking Behaviour, Lung Function, and Chronic Obstructive Pulmonary Disease (UK BiLEVE): A Genetic Association Study in UK Biobank," *Lancet Respiratory Medicine* 3 (2015): 769–781.

9. I. Sayers, C. John, J. Chen, and I. P. Hall, "Genetics of Chronic Respiratory Disease," *Nature Reviews Genetics* 25 (2024): 534–547.

10. N. Shrine, A. G. Izquierdo, J. Chen, et al., "Multi-Ancestry Genome-Wide Association Analyses Improve Resolution of Genes and Pathways Influencing Lung Function and Chronic Obstructive Pulmonary Disease Risk," *Nature Genetics* 55 (2023): 410–422.

11. M. J. Machiela and S. J. Chanock, "LDlink a Web-Based Application for Exploring Population-Specific Haplotype Structure and Linking Correlated Alleles of Possible Functional Variants," *Bioinformatics* 31, no. 21 (2015): 3555–3557.

12. H. Thorvalsdottir, J. T. Robinson, and J. P. Mesirov, "Integrative Genomics Viewer (IGV) High Performance Genomics Data Visualization and Exploration," *Briefings in Bioinformatics* 14, no. 2 (2013): 178–192, https://doi.org/10.1093/bib/bbs017.

13. R. J. Packer, A. T. Williams, W. Hennah, et al., "DeepPheWAS: An R Package for Phenotype Generation and Association Analysis for Phenome-Wide Association Studies," *Bioinformatics* 39 (2023): 4.

14. V. P. Dubois, D. Zotova, K. M. Parkins, et al., "Safe Harbor Targeted CRISPR-Cas9 Tools for Molecular Imaging of Cells in Living Subjects," *CRISPR Journal* 1 (2018): 440–449.

15. T. Rosner, S. Kahle, F. Montenegro, et al., "Immune Effector Functions of Human IgG2 Antibodies Against EGFR," *Molecular Cancer Therapeutics* 18, no. 1 (2019): 75–88, https://doi.org/10.1158/1535-7163.MCT-18-0341.

16. M. A. Wheeler, I. C. Clark, E. C. Tjon, et al., "MAFG-Driven Astrocytes Promote CNS Inflammation," *Nature* 578 (2020): 593–599.

17. A. Krämer, J. Green, J. Pollard, Jr., and S. Tugendreich, "Causal Analysis Approaches in Ingenuity Pathway Analysis," *Bioinformatics* 30 (2014): 523–530.

18. E. A. Oczypok, T. N. Perkins, and T. D. Oury, "All the "RAGE" in Lung Disease: The Receptor for Advanced Glycation Endproducts (RAGE) is a Major Mediator of Pulmonary Inflammatory Responses," *Paediatric Respiratory Reviews* 23 (2017): 40–49.

19. K. Waseda, N. Miyahara, A. Taniguchi, et al., "Emphysema Requires the Receptor for Advanced Glycation End-Products Triggering on Structural Cells," *American Journal of Respiratory Cell and Molecular Biology* 52, no. 4 (2015): 482–491.

20. L. Wu, L. Ma, L. F. Nicholson, and P. N. Black, "Advanced Glycation End Products and Its Receptor (RAGE) Are Increased in Patients With COPD," *Respiratory Medicine* 105, no. 3 (2011): 329–336.

21. M. A. Queisser, F. M. Kouri, M. Konigshoff, et al., "Loss of RAGE in Pulmonary Fibrosis: Molecular Relations to Functional Changes in Pulmonary Cell Types," *American Journal of Respiratory Cell and Molecular Biology* 39, no. 3 (2008): 337–345.

22. C. Liu, P. Li, J. Zheng, Y. Wang, W. Wu, and X. Liu, "Role of Necroptosis in Airflow Limitation in Chronic Obstructive Pulmonary Disease: Focus on Small-Airway Disease and Emphysema," *Cell Death Discovery* 8 (2022): 363.

23. Y. Yao, J. Zhou, X. Diao, and S. Wang, "Association Between Tumor Necrosis Factor-Alpha and Chronic Obstructive Pulmonary Disease: A Systematic Review and Meta-Analysis," *Therapeutic Advances in Respiratory Disease* 13 (2019): 1753466619866096.

24. A. Arida, A. D. Protogerou, G. D. Kitas, and P. P. Sfikakis, "Systemic Inflammatory Response and Atherosclerosis: The Paradigm of Chronic Inflammatory Rheumatic Disease," *International Journal of Molecular Sciences* 19 (2018): 1890.

25. M. Dirchwolf and A. E. Ruf, "Role of Systemic Inflammation in Cirrhosis: From Pathogenesis to Prognosis," *World Journal of Hepatology* 7, no. 16 (2015): 1974–1981.

26. E. J. D. Oudijk, J. W. J. Lammers, and L. Koenderman, "Systemic Inflammation in Chronic Obstructive Pulmonary Disease," *European Respiratory Journal* 22, no. 46 (2003): 5s–13s.

27. V. Mirakaj and P. Rosenberger, "Immunomodulatory Functions of Neuronal Guidance Proteins," *Trends in Immunology* 38, no. 6 (2017): 444–456.

28. N. Tiwari, A. S. Marudamuthu, Y. Tsukasaki, M. Ikebe, J. Fu, and S. Shetty, "p53- and PAI-1-Mediated Induction of C-X-C Chemokines and CXCR2: Importance in Pulmonary Inflammation due to Cigarette Smoke Exposure," *American Journal of Physiology. Lung Cellular and Molecular Physiology* 310, no. 6 (2016): L496–L506.

29. Y. Cheng, X. Ma, Y. Wei, and X. Wei, "Potential Roles and Targeted Therapy of the CXCLs/CXCR2 Axis in Cancer and Inflammatory Diseases," *BBA—Reviews on Cancer* 1871 (2019): 289–312.

30. L. A. Al-Alwan, Y. Chang, A. Mogas, et al., "Differential Roles of CXCL2 and CXCL3 and Their Receptors in Regulating Normal and Asthmatic Airway Smooth Muscle Cell Migration," *Journal of Immunology* 191, no. 5 (2013): 2731–2741.

31. J. Zhang and C. Bai, "The Significance of Serum Interleukin-8 in Acute Exacerbations of Chronic Obstructive Pulmonary Disease," *Tanaffos* 17, no. 1 (2018): 13–21.

32. E. Dima, O. Koltsida, P. Katsaounou, et al., "Implication of Interleukin (IL)-18 in the Pathogenesis of Chronic Obstructive Pulmonary Disease (COPD)," *Cytokine* 74, no. 2 (2015): 313–317.

33. C. A. Ojiaku, E. J. Yoo, and R. A. Panettieri, Jr., "Transforming Growth Factor β1 Function in Airway Remodeling and Hyperresponsiveness. The Missing Link?," *American Journal of Respiratory Cell and Molecular Biology* 56, no. 4 (2017): 432–442.

34. E. T. Osei, C.-A. Brandsma, W. Timens, I. H. Heijink, and T.-L. Hackett, "Current Perspectives on the Role of Interleukin-1 Signaling in the Pathogenesis of Asthma and COPD," *European Respiratory Journal* 55 (2019): 1900563, https://doi.org/10.1183/13993003.00563-2019.

35. Y. Tan, Y. Qiao, Z. Chen, et al., "FGF2, an Immunomodulatory Factor in Asthma and Chronic Obstructive Pulmonary Disease (COPD)," *Frontiers in Cell and Developmental Biology* 8 (2020): 223.

36. M. Aziz and K. S. Yadav, "Pathogenesis of Atherosclerosis," *Medical Clinics of North America* 2 (2016): 3.

37. M. Back, A. Yurdagul, I. Tabas, K. Oorni, and P. Kovanen, "Inflammation and Its Resolution in Atherosclerosis: Mediators and Therapeutic Options," *Nature Reviews Cardiology* 16 (2019): 389–406.

38. K. Moore, F. Sheedy, and E. Fisher, "Macrophages in Atherosclerosis: A Dynamic Balance," *Nature Reviews Immunology* 13, no. 10 (2013): 709–721.

39. S. Apostolakis, K. Vogiatzi, V. Amantidou, and D. A. Spandidos, "Interleukin 8 and Cardiovascular Disease," *Cardiovascular Research* 3, no. 1 (2009): 353–360.

40. S. A. Huber, P. Sakkinen, D. Conze, N. Hardin, and R. Tracy, "Interleukin-6 Exacerbates Early Atherosclerosis in Mice," *Arteriosclerosis, Thrombosis, and Vascular Biology* 19 (1999): 2364–2367.

41. Y. Kamari, A. Shaish, S. Shemesh, et al., "Reduced Atherosclerosis and Inflammatory Cytokines in Apolipoprotein-Edeficient Mice Lacking Bone Marrow-Derived Interleukin-1alpha," *Biochemical and Biophysical Research Communications* 405 (2011): 197–203.

42. H. Kirii, T. Niwa, Y. Yamada, et al., "Lack of Interleukin-1beta Decreases the Severity of Atherosclerosis in ApoE-Deficient Mice," *Arteriosclerosis, Thrombosis, and Vascular Biology* 23 (2003): 656–660.

43. Z. Mallat, A. Corbaz, A. Scoazec, et al., "Expression of Interleukin-18 in Human Atherosclerotic Plaques and Relation to Plaque Instability," *Circulation* 104 (2001): 1598–1603.

44. Z. Mallat, A. Corbaz, A. Scoazec, et al., "Interleukin-18/Interleukin-18 Binding Protein Signaling Modulates Atherosclerotic Lesion Development and Stability," *Circulation Research* 89 (2001): E41–E45.

45. H. Schuett, R. Oestreich, G. H. Waetzig, et al., "Transsignaling of Interleukin-6 Crucially Contributes to Atherosclerosis in Mice," *Arteriosclerosis, Thrombosis, and Vascular Biology* 32 (2012): 281–290.

46. S. C. Whitman, P. Ravisankar, and A. Daugherty, "Interleukin-18 Enhances Atherosclerosis in Apolipoprotein $E^{-/-}$ Mice Through Release of Interferon-Gamma," *Circulation Research* 90 (2002): E34–E38.

47. E. Folestad, A. Kunath, and D. Wagsater, "PDGF-C and PDGF-D Signaling in Vascular Disease and Animal Models," *Molecular Aspects of Medicine* 62 (2018): 1–11.

48. P. A. Owens and N. Mackman, "Role of Tissue Factor in Atherothrombosis," *Current Atherosclerosis Reports* 14, no. 5 (2012): 394–401.

49. Y. Zhang, W. Zhang, K. Wang, T. Li, S. Song, and B. Yuan, "Expression of Platelet-Derived Growth Factor in the Vascular Walls of Patients With Lower Extremity Arterial Occlusive Disease," *Experimental and Therapeutic Medicine* 9 (2015): 1223–1228.

50. J. Li and K. Ley, "Lymphocyte Migration Into Atherosclerotic Plaque," *Arteriosclerosis, Thrombosis, and Vascular Biology* 35, no. 1 (2015): 40–49.

51. J. K. Quint and J. A. Wedzicha, "The Neutrophil in Chronic Obstructive Pulmonary Disease," *Journal of Allergy and Clinical Immunology* 119, no. 5 (2007): 1065–1071.

52. A. K. Rana, Y. Li, Q. Dang, and F. Yang, "Monocytes in Rheumatoid Arthritis: Circulating Precursors of Macrophages and Osteoclasts and, Their Heterogeneity and Plasticity Role in RA Pathogenesis," *International Immunopharmacology* 65 (2018): 348–359.

53. A. A. Johnson, K. Akman, S. R. Calimport, D. Wuttke, A. Stolzing, and J. P. de Magalhães, "The Role of DNA Methylation in Aging, Rejuvenation, and Age-Related Disease," *Rejuvenation Research* 15, no. 5 (2012): 483–494.

54. F. H. Xiao, H. T. Wang, and Q. P. Kong, "Dynamic DNA Methylation During Aging: A "Prophet" of Age-Related Outcomes," *Frontiers in Genetics* 10 (2019): 107.

55. F. H. Xiao, Q. P. Kong, B. Perry, and Y. H. He, "Progress on the Role of DNA Methylation in Aging and Longevity," *Briefings in Functional Genomics* 15 (2016): 454–459.

56. P. A. Dugué, J. K. Bassett, J. E. Joo, et al., "Association of DNA Methylation-Based Biological Age With Health Risk Factors and Overall and Cause-Specific Mortality," *American Journal of Epidemiology* 187, no. 3 (2018): 529–538.

57. A. M. Boots, A. B. Maier, P. Stinissen, P. Masson, R. J. Lories, and F. De Keyser, "The Influence of Ageing on the Development and Management of Rheumatoid Arthritis," *Nature Reviews Rheumatology* 9, no. 10 (2013): 604–613.

58. C.-A. Brandsma, M. de Vries, R. Costa, R. R. Woldhuis, M. Konigshoff, and W. Timens, "Lung Ageing and COPD: Is There a Role for Ageing in Abnormal Tissue Repair?," *European Respiratory Review* 26 (2017): 170073.

59. F. Sheedfar, S. Di Biase, D. Koonen, and M. Vinciguerra, "Liver Diseases and Aging: Friends or Foes?," *Aging Cell* 12 (2013): 950–954.

60. J. C. Wang and M. Bennett, "Aging and Atherosclerosis. Mechanisms, Functional Consequences, and Potential Therapeutics for Cellular Senescence," *Circulation Research* 111 (2012): 245–259.

61. C. Franceschi, M. Bonafè, S. Valensin, et al., "Inflamm-Aging. An Evolutionary Perspective on Immunosenescence," *Annals of the New York Academy of Sciences* 908, no. 1 (2000): 244–254.

62. K. Krakowiak and H. J. Durrington, "The Role of the Body Clock in Asthma and COPD," *Pulmonary Therapy* 4 (2018): 29–43.

63. I. Schmutz, J. A. Ripperger, S. Baeriswyl-Aebischer, and U. Albrecht, "The Mammalian Clock Component PERIOD2 Coordinates Circadian Output by Interaction With Nuclear Receptors," *Genes & Development* 24 (2009): 345–357.

64. H. Viswambharan, J. M. Carvas, V. Antic, et al., "Mutation of the Circadian Clock Gene PER2 Alters Vascular Vascular Endothelial Function," *Circulation* 115 (2007): 2188–2195.

65. S. Xiaoli, D. Chen, K. Yang, Q. Zhao, D. Zhao, and X. Lv, "The Circadian Clock Gene PER2 Plays an Important Role in Tumor Suppression Through Regulating Tumor-Associated Genes in Human Oral Squamous Cell Carcinoma," *Oncology Reports* 38, no. 1 (2017): 472–480.

66. H. Yao, I. K. Sundar, Y. Huang, et al., "Disruption of Sirtuin 1-Mediated Control of Circadian Molecular Clock and Inflammation in Chronic Obstructive Pulmonary Disease," *American Journal of Respiratory Cell and Molecular Biology* 53, no. 6 (2015): 782–792.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.