

Chromatin accessibility landscapes of large and small airway cells annotate multiple COPD susceptibility GWAS regions

Christopher J. Benway, Fei Du, Michael H. Cho, Edwin K. Silverman, Xiaobo Zhou
Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA

Abstract

Genome-wide association studies (GWAS) have identified hundreds of genome-wide significant loci for respiratory disease, including Chronic Obstructive Pulmonary (COPD), asthma, pulmonary fibrosis, lung function and COPD-related phenotypes. However, identification of causal variants and functional annotation in the appropriate cell type at these loci remain a major challenge.

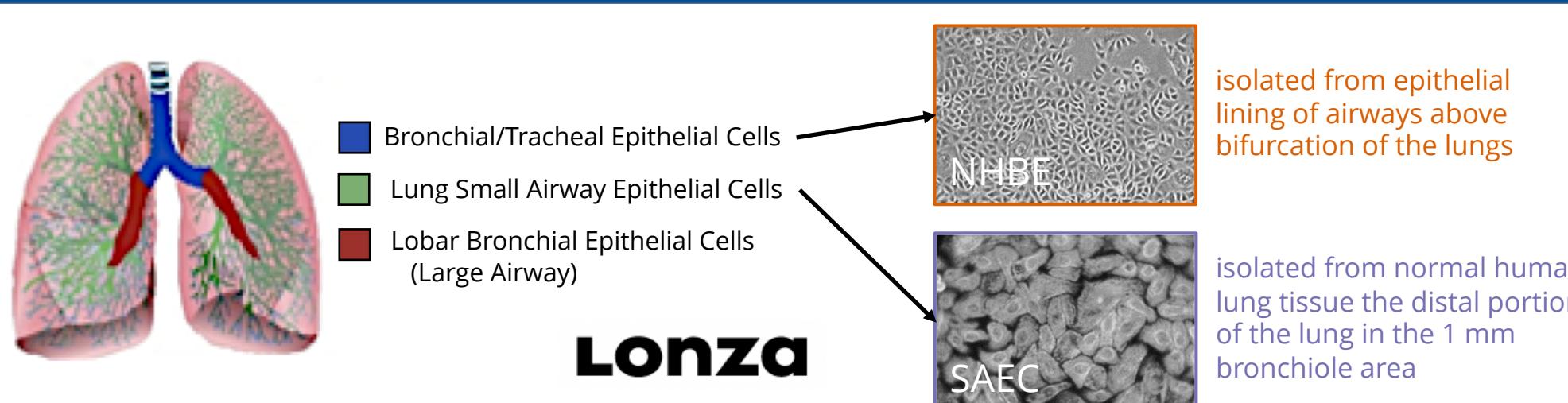
To gain a better understanding of COPD and other respiratory diseases, we sought to functionally annotate GWAS loci using chromatin accessibility profiling of small airway and bronchial epithelial cells from healthy human lungs. We performed Omni-ATAC-seq in small airway epithelial cells (SAEC) from two biological replicates and bronchial epithelial cells (NHBE) in one biological replicate as well as the 16HBE human bronchial epithelial cell line. We confirmed that Omni-ATAC-seq outperforms Fast-ATAC-seq in 16HBE cells with regard to mtDNA contamination, signal-to-noise, library complexity, and enrichment in annotated genomic regions such as transcription start sites and CTCF sites.

We tested for enrichment of 22 COPD susceptibility GWAS loci in the lung epithelial ATAC-seq peaks. Sets of SNPs within GWAS loci were generated either by a heuristic or Bayesian fine-mapping approach. For the heuristic approach, LD proxy SNPs (LD $r^2 > 0.8$) were determined for each lead SNP based on 1000 Genomes data. For the Bayesian fine-mapping approach, candidate causal variants for each locus were determined using the Probabilistic Identification of Causal SNPs (PICS) algorithm.

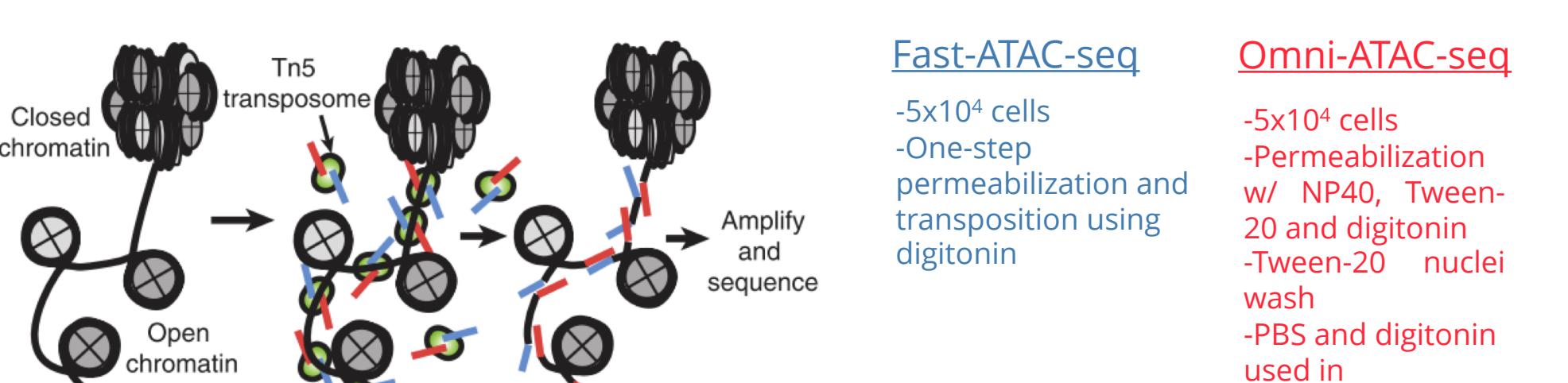
To aid in the functional annotation of COPD GWAS loci, we compared chromatin accessibility peaks and signals of these different cell types. Both primary lung subtypes were highly correlated with one another, but the 16HBE bronchial epithelial cell line was only moderately correlated with its primary counterpart. We identified 11,363 differentially accessible peaks ($FDR < 0.05$) between large and small airway epithelial cells. Further, we profiled genome-wide foot-printing of 533 transcription factors using the Protein Interaction Quantification (PIQ) computational method.

Our work confirms the performance of Omni-ATAC-seq, and suggests cell lines substantially differ from primary cells in chromatin accessibility. These factors should be considered when mapping respiratory disease loci. Further work identifying peak enrichment in fine-mapping regions is ongoing.

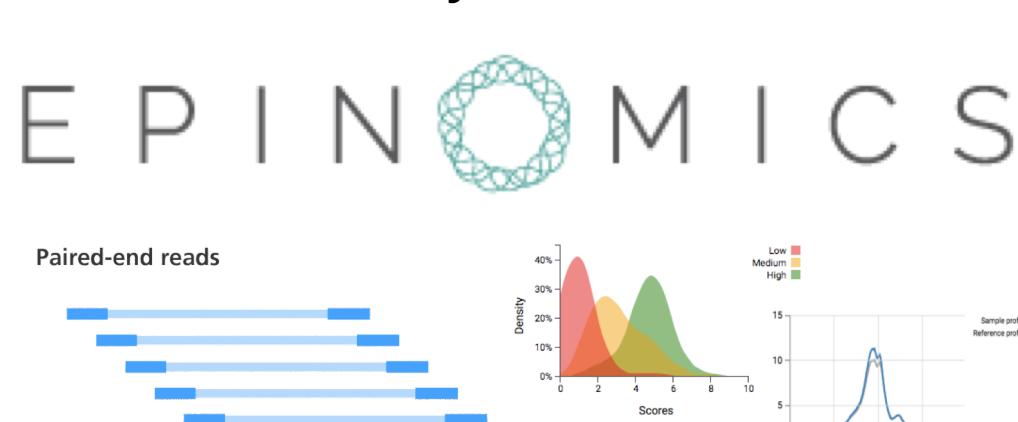
Methods



Omni-ATAC-seq optimization and comparison with Fast-ATAC-seq were conducted using the 16HBE human bronchial epithelial cell line originally established from normal bronchus tissue taken at lobectomy for squamous cell carcinoma (ATCC). This cell line has been transformed with the recombinant retrovirus LXS16E6E7 containing the human papilloma virus (HPV) E6/E7 gene. Primary normal human bronchial epithelial (NHBE) and small airway epithelial (SAEC) cells were obtained from Lonza.



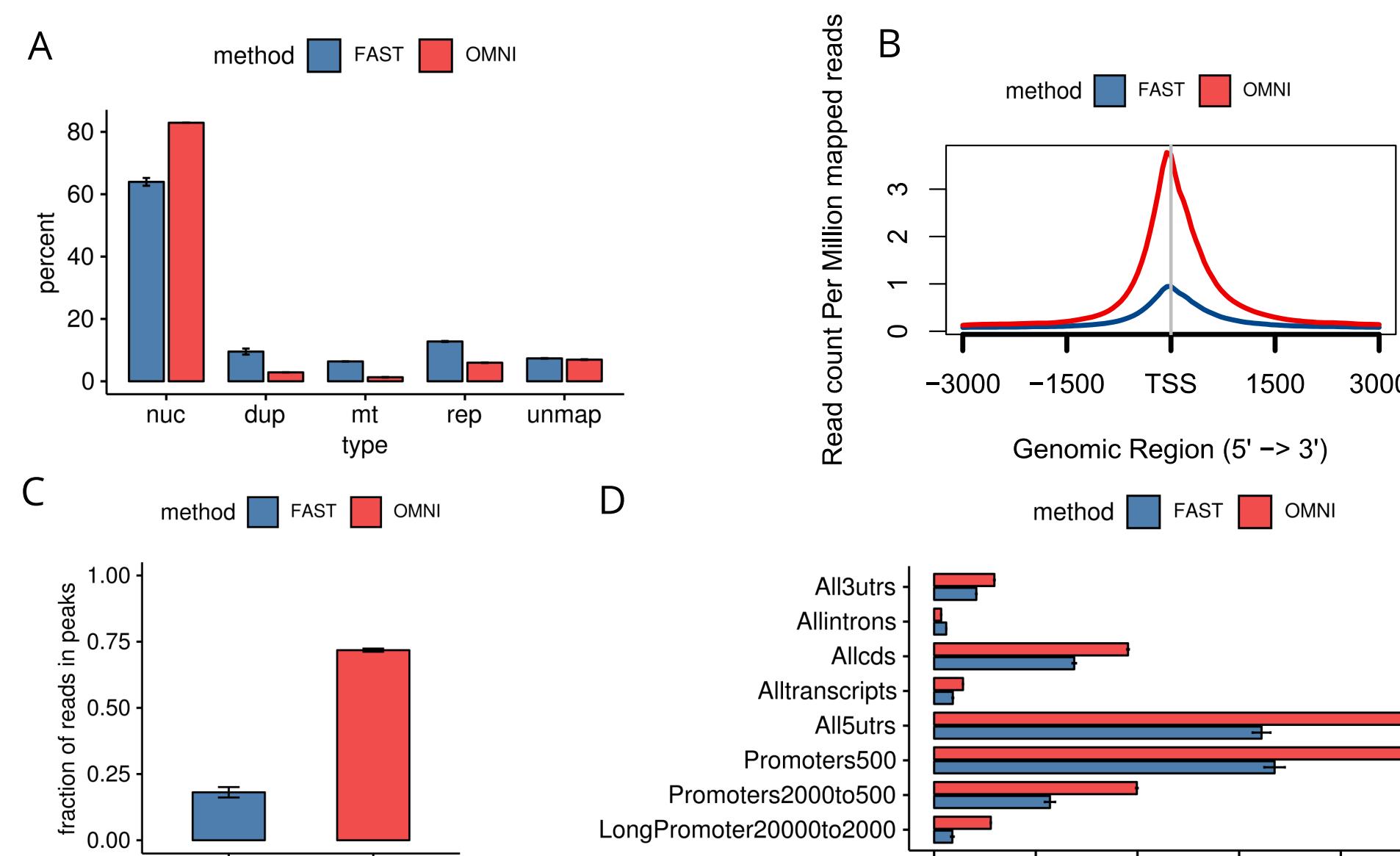
Fast-ATAC-seq and Omni-ATAC-seq were performed as described in Corces et al. 2016 and Corces et al. 2017, respectively^{4,5}. Briefly, cell lines or primary cells were cultured in appropriate serum-free media and harvested by trypsinization at sub-confluence. 50,000 cells were permeabilized and transposed using the Nextera Tn5 Library Kit (Illumina).



ATAC-seq bioinformatic analysis was performed using the Epionomics end-to-end platform (alignment, peak calling, and footprinting). Subsequent analysis and visualization was conducted in RStudio using the packages (Gviz, esATAC, ChIPseeker, ChIPpeakAnno, clusterProfiler, DiffBind, and DESeq2). COPD GWAS finemapping was conducted using summary statistics obtained from Hobbs et al 2017 and the PICS algorithm (Broad Institute)^{2,6}. Fisher exact test was performed using 'bedtools'.

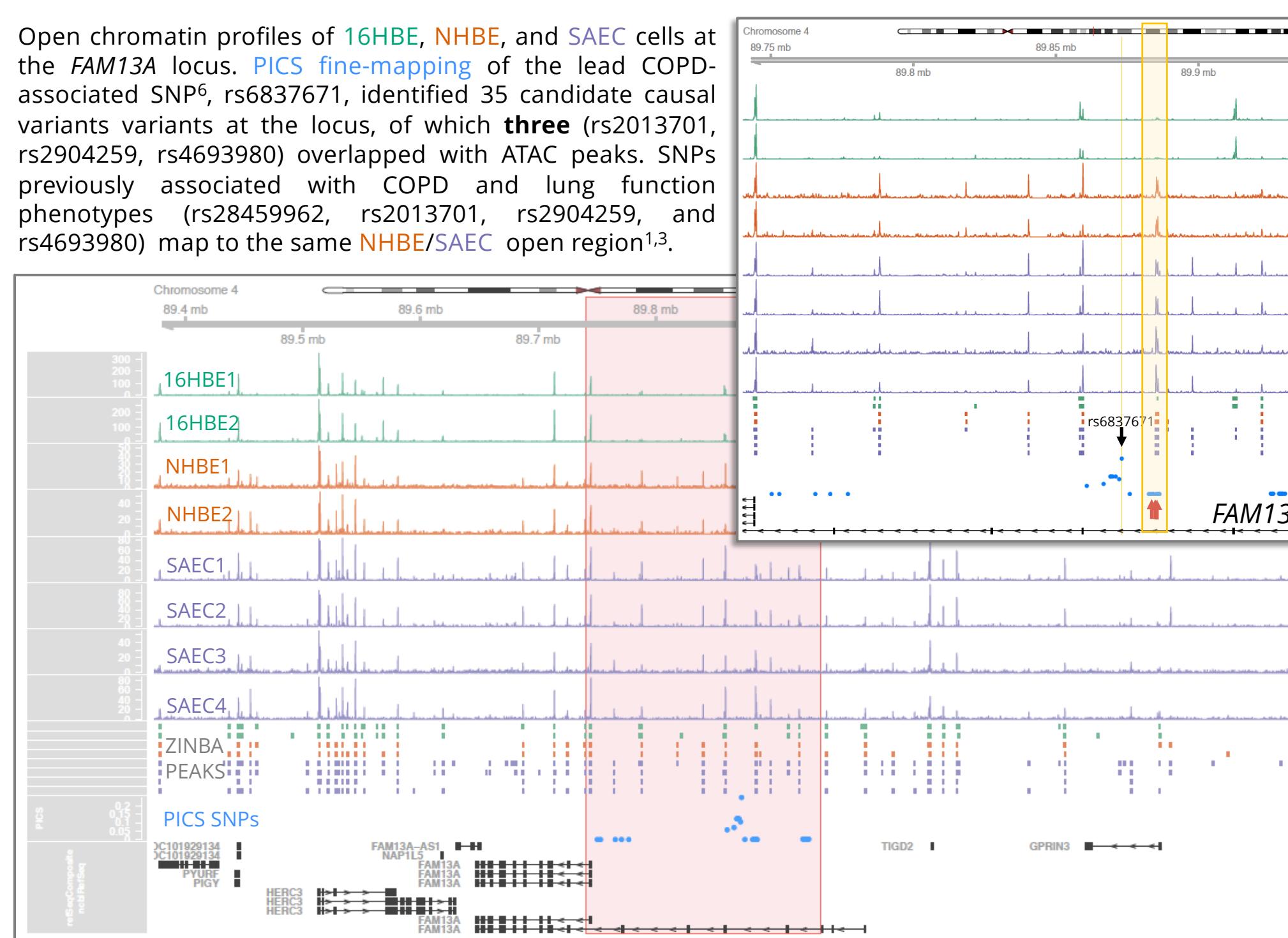
Results

Omni-ATAC-seq outperforms Fast-ATAC-seq in 16HBE human bronchial epithelial cell line

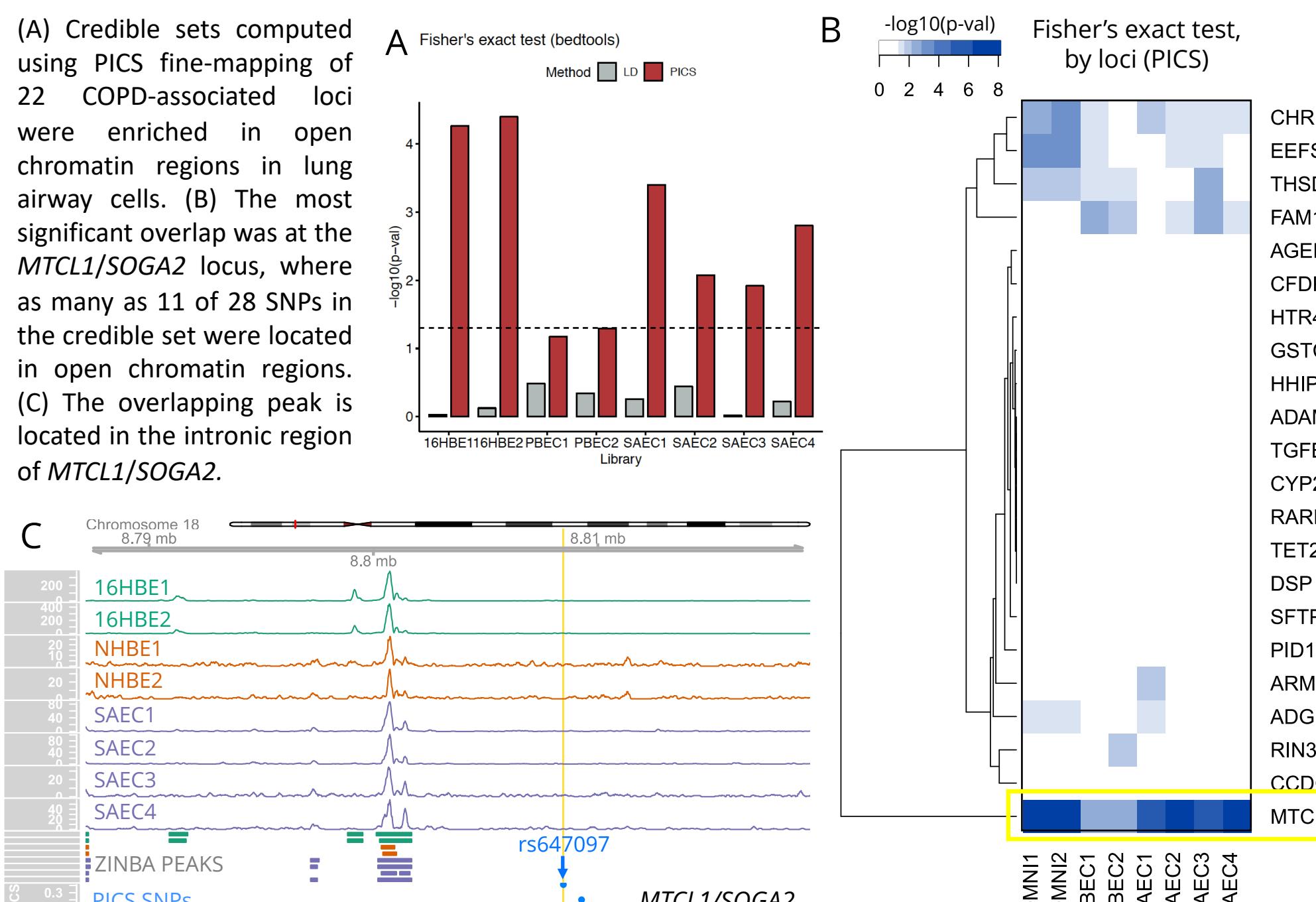


In 16HBE human bronchial epithelial cell line, the optimized **Omni-ATAC-seq** protocol, compared to **Fast-ATAC-seq**: (A) increased nuclear reads and decreased contaminant mitochondrial reads sequenced, (B) increased signal-to-background ("fraction of reads in peaks"), (C) increased signal at transcription start sites, and (D) increased enrichment of reads mapped to transcripts, promoters, and distal promoters.

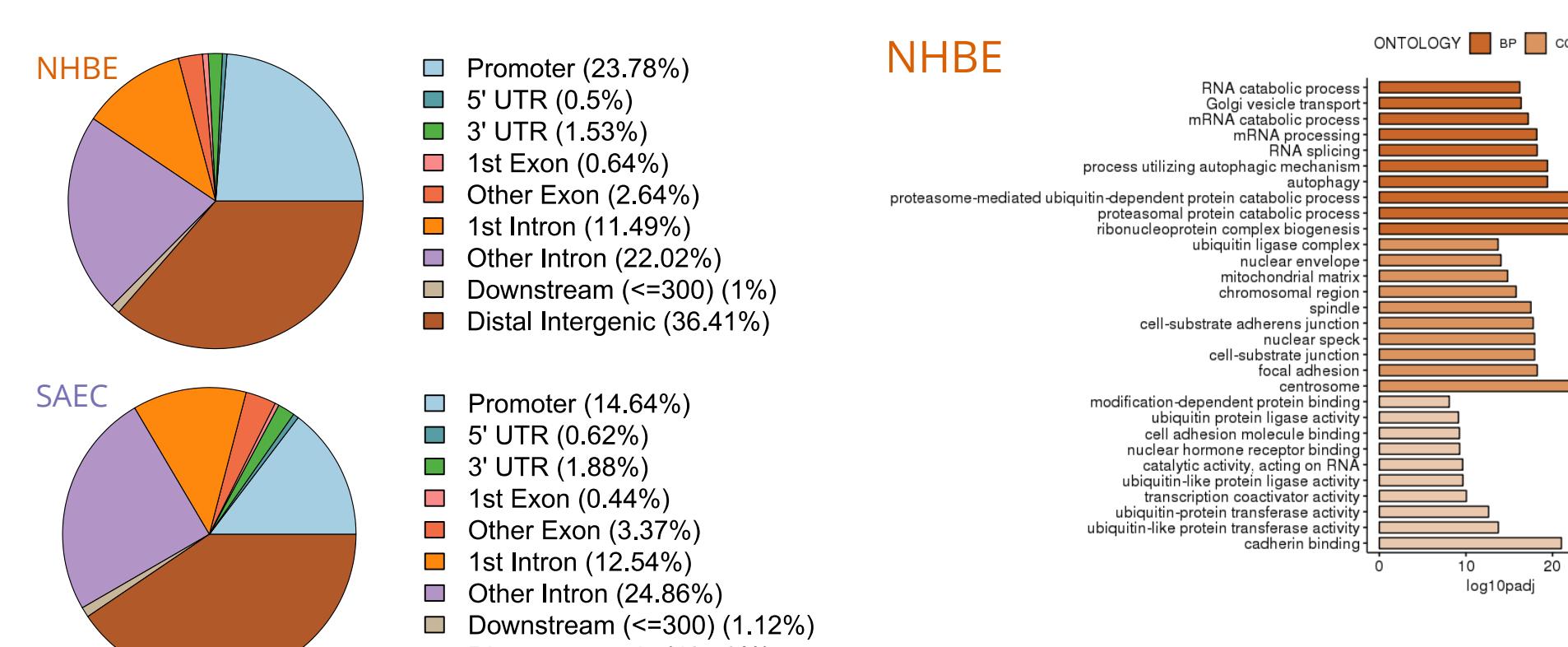
Omni-ATAC-seq of primary lung airway epithelial cells functionally annotates COPD GWAS Loci



Accessible chromatin in primary lung airway cells is enriched for COPD-associated variants

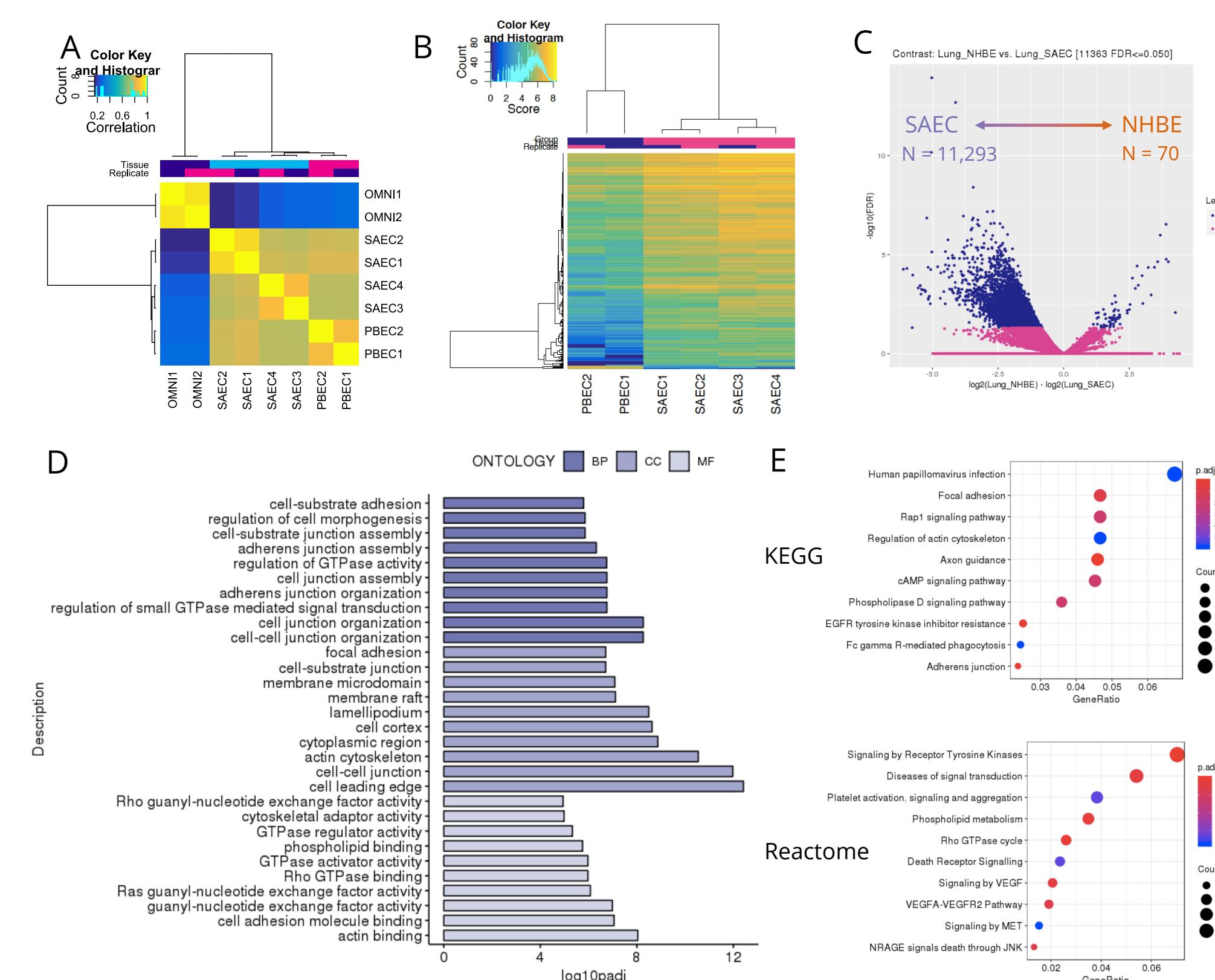


Genomic annotation and gene ontology analysis of accessible chromatin in primary lung epithelial cells



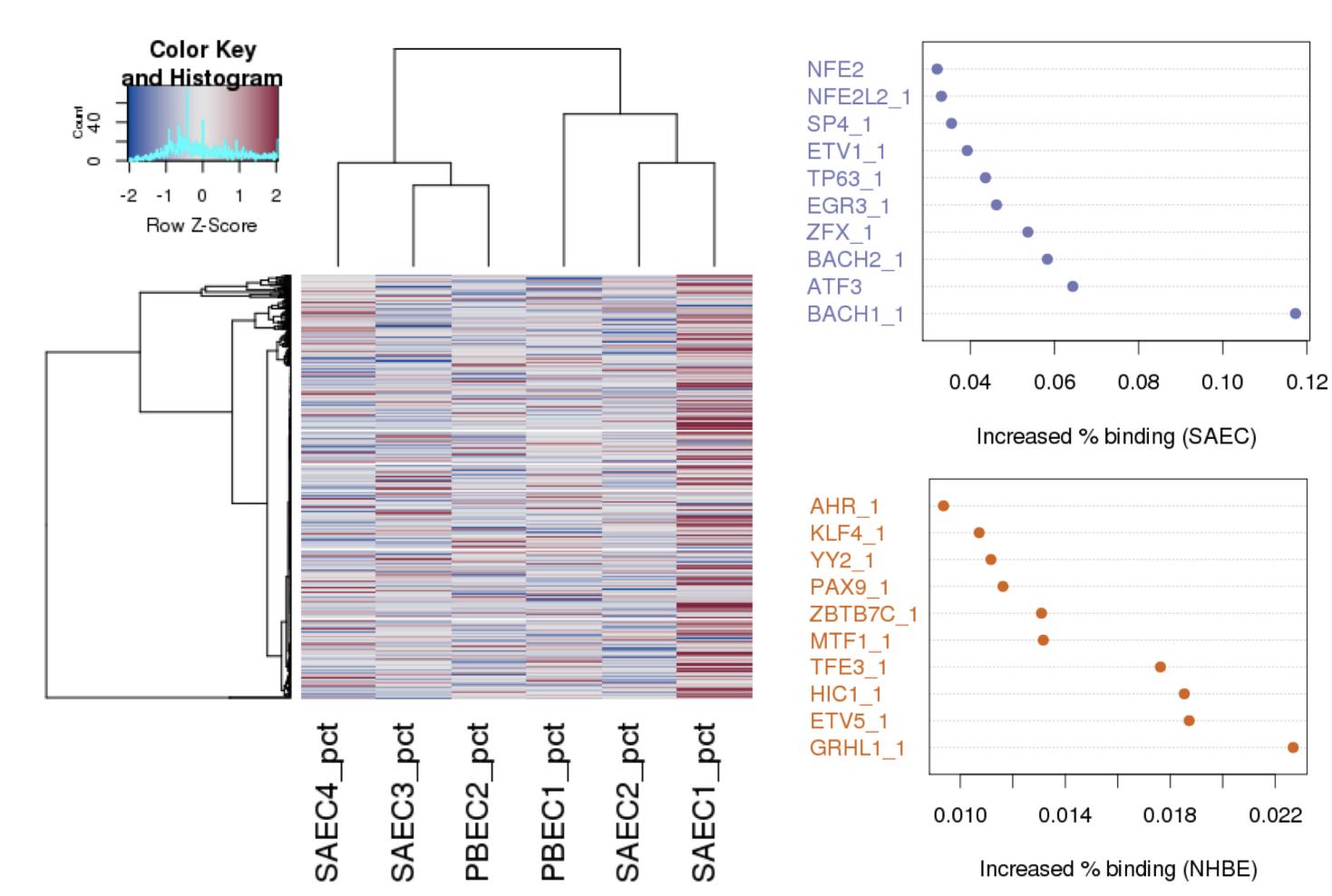
Left, open chromatin peaks in both **NHBE** and **SAEC** cell types were primarily located in promoter or intronic regions of transcripts. Representative pie charts of the libraries are depicted. Across replicates the proportions were comparable between cell types. Right, GO analysis for the genes nearest promoter and intron peaks identifies the most enriched biological processes, cellular compartments, and molecular functions in lung airway epithelial cells.

Differential chromatin accessibility in lung airway epithelium cells



(A) Pearson correlation of chromatin accessibility landscapes in primary human lung airway cells versus the **16HBE** bronchial epithelial cell line. (B) Heatmap of differentially accessible peaks in the primary lung cell types. (C) Volcano plot of differentially accessible peaks demonstrating that distribution of significant peaks is largely comprised of **SAEC** peaks. (D) Gene ontology analysis for **SAEC** peaks located in promoter or intron regions; BP = biological process, CC = cellular compartment, MF = molecular function. (E) Pathway enrichment analysis for **SAEC** peaks located in promoter or intron regions.

Genome-wide TF foot-printing profiles of lung airway



Genome-wide transcription factor (TF) profiling of 533 proteins reveals sample-specific occupancy profiles (Z-scores represent percent of total sites bound, scaled by row). **BACH1** transcription factor displayed 12% increased binding in **SAEC** compared to **NHBE**.

Conclusions

- Omni-ATAC-seq is a robust method for assaying chromatin accessibility in primary lung airway cell types.** We report that Omni-ATAC-seq in these cells outperformed previous ATAC protocols in terms of enrichment of nuclear reads, reduction of mitochondrial reads, increased signal-to-background, increased enrichment at transcription start sites and annotated transcript regions.
- Omni-ATAC-seq facilitates the functional annotation of candidate causal COPD GWAS SNPs.** Within GWAS loci associated with COPD disease status and related lung function phenotypes, we observed several open chromatin regions in lung airway cells flanking lead SNPs. Using probabilistic fine-mapping we also identified several candidate causal SNPs directly overlapping ATAC peaks in COPD associated loci.
- Differential accessibility analysis revealed differential open chromatin profiles in small and large airway cells.** Despite the overall similarity of the open chromatin profiles between cell types, we identified over 11,000 differentially open regions in small airway cells. Ontology analysis suggests these elements regulate cell junction and adhesion processes and actin binding.
- Transcription factor profiling at AT peaks reveals differential binding.** **BACH1** demonstrated increased binding in small airway compared to large airway cells.

References

- Castaldi, P. et al. Identification of Functional Variants in the FAM13A COPD GWAS Locus by Massively Parallel Reporter Assays. *American Journal of Respiratory and Critical Care Medicine* (2018). doi:10.1164/rccm.201802-0337oc
- Farh, K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518, 337-343 (2014).
- S. et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genetics* 16, (2015).
- Corces, M. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods* 14, 959-962 (2017).
- Corces, M. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature Genetics* 48, 1193-1203 (2016).
- Hobbs, B. et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nature Genetics* 49, 426-432 (2017).