Expanded View Figures

Figure EV1. Genome-wide DNA methylation changes occur early in human lung fibroblasts in COPD and progress with disease development.

A Characteristics of the lung tissue donors used in this study for fibroblast isolation. Lung function between COPD and smoker control group (no COPD) is significantly different (*P-value < 0.05; non-parametric unpaired Kruskal–Wallis test).

B Examples of hematoxylin and eosin (H&E) staining of tissue samples from the study cohort: no COPD, COPD (I), COPD (II), COPD (III) and COPD (IV). Scale bars: 0.1 mm.

C, D Validation of the purity of isolated fibroblast by FACS (C) and immunofluorescence (D).

C Left, contour plot of lung live-single cell suspensions from COPD (green) and no COPD (blue) donors gated as indicated above. Right, contour plot of trypsinized fibroblasts from COPD (green) and no COPD (blue) donors. EpCAM-PE was used as an epithelial marker and CD45-Bv605 as an immune marker.

D Immunofluorescence staining of fibroblasts isolated from the lungs of COPD (top) and no COPD (bottom) donors and stained with antibodies against vimentin (VIM, mesenchymal marker) and alpha smooth-muscle actin (αSMA, myofibroblast marker). DAPI was used to counterstain the nuclei. Scale bars: 0.1 mm.

E Read coverage at single CpG sites in each sample.

F, G Histograms showing the number of CpGs per DMR (F) and the DMR size distribution (G). Median values are indicated and highlighted by a vertical line.

H Scatter plot showing correlation of average methylation differences in COPD between DMRs identified in this study and published 450 k methylation array data of parenchymal fibroblasts (Clifford *et al*, 2018). Each dot represents one DMR covered by at least three CpG probes of the array. The blue diagonal represents the linear regression. Shaded areas are confidence intervals of the correlation coefficient at 95%.

I Hierarchical clustering of all samples based on 6,279 DMRs identified between no COPD and COPD (II–IV).

J Functional annotation of genes located next to hypomethylated DMRs using GREAT. Left, hits were sorted according to the binominal P-value and the top 15 hits are shown. The adjusted P-value is indicated by the color code and the number of DMR-associated genes is indicated by the node size. Right, the percentage of DMRs associated to processes shown on the left belonging to clusters 1 (orange) and 3 (purple) (clusters defined in Fig 1H).

K Scatter plots showing correlation of average methylation obtained with T-WGBS and Mass Array from commercial COPD and smoker fibroblasts. Each dot represents one region with varying methylation levels. The blue diagonal represents the linear regression. Shaded areas are confidence intervals of the correlation coefficient at 95%. Correlation coefficients and P-values were calculated by the Pearson correlation method.

Data information: In (A), data points represent values for each donor and horizontal bars the group median. P-values: ** < 0.001, *** < 0.0001. One-way ANOVA non-parametric unpaired test was used (Kruskal–Wallis) within the GraphPad Prism software, correcting for multiple comparisons using Dunn's test. In (H) correlation coefficients and P-values were calculated by the Pearson correlation method. Exact P-values for each test: FEV_1, 0.0006; FEV_1/FVC: 0.0071. BMI, body mass index; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; ESI, emphysema score index.

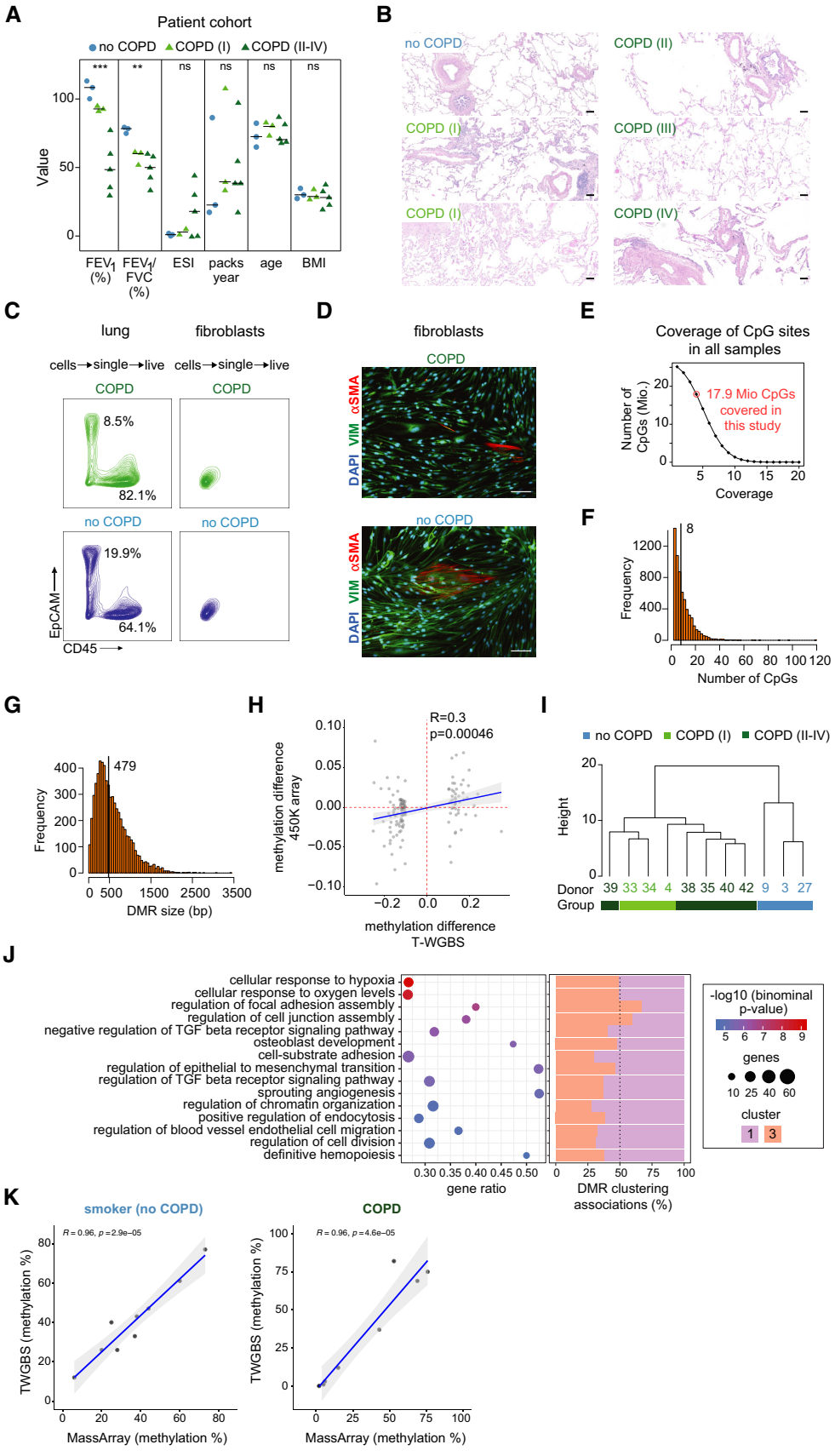

Figure EV1.

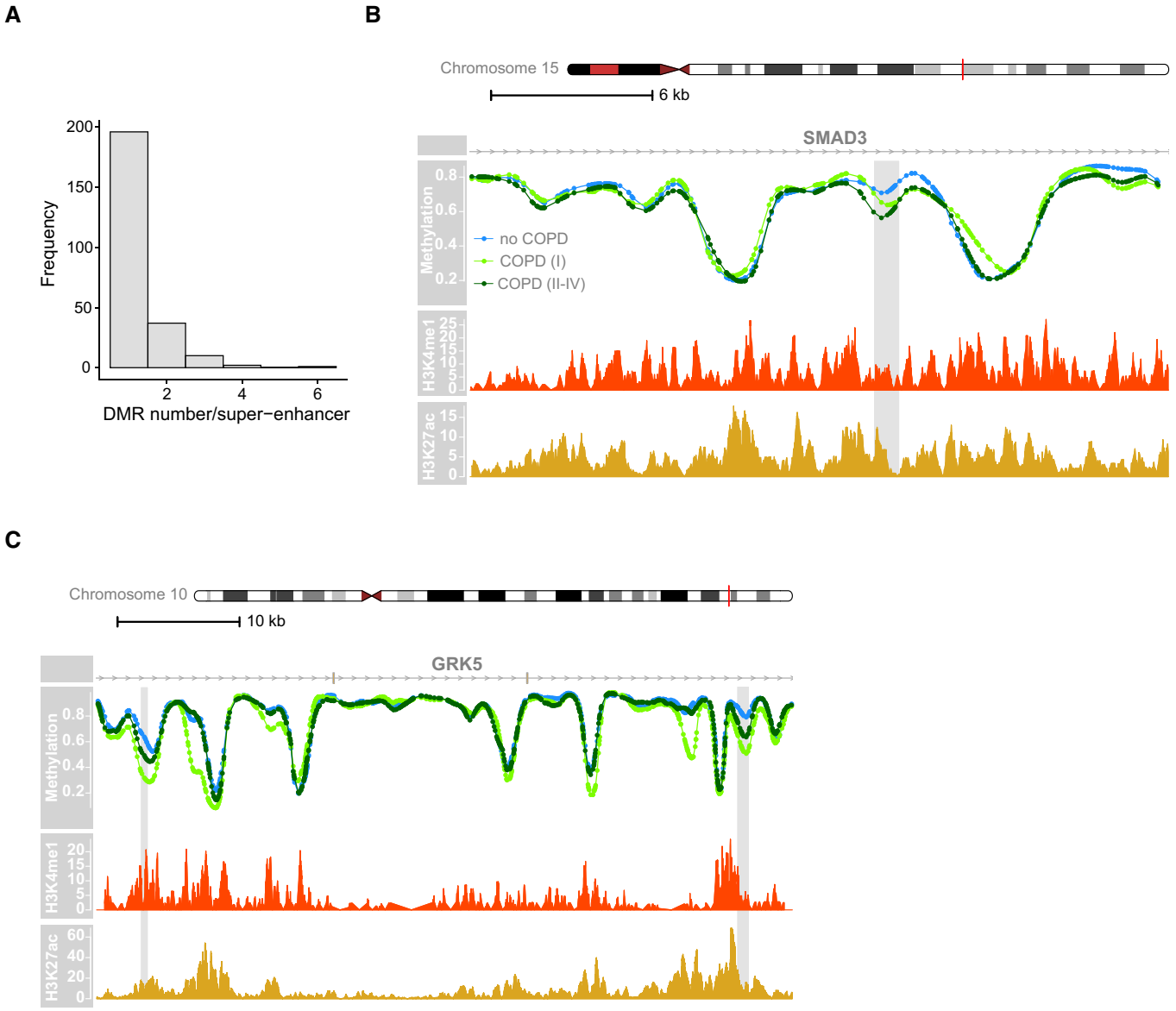

Figure EV2. DNA methylation changes occur at regulatory regions in primary human lung fibroblasts cells during COPD.

A Number of DMRs associated with super-enhancers.

B, C Genome browser views of two super-enhancer regions overlapping with identified DMRs (shaded in gray). Group median CpG methylation is shown for no COPD (blue), COPD (I) (light green), and COPD (II–IV) (dark green). At the bottom, the level of enhancer marks: H3K4me1 (ENCODE accession: ENCFF102BGI; red track) and H3K27ac (ENCODE accession: ENCFF386FDQ; orange track) is depicted as fold change over control.

Figure EV3. DNA methylation changes in primary human lung fibroblasts are accompanied by gene expression changes in COPD.

A Normalized RNA-seq read counts of four exemplary differentially expressed lincRNAs.

B, C Functional annotation of upregulated (B) and downregulated (C) DEGs. Top 5 enriched GO biological process terms are shown on top of each panel. Bottom panels show a more detailed overview of the biological processes enriched for upregulated (B) and downregulated (C) DEGs. Biological processes are connected, if DEGs associated with both processes are in common.

D Gene set enrichment analysis (GSEA) of down/up-regulated genes in COPD. Gene sets of significantly up- (red line) or down-regulated (blue) genes in human lung fibroblasts of COPD patients were compared to transcriptional changes in whole tissue RNA extractions (Kim *et al*, 2015). Genes were sorted and ranked by the expression fold change (log2(COPD/no COPD)) in the Kim *et al* data set. *P*-values were derived from permutation analysis in the GSEA.

E Unsupervised principal component analysis (PCA) of all samples based on the 500 most variable expressed genes.

Data information: In (A), data points represent normalized counts from no COPD (blue) and COPD (II-IV, green). The group mean is shown as a black dot and the error bar represents the standard deviation. FDRs were calculated using DESeq2, which uses a negative binominal GLM (generalized linear model) and Wald statistics. *FDR < 0.05; **FDR < 0.01. Exact FDR-values are as follows: MALAT1, FDR = 0.00116; NEAT-1, FDR = 0.00259; LINC00856, FDR = 0.00291; LINC02057, FDR = 0.01836.
Source data are available online for this figure.

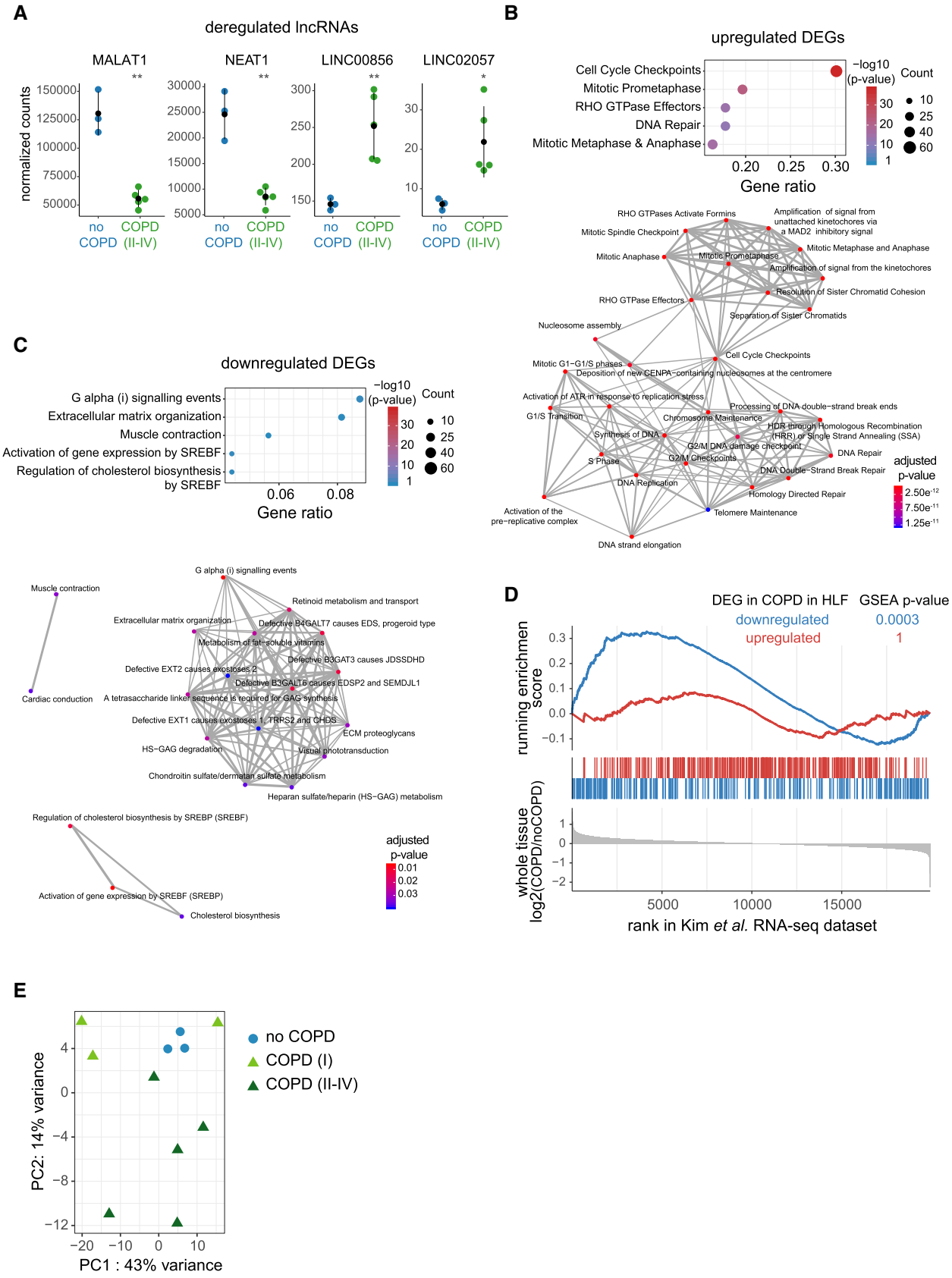

Figure EV3.

Figure EV4. Integrative data analysis reveals epigenetically regulated genes in COPD fibroblasts.

A Representative methylation profiles and DMRs (shaded in gray). Group median CpG methylation is shown for no COPD (blue), COPD (I) (light green), and COPD (II–IV) (dark green). RefSeq annotated genes and CpG islands are indicated.

B Location of hyper- and hypomethylated DMRs relative to the TSS of DEGs in downregulated (left) and upregulated (right) genes.

C Fraction of genes associated with at least one DMR in the proximity ($\pm$ 4 kb) of the TSS are indicated in blue. Genes were split into not significantly changed genes (left bar, FALSE) and DEGs (right bar, TRUE). DMRs are significantly enriched at DEGs.

D Biological processes enriched at DMR associated DEGs. DEGs nodes (colored bubbles) are connected with their associated processes. DEGs and integration clusters are indicated by squares and circles, respectively.

E Examples of genes showing correlation between gene expression and methylation. Scatter plots showing positive (left, AQP3) and negative (right, LPXN) correlation between gene expression and methylation of promoter-associated DMR. Each dot represents an individual donor. Dots are color coded according to disease state. Gene expression is illustrated as normalized counts and methylation is illustrated as average beta value of the corresponding DMR. Linear regression analysis (black line) and 95% confidence interval are shown (gray area).

Data information: In (C), P-value $= 1.3 \times 10^{-7}$ (Fisher's exact test). In (E) correlation coefficients and P-values were calculated by the Pearson correlation method.

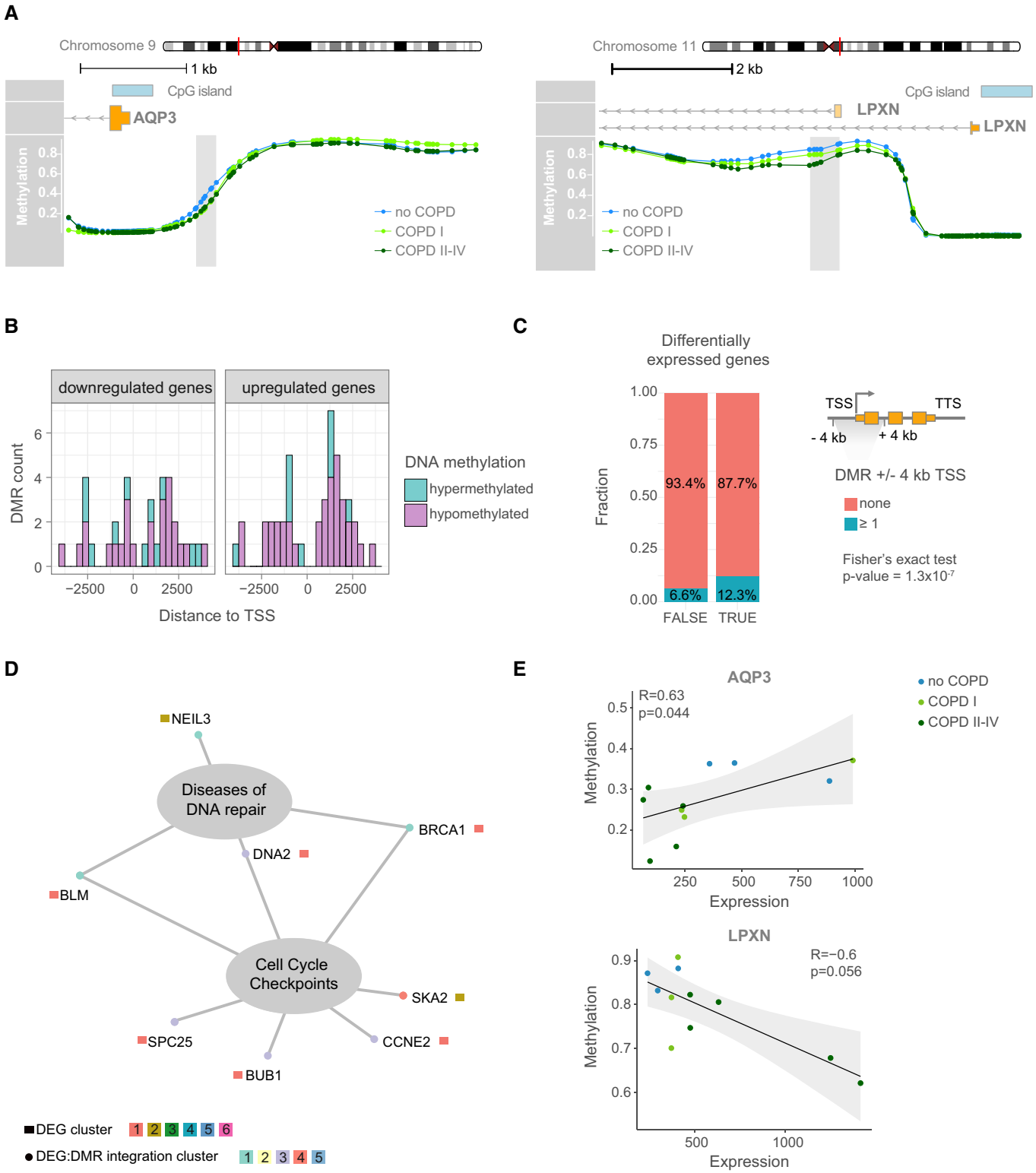

Figure EV4.

Figure EV5. siRNA-based phenotypic screens in normal and COPD primary human lung fibroblasts identify multiple candidate genes regulating COPD phenotypes.

A Distribution of candidates selected for functional validation among different categories (related to Fig 5C). The selected candidates were divided into seven categories: (i) epigenetic players dysregulated on the transcriptional level, (ii) DMR motif includes transcription factors with binding sites overrepresented in DMRs, (iii) upstream regulators identified by the Ingenuity Pathway Analysis, (iv) DMR-DEG correlation represents differentially expressed genes correlated with epigenetic changes in their promoter, (v) top downregulated DEGs, (vi) top upregulated DEGs, and (vii) lncRNA for the top dysregulated long non-coding RNAs. In addition, three assay controls (ACTA2, TGFβR1 and BRD4) were included.

B Scatterplots showing the correlation of the data obtained from two different NHLF (left) and two DHLF (right) donors in the phenotypic screens using col1 and proliferation/nuclei count as readouts.

C Examples of epigenetic hits from the siRNA screen for each indicated readout. Each dot represents a unique candidate tested, blue and red dots represent significant hits (|SSMD values| ≥ 2 are shown in lighter shade and |SSMD values| ≥ 3 in stronger shade). Assay controls are labeled in black and epigenetic factors identified as strong positive hits (|SSMD values| ≥ 3) in fibroblast to myofibroblast transition and cell proliferation processes are labeled in blue or red depending on the effect on each readout upon KD.

D Dot plots showing examples of positive hits with significant differences between NHLFs (2 donors, 2 biological replicates each, shown in green) and DHLFs (3 donors, 2 biological replicates each, shown in orange) in col1 and proliferation/nuclei count readout. Screen readout was normalized to the corresponding non-targeting siRNA control (NTC). TGFβR1 and ACTA2 represent screen controls. The results of all replicates are shown.

E Relative changes in gene expression ($2^{-\Delta\Delta Ct}$) for LPXN, AQP3, GLI4, TGFβR1 and ACTA2 upon gene knockdown using the indicated siRNAs in human lung primary fibroblasts. Gene expression was measured by RT–qPCR using non-target control (NTC) samples as reference and RLP0 (LPXN, AQP3, GLI4) or RNA-polymerase II (TGFβR1, ACTA2) as housekeeping gene. Each dot represents the mean of two technical replicates from at least 3 independent experiments. Bars show median and error bars denote 95% CI of at least three biological replicates ($n > 3$).

Data information: In (B) *P*-value was derived from linear regression analysis and the Pearson correlation coefficient (*R*) is indicated. In (D) statistical evaluation was performed with unpaired two-tailed student's *t*-test. **P*-value < 0.05; ***P*-value < 0.01. Black dots denote the means and error bars represent standard deviation. In (E) Paired *t*-test, FDR corrected using two-stage set-up method of Benjamini, Krieger, and Yekutieli. **P*-value < 0.05; ***P*-value < 0.01; *****P*-value < 0.00001. Exact *q*-values for the specific gene-knockdowns: LPXN, 0.00045; AQP3, 0.000549; GLI4, 0.003146; TGFBR1, 0.000005; ACTA2, < 0.000001.
Source data are available online for this figure.

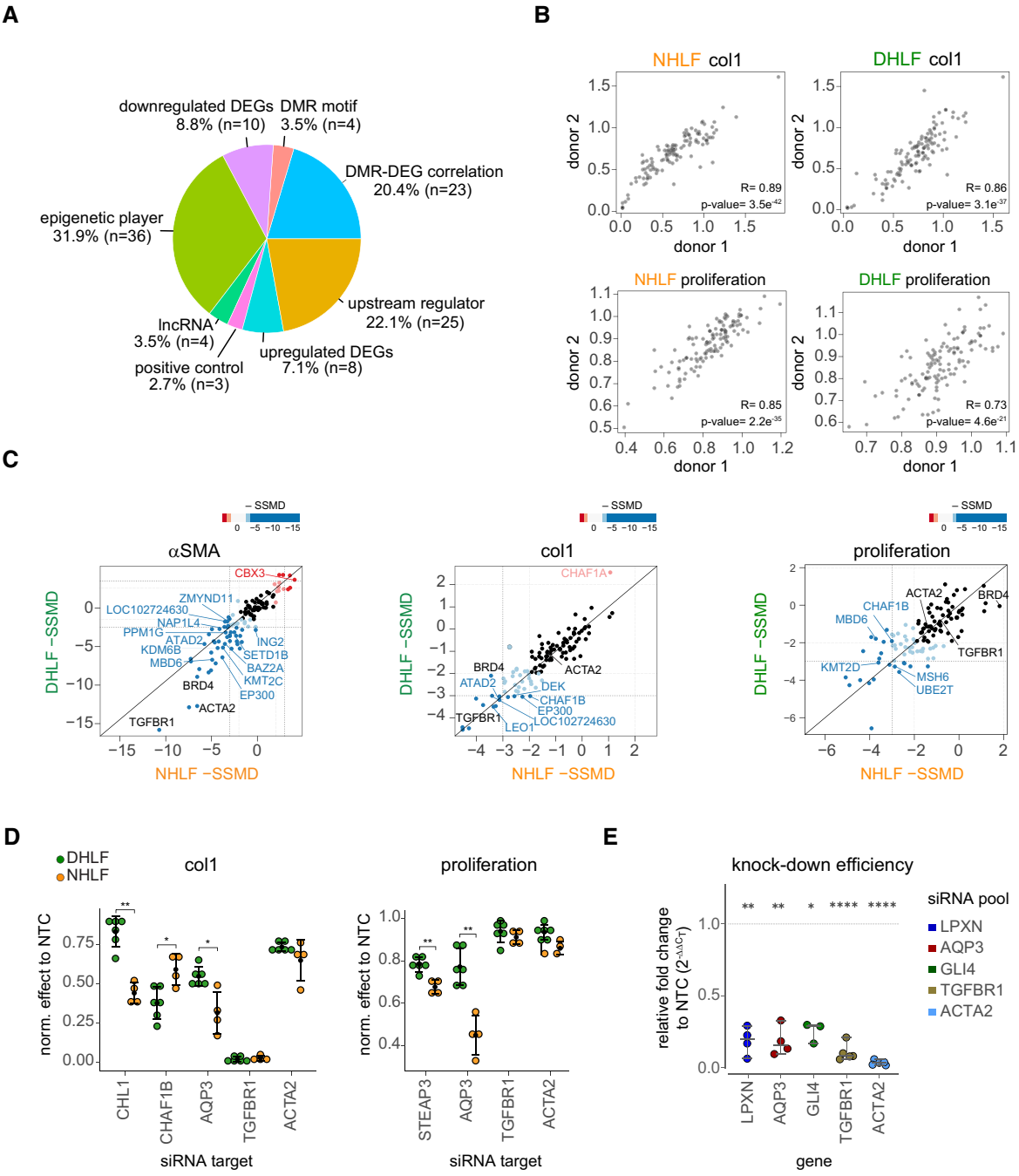

Figure EV5.