University of Groningen

Cell-type eQTL deconvolution of bronchial epithelium through integration of single-cell and bulk RNA-seq

Qi, Cancan; Berg, Marijn; Chu, Xiaojing; Timens, Wim; Kole, Tessa; van den Berge, Maarten; Xu, Cheng Jian; Koppelman, Gerard H.; Nawijn, Martijn C.; Li, Yang

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Allergy EUROPEAN JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY · EAACI · WILEY

Cell-type eQTL deconvolution of bronchial epithelium through integration of single-cell and bulk RNA-seq

To the Editor,

Expression quantitative trait loci (eQTL) analyses may help to understand the function of disease-associated genetic risk factors. However, eQTL studies obtained from bulk RNA-seq data are limited by the lack of cell-type resolution. In this study, we hypothesized that eQTL deconvolution[1] of bulk RNA-seq datasets may help to identify cell-type-associated regulation of gene expression by asthma-associated genetic variants in bronchial epithelial cell subsets.

We present a tool for predicting bronchial cell proportions in bulk bronchial RNA-seq datasets, called "BronchiCellPred" (https://git.web.rug.nl/GRIAC/BronchiCellPred). The description of the methods can be found in the Appendix S1. Relative expression of the signature genes in each cell type of this tool is shown in Figure S1. We used two single-cell (sc-) RNA-seq datasets as validations to evaluate the performance of this tool, and found that the estimated cell proportions were significantly correlated with the measured frequencies for most of the cell types with spearman correlation $R > .7$ and p values <.05, except for the rare cell-type ionocyte (Figure 1B,C).

We then applied eQTL deconvolution[1] to bulk RNA-seq data of bronchial biopsies of 146 subjects from the INDURAIN cohort (Figure S2), to identify cell-type-associated *cis*-eQTLs of asthma-associated genetic variants,[2,3] in three main epithelial cell types (ciliated, basal and secretory cells). In total, we identified 14 decon-eQTLs (p value <.001, Table S1). For 11/14 eQTLs, decon-QTL associations were more significant than bulk-eQTL (Table S2). These decon-eQTLs were replicated using an independent bronchial epithelial single-cell-eQTL dataset (N = 28). Three eQTLs were replicated with p value <.05 (rs4749894~*CALML5* in basal cells, rs8103278~*FOXA3* in secretory cells, and rs10876864~*RPS26* in basal cells) (Figure 2, Table S2). All the study protocols were approved by the local medical ethics committee. All subjects gave their written informed consent.

The eQTL analysis of asthma SNPs in bronchial tissue is still limited, especially at the cell-type resolution.[2] To the best of our knowledge, this is report of the first cell-type-associated eQTL of asthma-related SNPs from bronchial biopsies by a deconvolution method, confirmed by single-cell eQTL. These approaches may help us to identify the potential targeted genes and cell types for drug discovery and improve our understanding of the regulation of genetic variants on the disease development.

Two of the three replicated decon-eQTLs have previously been identified as bulk-eQTLs by GTEx (https://gtexportal.org/), including rs10876864~*RPS26* which were identified in multiple tissues including lung tissue, and rs8103278~*FOXA3* which were identified in minor salivary gland. Our study revealed the eQTL effect in cell types relevant to asthma. *FOXA3* is a transcription factor expressed in goblet cells in the airways, which induces mucus production and goblet cell differentiation and may be involved in antiviral responses of epithelial cells.[4] In our study, we showed that the asthma-associated risk allele was associated with increased *FOXA3* expression in secretory epithelial cells, a cell-type annotation that encompasses both club and goblet cells,[5] thereby providing a molecular mechanism for goblet cell hyperplasia and increased mucus production in asthma. *CALML5*, expressed in bronchial epithelial cells and keratinocytes, is important for epithelial differentiation and barrier function.[6] We identified a basal cell-specific increased expression of *CALML5* associated with the asthma-associated variant, which may play a role in epithelial repair and differentiation, and thus, in maintaining and/or restoring epithelial integrity in asthma.

The accuracy of cell-type estimation is important for eQTL deconvolution.[1] Most of the cell types can be well estimated using our methods, which was supported by the validation analyses. However, the method may not have enough power to predict rare cell type such as ionocytes. This should be kept in mind when filtering on the estimated cell-type proportion to be applied in downstream analysis.

Many of the decon-eQTL genes showed low expression in scRNA-seq data (Figure S3), which makes it difficult to identify these cell-type eQTL by sc-eQTL. This indicates that the deconvolution methods may even be more powerful to identify these cell-type eQTL effects. As shown in our results, we were able to identify the cell-type eQTL effect with a limited sample size, some of which can be replicated by a sc-eQTL method for genes that had sufficiently high expression in the scRNA-seq datasets.

However, there are still some limitations in this study: 1) the sample size of bulk data is small. We were only able to assess the three main epithelial cell types, and we applied a suggestive p
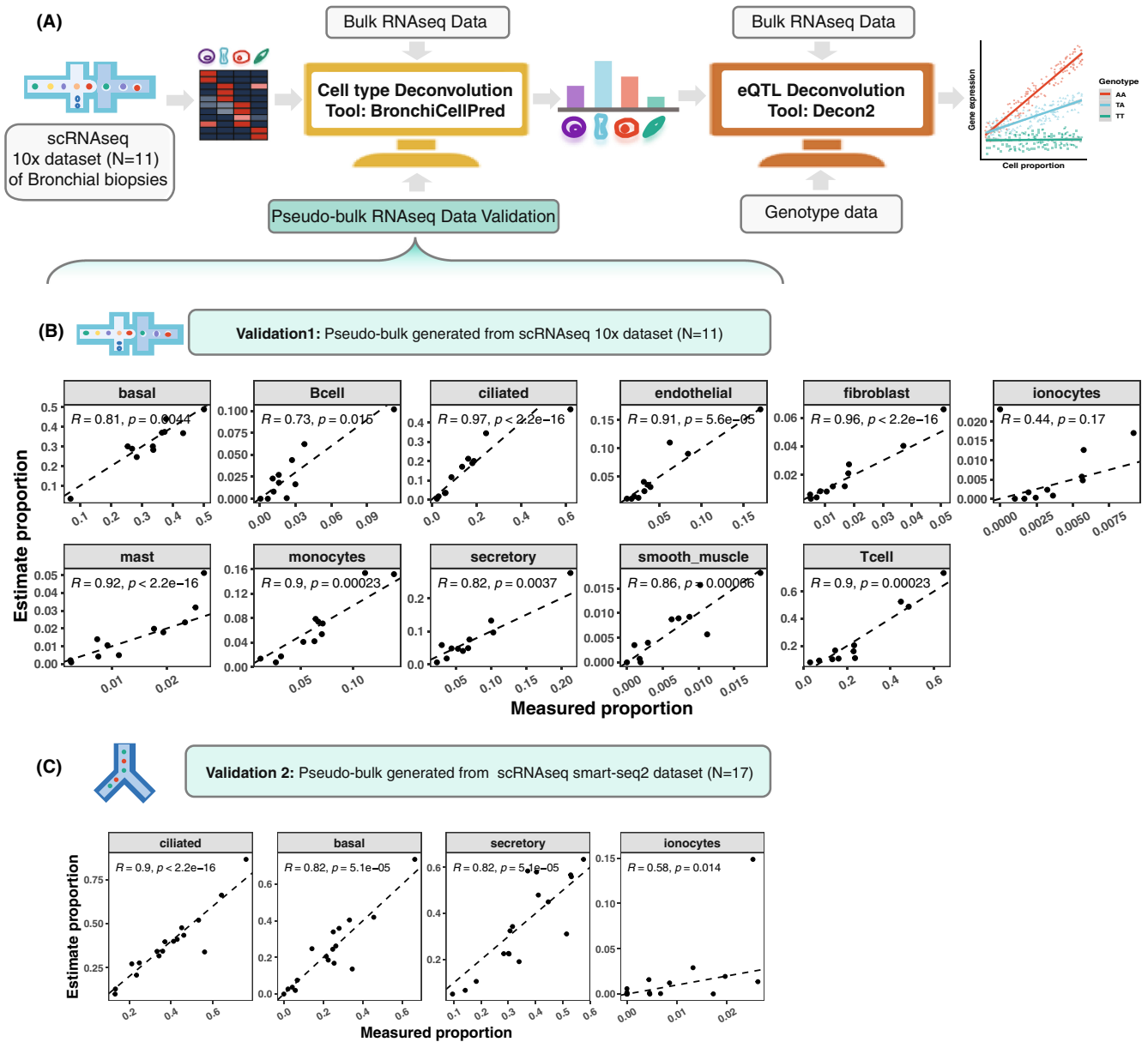
FIGURE 1 (A) Study design. Gene signature matrices were generated from a scRNA-seq dataset of bronchial biopsies measured by Illumina 10x platform. We present a tool "BronchiCellPred" for users to estimate bronchial cell proportions from bulk RNA-seq data. We used an eQTL deconvolution tool "Dencon2", with estimated cell proportions, bulk RNA-seq data and matched genotype data as the input, to identify the cell-type eQTL effect in bronchial cells. (B) Scatter plots showing the correlation between measured cell proportion and estimated cell proportion in 10x scRNA-seq data. To evaluate the performance of cell-type deconvolution, we generated two pseudo-bulk expression datasets from the 11 samples with 10x scRNA-seq data (validation 1, B) and the 17 samples that only have Smartseq2 scRNA-seq data (validation 2, C), respectively. Spearman correlation coefficient (R) and p value were calculated for each cell type. There were 11 cell types identified by 10x dataset. (C) Scatter plots showing the correlation between measured cell proportion and estimated cell proportion in Smartseq2 dataset, less cell types (N = 5) were identified with 4 cell types overlapped with the 10x dataset that were shown

value threshold considering the limited power. Here, we demonstrated the eQTL deconvolution analysis by focusing on the three abundant cell types. The method can be applied to include less abundant cell subpopulations when the sample size increase. 2) The identified eQTL genes were lowly expressed in scRNA-seq data in the replication dataset, leading to low power in our validation design.

In conclusion, this study provides the first proof of concept of using eQTL deconvolution method to identify cell-type eQTL for asthma-associated genetic variants in bronchial biopsies. The method to deconvolute cell types from the bronchial airways is openly available for future use. The eQTL deconvolution analyses are likely to be a powerful tool to better understanding cell type-specific effects of asthma-associated genetic variants, thereby providing a
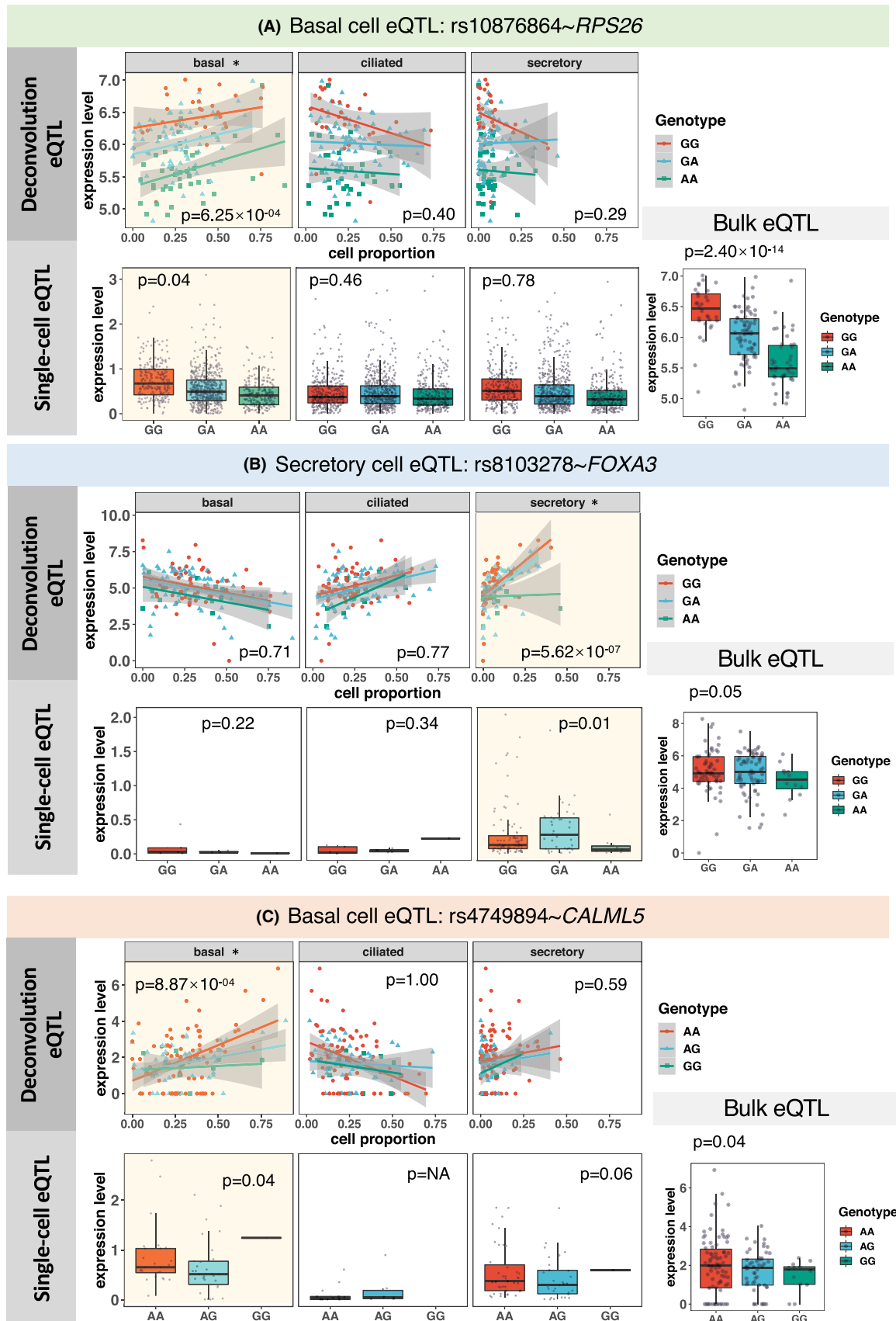
FIGURE 2 Cell-type eQTL identified by eQTL deconvolution. Cell-type eQTL replicated by sc-eQTL. In each small panel, top scatter plots showing the deconvoluted eQTLs in basal, ciliated and secretory cell; bottom left, three boxplots showing replicated sc-eQTLs in three cell types; and bottom right, boxplots showing the bulk-eQTL in bronchial biopsies

broader overview of the molecular mechanisms of asthma development and can be applied on other large bulk RNA-seq datasets in respiratory disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Cancan Qi[1,2,3,4,5]
Marijn Berg[2,6]
Xiaojing Chu[5,7,8]
Wim Timens[2,6]
Tessa Kole[2,9]
Maarten van den Berge[2,9]
Cheng-Jian Xu[4,5,8,10]
Gerard H. Koppelman[1,2]
Martijn C. Nawijn[2,6]
Yang Li[4,5,8]

[1]Department of Pediatric Pulmonology and Pediatric Allergology, University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, University of Groningen, Groningen, The Netherlands
[2]University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, The Netherlands
[3]Microbiome Medicine Center, Division of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, China
[4]Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands
[5]Centre for Individualized Infection Medicine (CiiM), a Joint Venture Between Hannover Medical School and Helmholtz Centre for Infection Research, Hannover, Germany
[6]Department of Pathology & Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[7]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[8]TWINCORE, Centre for Experimental and Clinical Infection Research, a Joint Venture Between Hannover Medical School and Helmholtz Centre for Infection Research, Hannover, Germany
[9]Department of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
[10]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany

Correspondence
Yang Li, Centre for Individualized Infection Medicine (CiiM), Feodor-Lynen-Street 7, 30625 Hannover, Germany.
Email: yang.li@helmholtz-hzi.de

ORCID

Cancan Qi https://orcid.org/0000-0003-3825-5802
Cheng-Jian Xu https://orcid.org/0000-0003-1586-4672

REFERENCES

1. Aguirre-Gamboa R, de Klein N, di Tommaso J, et al. Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics*. 2020;21:243.
2. El-Husseini ZW, Gosens R, Dekker F, Koppelman GH. The genetics of asthma and the promise of genomics-guided drug target discovery. *Lancet Respir Med*. 2020;8:1045-1056.
3. Han Y, Jia Q, Jahani PS, et al. Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat Commun*. 2020;11:1776.
4. Chen G, Korfhagen TR, Karp CL, et al. Foxa3 induces goblet cell metaplasia and inhibits innate antiviral immunity. *Am J Respir Crit Care Med*. 2014;189:301-313.
5. Vieira Braga FA, Kar G, Berg M, et al. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med*. 2019;25:1153-1163. http://www.nature.com/articles/s41591-019-0468-5 Accessed June 24, 2019.
6. Vos JB, Datson NA, van Kampen AH, et al. A molecular signature of epithelial host defense: comparative gene expression analysis of cultured bronchial epithelial cells and keratinocytes. *BMC Genomics*. 2006;7:9.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.