VILNIUS UNIVERSITY

Rapolas
Žilionis

Defining cell populations and their conservation between human and mouse in airway epithelium and lung cancer by single cell RNA sequencing

DOCTORAL DISSERTATION

Natural Sciences,
Biochemistry **N 004**

VILNIUS 2019

This thesis was prepared during doctoral studies at Vilnius University from 2014 to 2018. Research described in this thesis was performed in Allon Klein laboratory at Harvard Medical School and Linas Mažutis laboratory at the Life Sciences Center of Vilnius University.

Academic supervisor – Dr. Linas Mažutis (Vilnius University, Natural Sciences, Biochemistry – N 004).

Academic consultant – Dr. Allon Klein (Harvard Medical School, USA, Natural Sciences, Physics – N 002).

Dissertation Defence Panel:
Chairman – Assoc. prof. dr. Gintaras Valinčius (Vilnius University, Natural Sciences, Biochemistry – N 004).

Members**:**
Assoc. prof. dr. Daiva Dabkevičienė (Vilnius University, Natural Sciences, Biochemistry – N 004);

Prof. dr. Skirmantas Kriaučionis (Oxford University, Natural Sciences, Biochemistry – N 004);

Assoc. prof. dr. Algirdas Žiogas (Centre for Innovative Medicine, Natural Sciences, Biochemistry – N 004);

Prof. dr. Aurelija Žvirblienė (Vilnius University, Natural Sciences, Biochemistry – N 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 2PM on 20 September 2019 in room R101 of the Life Sciences Center, Vilnius University.
Address: Saulėtekio al. 7, LT-10257 Vilnius, Lithuania
Tel. +370 522 34420; e-mail: info@gmc.vu.lt

The text of this dissertation can be accessed at the Vilnius University Library as well as on the website of Vilnius University:
www.vu.lt/lt/naujienos/ivykiu-kalendorius

VILNIAUS UNIVERSITETAS

Rapolas
Žilionis

Ląstelių populiacijų ir jų atitikimo tarp pelės ir žmogaus nustatymas kvėpavimo takų epitelyje ir plaučių navikuose taikant pavienių ląstelių RNR sekoskaitą

DAKTARO DISERTACIJA

Gamtos mokslai,
Biochemija **N 004**

VILNIUS 2019

Disertacija rengta 2014–2018 metais Vilniaus universitete ir Harvardo medicinos mokykloje.

Mokslinis vadovas – Dr. Linas Mažutis (Vilniaus universitetas, gamtos mokslai, biochemija N – 004).

Mokslinis konsultantas – Dr. Allon Klein (Harvardo medicinos mokykla, JAV, gamtos mokslai, fizika, N – 002).

Gynimo taryba:
Pirmininkas – **doc. dr. Gintaras Valinčius** (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).

Nariai:
doc. dr. Daiva Dabkevičienė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004);

prof. dr. Skirmantas Kriaučionis (Oksfordo universitetas, gamtos mokslai, biochemija – N 004);

doc. dr. Algirdas Žiogas (Inovatyvios medicinos centras, gamtos mokslai, biochemija – N 004);

prof. dr. Aurelija Žvirblienė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004).

Disertacija bus ginama viešame Gynimo tarybos posėdyje 2019 m. rugsėjo 20 d. 14 val. Vilniaus universiteto Gyvybės mokslų centro R101 auditorijoje.

Adresas: Saulėtekio al. 7, LT-10257 Vilnius, Lietuva.
Tel. +370 522 34420; el. paštas: info@gmc.vu.lt

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: https://www.vu.lt/naujienos/ivykiu-kalendorius.

To my dear wife Vilija.

ABSTRACT

2015 marks the advent of high-throughput single cell RNA sequencing (scRNAseq) technologies, which enabled measuring gene expression levels in tens of thousands of cells in a single experiment. One of the methods, inDrops, elegantly harnessed the throughput of droplet microfluidics, and received immediate interests both from the direct environment which witnessed its development as well as a broader community. However, the complexity of the multi-step procedure dependent on custom-made reagents called for a detailed public protocol before it could be widely adopted. In this thesis, I present efforts to improve the accessibility of inDrops, as well as the application of the technology to study two biological systems: the airway epithelium and lung tumor-infiltrating myeloid cells.

In both studies, scRNAseq was used to generate unbiased atlases of cell types in the sample. In the case of airway epithelium, we entered the project seeking a better understanding of the heterogeneity of basal cells, which are the stem cells of the niche, and how the homeostatic tissue composition is restored following injury. While our results cover this original motivation, the focus of the study turned in response to the discovery of a rare and previous uncharacterized cell type, which we called pulmonary ionocytes. We further determined the location of these cells in the airway, gained insight into their specification, and established their relevance to cystic fibrosis. While the function of pulmonary ionocytes is still to be fully understood, we anticipate that their identification will better inform cystic fibrosis research.

In the second study, we sought to characterize the diversity of tumor-infiltrating myeloid cells, estimate the conservation of observed states between human and mouse, and summarize our results in a resource that would serve as a starting point for mechanistic follow ups. Our analysis revealed a one-to-one human-mouse match in dendritic cell and monocyte subsets, partial conservation of neutrophil states, and species differences in macrophages. Neutrophil diversity was largely novel, and a subset of tumor-specific neutrophils was established to have tumor-promoting functions as part of a third study I had the chance to contribute to. The goal of providing a resource was fulfilled by enabling access to our data in multiple formats, including tools for interactive exploration. We hope that our study will contribute to a better understanding of species differences, which in turn would allow focusing efforts in drug and disease mechanism research to human-relevant aspects of mouse biology.

TABLE OF CONTENTS

1. ABBREVIATIONS

ALI	air-liquid interface
BC	basal cells
BHB	barcoding hydrogel bead
cDC	classical dendritic cell
cDNA	copy DNA
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
COPD	chronic obstructive pulmonary disease
CPM	counts per million
DC	dendritic cell
dpi	days post-injury
FACS	fluorescence assisted cell sorting
GEMM	genetically engineered mouse model
GEO	gene expression omnibus
HB	hydrogel bead
HBEC	human bronchial epithelial cell
HMS	Harvard Medical School
HSC	hematopoietic stem cell
ICB	immune checkpoint blockade
IFN-γ	interferon gamma
KP	$Kras^{G12D/+};Trp^{53-/-}$
LPS	lipopolysaccharide
MCC	mucociliary clearance
NK	natural killer
NSCLC	non-small cell lung cancer
PBA	Population Balance Analysis
PCD	primary ciliary dyskinesia
pDC	plasmacytoid dendritic cell
PNEC	pulmonary neuroendocrine cell
RT	reverse transcription
scRNAseq	single cell RNA sequencing
TIMs	tumor-infiltrating myeloid cells

2. PUBLICATIONS AND CONTRIBUTIONS

2.1. Publications forming the base of this thesis

I. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM[#], Mazutis L[#]. Single-cell barcoding and sequencing using droplet microfluidics. *Nature Protocols*. 2017 Jan;12(1):44–73.

My main contribution to this article consists of experimental work, writing the protocol and incorporating several changes made to the original description of inDrops. I described the inDrops procedure and anticipated results. I acquired data for the anticipated results section, performed the reverse transcription reaction optimization that lead to improved cell barcoding conditions. I also optimized and described a liquid hander protocol that is used to synthesize barcoded hydrogel beads.

II. Plasschaert LW[*], **Zilionis** R[*], Choo-Wing R, Savova V, Knehr J, Roma G, Klein AM[#], Jaffe AB[#]. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*. 2018 Aug;560(7718):377–81.

I performed all scRNAseq experiments and the analysis of resulting data. I prepared figures involving scRNAseq data that are presented in the article. I also participated in writing the manuscript and I am currently curating the scRNAseq data that was made publicly available as a resource.

III. **Zilionis** R[*], Engblom C[*], Pfirschke C*, Savova V*[#], David Zemmour, Saatcioglu HD, Krishnan I, Maroni G, Meyerovitz CV, Kerwin CM, Choi S, Richards WG, De Rienzo A, Tenen DG, Bueno R, Levantini E, Pittet MJ[#], Klein AM[#]. Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity*. 2019 May 21;50(5):1317– 34.e10.

My contribution is the same as for Paper II. In addition, I played an active role in defining the scope and goals of this study.

IV. Engblom C[*], Pfirschke C[*], **Zilionis R**, Da Silva Martins J, Bos SA, Courties G, Rickelt S, Severe N, Baryawno N, Faget J, Savova V, Zemmour D, Kline J, Siwicki M, Garris C, Pucci F, Liao HW, Lin YJ, Newton A, Yaghi OK, Iwamoto Y, Tricot B, Wojtkiewicz GR, Nahrendorf M, Cortez-Retamozo V, Meylan E, Hynes RO, Demay M, Klein K, Bredella MA, Scadden DT, Weissleder R, Pittet MJ[#]. Osteoblasts remotely supply lung tumors with cancer-promoting SiglecFhigh neutrophils. *Science*. 2017 Dec 1;358(6367).

I performed neutrophil scRNAseq data analysis and patient survival analysis presented in Figs. 5 and S19, wrote the respective methods sections, and reviewed and edited relevant results sections.

[*]These authors contributed equally.

[#]Corresponding authors.

2.2. Publication not included in this thesis

V. Galinis R[*], Stonyte G[*], Kiseliovas V, **Zilionis R**, Studer S, Hilvert D, Janulaitis A, Mazutis L[#]. DNA Nanoparticles for Improved Protein Synthesis In Vitro. *Angewandtle Chemie*. 2016 Feb 24;55(9):3120–3.

VI. Derr A, Yang C[*], **Zilionis R**[*], Sergushichev A, Blodgett DM, Redick S, Bortell R, Luban J, Harlan DM, Kadener S, Greiner DL, Klein AM, Artyomov MN, Garber M[#]. End Sequence Analysis Toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. *Genome Research*. 2016 Oct;26(10):1397– 410.

VII. Hrvatin S[*], Hochbaum DR[*], Nagy MA[*], Cicconet M, Robertson K, Cheadle L, **Zilionis R**, Ratner A, Borges-Monroy R, Klein AM, Sabatini BL[#], Greenberg ME[#]. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nature Neuroscience*. 2018 Jan;21(1):120–9.

VIII. Tusi BK[*], Wolock SL[*], Weinreb C, Hwang Y, Hidalgo D, **Zilionis R**, Waisman A, Huh JR, Klein AM[#], Socolovsky M[#]. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*. 2018 Mar 1;555(7694):54–60.

IX. Zemmour D, **Zilionis R**, Kiner E, Klein AM, Mathis D[#], Benoist C[#]. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nature Immunology.* 2018 Mar;19(3):291–301.

X. DiSpirito JR[*], Zemmour D[*], Ramanan D, Cho J, **Zilionis R**, Klein AM, Benoist C[#], Mathis D[#]. Molecular diversification of regulatory T cells in nonlymphoid tissues. *Science Immunology*. 2018 Sep 14;3(27).

XI. Jones KB, Furukawa S[*], Marangoni P[*], Ma H, Pinkard H, D'Urso R, **Zilionis R**, Klein AM, Klein OD. Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell.* 2019 Jan 3;24(1):183–92.e8.

XII. Sharir A, Marangoni P[*], **Zilionis R**[*], Wan M, Wald T, Hu J, Kawaguchi K, Castillo-Azofeifa D, Epstein Leo, Harrington K, Pagella P, Mitsiadis T, Siebel CW, Klein AM[#] and Klein OD[#]. A large pool of actively cycling progenitors orchestrates self-renewal and injury repair of an ectodermal appendage. Accepted to Nature Cell Biology.

[*]These authors contributed equally.

[#]Corresponding authors.

In Paper V, I performed initial experiments showing that the addition of pyrophosphate (PPi) to a solution containing $Mg2^{+}$ ions and plasmid DNA was sufficient to obtain precipitates of DNA:PPi:Mg complexes.
In Papers VI-X, I performed either all (VI, IX and X) or part of microfluidic single-cell transcriptome barcoding experiments using inDrops.

In Papers XI and XII, I performed the single cell RNA seq data analysis. For Paper XII, I am also curating the resources and data which were made public, including a GitHub repository with python code I wrote and used for the analyses shown in the paper (github.com/rapolaszilionis/Sharir_et_al_2019). Although not included in the main part of this thesis, the study described in Paper XII represents a major effort in the context of my doctorate.

3. INTRODUCTION

3.1. About the structure of this thesis

This thesis is composed of 3 stories, supported by 4 publications listed in Chapter 2.1.

Story I, supported by Paper I, focuses on making a new method readily accessible to a broader scientific community. The method in question is inDrops, a single-cell RNA sequencing (scRNAseq) approach developed by my PhD supervisors Linas Mažutis and Allon M. Klein, and published in Cell in 2015 (1). Making inDrops readily accessible involved: mastering all steps of the approach myself, optimizing steps that resulted in inconsistent results, adopting and integrating improvements made by colleagues, and providing detailed step-by-step instructions in a strictly organized Nature Protocols publication. InDrops is also the unifying theme of the 3 stories composing this thesis, as Story II and III harness the technique to gain insight into two biological systems.

Story II, supported by Paper II, describes the discovery of pulmonary ionocytes, a novel cell type in the airway epithelium. This work is a collective effort of two laboratories. It was led by Lindsey Plasschaert from Aron Jaffe lab at Novartis as expert in airway biology, and myself as member of Allon Klein lab at HMS as expert in scRNAseq. This discovery exemplifies one of the key strengths of inDrops and scRNAseq in general: *de novo* characterization of the cellular composition of tissues. Therefore, even in well studied systems like the airway epithelium scRNAseq can reveal unappreciated cell types. The Results Chapter 6.2 elaborates further on the findings in the paper. Here, I would like to emphasize that pulmonary ionocytes could be found both in human and mouse airway, this way representing a conserved cell type across the two species.

Story III, supported by Paper III, elaborates specifically on the question of cell type conservation between human and mouse. This question is motivated by the pivotal role of mouse model in human-focused research. In this study, we focused on the immune context of non-small cell lung cancer (NSCLC), by first addressing the unbiased population structure of tumor infiltrating immune cells, and then identify cell states with counterparts between species. My role in the study was performing scRNAseq experiments and analyzing the resulting data. This work was performed in close collaboration with an expert lab in the field of tumor immunology: Pittet lab at Massachusetts General Hospital. It was co-led with Camilla Engblom, Christina Pfirschke and Virginia Savova.

One of the insights of Story III was the existence of previously unappreciated tumor-specific neutrophil states. Although at lower resolution, these states are already described in Paper IV, where my contribution was to identify genes characteristic to tumor-specific neutrophils, and related their expression to patient survival. Apart from these analyses, Paper IV should be regarded as a separate study, and its key insights are generally not related to this thesis.

3.2. scRNAseq-specific background (Story I)

3.2.1. Single cell RNA sequencing enables *de novo* identification of cell types

The primary application of single cell RNA seq (scRNAseq) is cataloguing of cells that compose a given sample. Knowing the expression level of every gene in thousands of individual cells enables a data-driven identification of groups of cells sharing similar gene expression. Importantly, because all rather than *a priori* chosen genes are assayed, the identification of cell types is not biased by the experimentalist and enables a truly naïve approach to the tissue being studied.

From a technical point of view, as with conventional RNA sequencing, current scRNAseq protocols universally involve an initial reverse transcription (RT) of mRNA (2, 3). It is followed by amplification of the resulting cDNA to obtain a library of double stranded DNA fragments compatible with high-throughput sequencing platforms. As discussed in the following paragraph, exploiting the initial RT reaction to introduce cell-specific barcodes was an important breakthrough in the evolution of scRNAseq methods.

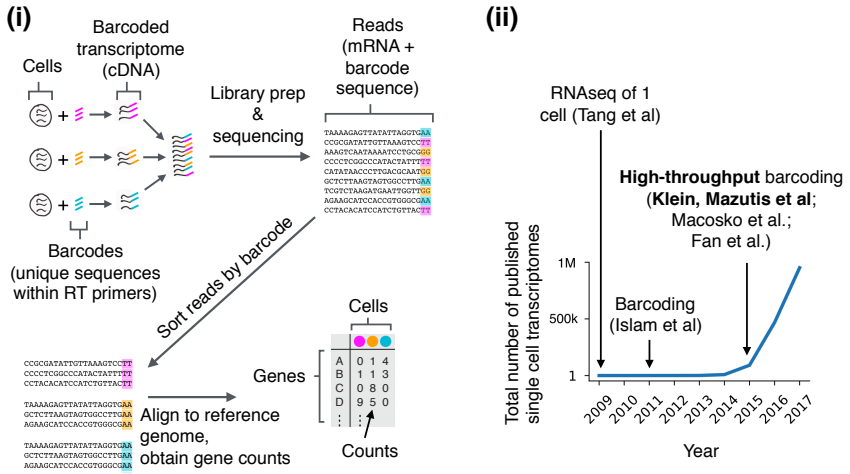

Figure A. Milestones of scRNAseq technology evolution. (i) The principle of multiplexing: introducing barcodes in the form of unique nucleotide sequences allows to pool the material from multiple cells right after the initial RT reaction. Further sequencing library preparation is performed treating all the material as a

single sample. The barcode sequence is revealed during sequencing and used to sort reads by cell of origin, as all reads originating from the same cell share the same barcode. The final result is a matrix containing the expression level of all genes in each cell in the sample. (ii) Cumulative plot showing the total number of single cell transcriptomes published since 2009. The plot reveals that although multiplexing was introduced in 2011, rapid growth in the total number of single cell transcriptomes was enabled by high-throughput barcoding solutions, inDrops being one of them. Data kindly provided by Valentine Svensson (2).

3.2.2. 6 years prior to inDrops

The first whole transcriptome profile obtained from a single cell was published in 2009. By developing efficient amplification, the authors overcame the technical limitation that high-throughput sequencing platforms need more material than obtained from a single cell (4). The next milestone was reached in 2011 by Islam *et al.*, who introduced early multiplexing **(Fig. A(i))** (5). This is the principle of incorporating barcodes in the form of unique nucleotide sequences into the cDNA during the first reverse transcription reaction of sequencing library preparation. For a given cell, all transcripts receive the same barcode, but barcodes are different between cells. Therefore, following the barcoding the material from multiple samples can be pooled for further processing in a single-tube reaction, which dramatically reduces the amount of labor and increases the amount of input material for subsequent enzymatic reactions in library preparation (2, 5). The barcode information is eventually retrieved during sequencing, and allows to sort transcript reads based on cell of origin. Since then multiple scRNAseq protocols were developed, including improvements such as introducing unique molecular identifier which allow to correct for amplification bias (5-7), or exploring different DNA amplification strategies (5, 8) (reviewed in (3)). However, it is 2015 that marks an explosion in the number of cells that could be sequenced in a single experiment (**Fig. A(ii)**)(2). This involved solving the challenge of efficiently coupling individual cells with unique sets of barcodes. In the original application of cell-barcoding by Islam *et al.* this was performed in 96-well plates, but this approach was not scalable. Several methods for high-throughput single-cell barcoding were published in 2015: two of them used microwells (9, 10), and two used droplets (1, 11) as reaction containers. A common feature was the immobilization of barcoding reverse transcription primers onto beads. This facilitate the creation of a wide barcode diversity through combinatorial split-and-pool synthesis (1, 11, 12). It also simplified the challenge of coupling a cell with a set of primers to one where a cell needs to be coupled with a bead. The following section describes how this is achieved using inDrops.

3.2.3. Barcoding in drops

Barcoding, or *in*dexing, of single cell transcriptomes with inDrops happens in *drops* (**Fig. 1 of Paper I**). The choice of drops as reaction containers made the

approach scalable and high-throughput. There is no fundamental constrain on how many droplets can be made, and barcoding droplets are made at 100 Hz. Even higher speed should be expected with further method development as in simpler setups droplet microfluidics allows droplet generation at speeds exceeding 10s of kHz (13).

The cell suspension, RT reagents, and barcode-bearing beads are injected into a microfluidic device, where they meet to form the aqueous phase which is split into droplets by a continuous flow of fluorinated oil containing a surfactant which stabilizes droplets. The resulting emulsion is collected in a standard 1.5 ml tube. The barcoding primers are attached to the beads by photocleavable linkers, which allows their release following encapsulation for an efficient homogeneous reaction, an aspect that turned out to be of crucial importance during the development of inDrops (1). After primer release, cell lysis and reverse transcription is initiated by transfer of the emulsion to 50 ℃. Following RT, the emulsion is broken, and further library preparation is performed.

Apart from the scalability offered by droplet microfluidics, a second strength of inDrops is the use of deformable hydrogel beads to deliver barcode-bearing primers. The squishy nature of such polyacrylamide beads allows to align them in the microfluidic channel and synchronize their arrival to achieve a nearly 100% occupancy of droplets with a single hydrogel bead. This means that every cell entering the microfluidic device receives a barcode. Cell delivery, by contrast, is Poissonian, and to avoid multiple cells entering one droplet their concentration needs to be kept at 1 cell in 10 drops. This is in contrast with Drop-Seq, the second powerful droplet-based scRNAseq method published back-to-back with inDrops in the same issue of Cell in 2015. In Drop-Seq, both bead and cell arrival is random (11). As a result, the majority of cells are lost, making the Drop-Seq unsuitable for samples of limited size.

The library preparation is based on the Cel-Seq protocol, which is distinguished by the use of *in vitro* transcription (IVT) as a strategy for amplification of cDNA (14, 15). Briefly, following RT the second cDNA strand is synthesized and serves as a template in an IVT reaction using the T7 promoter which is introduced together with barcodes (**Fig. 3 of Paper I**). The resulting amplified RNA is fragmented and reverse transcribed. Finally, PCR is used to introduce Illumina sequencer-compatible adaptor sequences at the ends of library molecules, yielding a sequencing-ready library.

3.2.4. Combinatorial barcoding of hydrogel beads

Figure 2 of Paper I summarizes the steps of hydrogel bead synthesis. The first part is the emulsification of a solution of acrylamide/bis-acrylamide and a

universal acrydate-modified "stub" primer. The resulting emulsion is polymerized off-chip, resulting in monodispersed hydrogel beads (HBs). The typical synthesis scale is in the millions of HBs.

The second part is combinatorial barcode synthesis involving two rounds of primer extension. First, hydrogel beads are distributed into 384 wells, each containing a barcoding oligo which anneals to the stub. Each well contains a different barcode. Following primer extension, beads from all wells are pooled together. The procedure is repeated with a second set of 384 barcodes, yielding a barcode diversity of 384^2.

3.3. Airway epithelium-specific background (Story II)

3.3.1. Mucociliary clearance and related pathologies

Mucociliary clearance (MCC) is the primary innate protection mechanisms of the airway. Key components are the viscous mucous loaded with antimicrobial compounds, a layer of liquid underneath it, and cilia, which are specialized organelles of ciliated cells within the airway epithelium. Cilia beat in an organized manner and constantly proper mucus with entrapped particles and pathogens from distal airway toward the mouth. The mucus is eventually either swallowed or expectorated. Defective MCC occurs pathological conditions like primary ciliary dyskinesia (PCD), chronic obstructive pulmonary disease (COPD), asthma, and cystic fibrosis (16).

PCD is a rare genetically heterogenous recessive disorder characterized by stiff, uncoordinated and/or ineffective cilia beat, which leads to compromised mucous removal. Symptoms include year-round wet cough and nasal congestion, infections of the airway, and infertility, the latter in particularly pronounced in men (17).

COPD is characterized by mucus overproduction, which leads to persistent airflow limitations. Mucus hypersecretion is a consequence of smoking, viral and bacterial infections, and inflammation (18). COPD is predicted to become the 3rd leading cause of death in 2030 by the World Health Organization (19). The disease is also characterized by tissue remodeling, including a disbalance in the cellular composition of the airway (20). As a specific example, goblet (mucus-secreting) cell hyperplasia and hypertrophy are observed in COPD (18). Another example is squamous metaplasia, where multiple layers of basal cells are observed, as opposed to a homeostatic single layer (21).

Cystic fibrosis (CF) is a rare life-limiting autosomal recessive disease. It affects multiple organs, with most of its morbidity today associated with the

airway (22). Historically, it was first described as a separate disease by Dorothy Hansine Andersen (23). Another milestone was reached in 1989 when 3 groups reported the identification of the specific gene mutation, causative of most CF cases: the deletion of 508th phenylalanine in a chloride channel, which authors called the cystic fibrosis transmembrane conductance regulator (CFTR) (24-26). Briefly, the pathophysiological cascade of cystic fibrosis in the airway involved the following steps: defective CFTR prevent ion transport into the airway lumen, concomitantly affecting the osmotic gradient and reducing water transport into the lumen. The lack of a water layer separating the mucus and the epithelium causes cilia to collapse compromising MCC. Next is a vicious cycle of airway obstruction by mucus, infection and inflammation (27, 28).

Despite the organ-level effects of CF being well studied, and the causative gene being identified, the localization of CFTR in the airway has been controversial. It has been reported to be mainly expressed in submucosal glands, with only a few CFTR-enriched cells in the surface epithelium (29). By contrast, another study determined the channel to be mainly present in surface epithelium ciliated cells with isolated single-cell enrichment (30). Story II of this thesis describes our insight into the identity of the CFTR expressing cell.

3.3.2. Cellular composition of the airway epithelium

The mammalian respiratory system is composed of two major compartments: the distal alveoli, where gas exchange occurs, and the proximal conducting airway, which hydrates the air and provides a first line of defense through mucociliary clearance (31, 32). The upper conducting airway is covered by a pseudostratified epithelium, where all cells rest on the basement membrane but not all reach the airway lumen. Those which do are called luminal, and those which do not are referred to as basal. Luminal cells are composed of ciliated, secretory, neuroendocrine, brush cells (21, 32, 33), and, as we now know from Story II of this thesis, rare pulmonary ionocytes.

Basal cells are the stem cells of the niche which self-renew and give rise to the luminal cell type. Across different epithelia they are generally thought of as double positive for cytokeratins (Krt) 5 and 14 (34). However, in the airway epithelium, basal cells are heterogenous in Krt5 and Krt14 expression. Only 20% of basal cells are Krt14^{+} in steady state, but following injury Krt14 expression is upregulated and all basal cells express it (35, 36). In their study using lineage tracing of Krt14^{+} cells in different conditions, Ghosh *et al.* concluded that under homeostasis Krt14^{+} basal cells are unipotent and only give rise to other basal cells. By contrast, following injury, Krt14^{+} cells were direct progenitor to club and ciliated cells (36).

Secretory cells are a heterogenous populations that secretes components of the protective respiratory tract fluids, including a variety of macromolecules, ions, and water (37). In the trachea and larger bronchi of mice, the main secretory cells are so-called club cells (32). It is appreciated that club cells are heterogeneous, e.g. as in their expression of secretoglobins SCGB3A1 and SCGB3A2 (38). Goblet cells are a type of secretory cells which is rare in laboratory mice grown in a pathogen-free environment but relatively abundant in human airway. The hallmark of goblet cells is mucous secretion and high production of the mucin MUC5AC, although both features also apply to club cells (33). Goblet cell numbers increase in response to inflammation, allergens and toxins (21).

As their name suggests, **ciliated cells** have cilia on their apical surface which perform the critical function of moving the liquid and mucous out of the airway (39). Under homeostasis, ciliated cells differentiate from club cells (40). However, as mentioned above, a direct basal-to-ciliated differentiation has been observed following injury (36).

Pulmonary neuroendocrine cells (PNECs) are rare innervated epithelial cells which secrete neuropeptides in response to stimuli, such as hypoxia and nicotine (32). PNECs modulate immune responses (41), and their secreted compounds are believed to regulate epithelial cells as well as the flow of blood and air. They potentially transmit stimuli to the central nervous system (32).

Brush cells are another rare epithelial cell type. Their name is based on the appearance of these cells that have a brush of microvilli (42). Brush cells have a chemosensory function and use a canonical taste transduction cascade to sense "bitter-tasting" irritants, and respond by releasing acetylcholine which elicits protective respiratory reflexes (43, 44).

3.4. Tumor-infiltrating myeloid cell-specific background (Story III)

3.4.1. Myeloid cells, their subsets, and functions

The distinction of myeloid and lymphoid cells. Traditionally, blood cells have been classified into two broach categories: lymphoid cells, composed of T, B, and natural killer (NK) cells, and myeloid cells, composed of polymorphonuclear (neutrophils, eosinophils, basophils) and mononuclear (monocytes, macrophages, dendritic cells [DCs]) cells (45). A broader definition of myeloid cells also includes megakaryocytes and erythroid cells (46). All blood cells are derived from a common hematopoietic stem cell (HSC), which sits at the top of a differentiation hierarchy, involving multi-, oligo- and unipotent progenitors with increasing lineage restriction. Under a classical model of hematopoiesis, the first lineage restriction is represented by a branching into

common lymphoid vs common myeloid progenitors (47), providing an ontogeny-based justification for the two broad categories. However, recent studies have challenged this view, providing evidence for alternative early hematopoietic hierarchies (48, 49). In this thesis, the term "myeloid cells" is used in the narrow sense and refers to mono- and polymorphonuclear phagocytes, without attempting a justification by ontogeny.

Macrophages are highly phagocytic cells that are involved in the clearance of dead and damaged cells. Through their cell-phagocytic activity they are also involved in tissue development and remodeling, e.g. during wound healing. Macrophages also play and important role detection and clearance of pathogens, and regulation of inflammation, from induction to resolution (50). By developmental original, macrophages are divided into monocyte-derived and tissue-resident. In 1968, van Furth and Cohn performed experiments showing that peritoneal macrophages were derived from blood monocytes (51), leading to a long-lasting perception that monocytes are precursors to macrophages found in tissues. While this is true, especially in response to insults including tumors (45), most macrophages found in tissues have an embryonic origin with replenishment relying on longevity and self-renewal, a now established understanding since the 2000s (52).

Tissue-resident macrophages adopt tissue-specific functions. For example, microglia cells responsible for synaptic remodeling, osteoclasts resorbing bones, and Kupffer cells clearing the blood of old erythrocytes in the liver are all examples of tissue-resident macrophages (52). Defining a nomenclature that would take into account the multiple dimension of macrophage diversity is an ongoing effort (53, 54).

Macrophage diversity is challenging to address, as plasticity in response to external stimuli is a hallmark of these cells (55). This plasticity is an additional layer of complexity to tissue-specific functions that tissue-resident macrophages adopt. Two best known macrophage states acquired in response to external stimuli are **classically** and **alternatively** activate macrophages and deserve a separate comment. The term "activate macrophage" dates back to the 1960s and the work of George Mackaness, who is attributed with coining the term, although according to his contemporaries he preferred the term "angry macrophage" (56, 57). The term was used to described macrophages with enhanced capability of mounting an immune response against intracellular pathogens. It was in 1992 that "activated macrophages" were renamed "classically activated", to accommodate the newly observed "alternative activation" (58). In 2000s, the term M1 and M2 were introduced to refer to classically and alternatively activated macrophages, respectively (59). This nomenclature mimicked that of T-helper (Th) cells and the

type of immunity they mediate through the secretion of cytokines. The main Th1 cytokine is interferon gamma (IFN-γ), which promotes cell mediated immunity against intracellular parasites and cancer cells. Th2 cytokines include interleukins 4, 5, 10, 13, 25 which promote anti-helminth immunity and promote tissue repair (60, 61). Importantly, "pure" M1 and M2 phenotypes are obtained *in vitro*, by exposing macrophages to IFN-γ and IL-4/13, respectively (62, 63). It is now well appreciated that this is an oversimplification of macrophage diversity *in vivo* in different disease contexts (53, 54, 62). I suspect that the M1/M2 paradigm will however retain its place in immunology: it exemplifies the degree of macrophage plasticity and represents two highly-relevant types of immune response.

Monocytes are cells mainly found in circulation, and are divided into inflammatory and patrolling subset. Inflammatory monocytes, also referred to a classical, are those that invade sites of insult and differentiate into macrophages. Patrolling, or non-classical monocytes, as their name suggests patrol the vasculature and perform surveillance of the endothelium (64).

There are two major subsets of **dendritic cells**: plasmacytoid and conventional (**pDC** and **cDC**). Classical dendritic cells are the archetypal professional antigen-presenting cells, which have a superior ability to prime naïve antigen-specific T cells. They are specialized to sample antigens in tissues and migrate to draining lymph nodes, where they mount specific T-cell immunity or tolerance to self-antigens (65). **Plasmacytoid DCs** are predominantly found in blood and lymphoid tissues. They show lower expression of molecules involved antigen presentation, and seem to be specialized at detecting viral infection, in response to which they produce large amounts of interferon alpha which induces an antiviral state in surround cells (65, 66).

Classical DCs are further divided into **cDC1** and **cDC2**. The former present a great capacity to cross-present antigens to cytotoxic T-cells, a process important in fighting intracellular pathogens such as viruses, and eliminating cancerous cells (67, 68). The latter have the complementary specialization of presenting antigens to helper T-cells, a mechanism targeted at extracellular pathogens (68).

The remaining myeloid cells are collectively referred to as polymorphonuclear cells, or granulocytes. The most abundant of them are **neutrophils**, which in general are the dominant cell type among healthy blood non-erythrocyte cells (50). They normally reside in the blood, are first-responders to injury and pathogens, and rapidly swarm to sites of insult, a process remarkably visualized *in vivo* using intravital microscopy in work by the Germain group (69). Today, apart from their potential to kill internalized and extracellular pathogens using the content of their granules, neutrophils are appreciated to play key roles

in inflammation, tissue repair, and homeostasis (70). It should be noted that neutrophils are also fierce phagocytes, an observation that dates back to the time of Elie Metchnikoff who is accredited from defining phagocytosis in the first place at the end of the 19th century. However, some researchers felt that neutrophils were later unfairly excluded from the professional phagocyte "club" (71).

The hallmark function of **eosinophils** is defense against helminths, with the direct killing mechanism involving degranulation of cytolytic granule content into the environment around the parasite (50). Together with **basophil** and **mast** cells, they are also important mediators of inflammation, and are implicated in allergic reactions (72).

3.4.2. On the role of myeloid cells in cancer

It is well appreciated that the immune system can control tumor by killing cancer cells or suppressing their growth. It can also promote tumor by selecting tumor cells that are best fit to survive and by establishing conditions that promote tumor growth (73). T cells play are major roles in these processes, and the understanding of their function in the context of cancer translated into efficient immunotherapies that benefit some patients (74). Myeloid cells remain less studied than T cells in the context of cancer. However, they present potential immunotherapeutic targets for a number of reasons. First, they are key player in eliciting a tumor antigen-specific cytotoxic T cell response, which requires cross-presentation of tumor antigens by antigen-presenting myeloid cells. This is in line with observations that the presence of the cross-presentation competent cDC1s is associated with increased patient survival and is critical for spontaneous tumor rejection (75). Second, myeloid cells mediate processes related to wound healing, which are hijacked by tumors to limit inflammation and support tumor growth (76). Third, even without assuming a particular mechanism, clinical data revealed a link between myeloid cell abundance in tumor biopsies and patient survival. The same myeloid cell types have been associate with both favorable and poor prognosis depending on the cancer type and the exact choice of markers to define that cell type (45, 77).

3.4.3. The KP tumor model

Studies described in Papers III and IV used a genetic lung adenocarcinoma model, where tumors are induced by activation of oncogenic Kras and deletion of Trp53 (therefore "KP")(78). A general strength of genetically engineered mouse models (GEMMs) is that they develop *de novo* tumors in a natural immune-proficient environment. This is in contrast with tumor cell line transplantation and xenograft models (79). GEMMs also have lower mutation load compared to

chemically induced models and therefore are less immunogenic (80). This can be viewed as a lack of resemblance to most human tumors, or as a model for a particularly challenging case. In their review, Politi and Pao (81) described an ideal GEMM of human cancer as featuring tumors that start from a small subset of cells within an organ by a genetic lesion found in the human counterpart and for which steps in tumor progression resemble those in human. The KP tumor model satisfies well this definition: it is induced in adult mice by adenoviral delivery of the Cre recombinase to specifically to lungs, and is driven by mutation to Kras and Trp53 as observed in patients. It also recapitulates features of the human non-small cell lung cancer (NSCLC) including lack of immunogenicity, and resistance to immuno- and chemotherapy (82). In Paper III and IV, we used a syngeneic orthotopic KP mouse model, with lung tumors initiated by intravenous injection of a KP tumor cell line. This variation of the model has been shown to share the same characteristics as the autochthonous version (82).

4. KNOWLEDGE GAPS ADDRESSED IN THIS THESIS

This section covers what knowledge gaps were addressed in each of the 3 stories of this thesis, explaining the scientific novelty of the work.

Story I: Accessibility and refinement of inDrops technology. Upon its publication in 2015, the inDrops technology for scRNAseq received immediate interest from the scientific community which was anticipating a scalable scRNAseq approach. However, the adoption of the inDrops technology remained challenging, as a detailed step-by-step protocol was not publicly available. Together with colleagues, I provided a detailed description of the approach, documenting the original procedure by dr. Klein and dr. Mažutis (1), and incorporating further improvements at multiple steps. I believe that this contributed to the technology being more readily adopted by academic laboratories, and provided a starting point for further development of the technology.

Story II: The discovery of pulmonary ionocytes and their characterization. This study involves a truly serendipitous discovery of a cell type in the airway epithelium, therefore some of the knowledge gaps addressed in this study are appreciated in retrospective.

In entering this project, the upper airway epithelium was already a well-studied biological system, known to be composed of specialized cell types derived from a transcriptionally heterogeneous basal cell population. This prior knowledge was based on evidence gathered by studies using diverse techniques, including microscopy, lineage tracing, and injury models (21). Such techniques could have overlooked cell types lacking a distinct morphology and/or known markers. Therefore, a truly comprehensive and unbiased view of cell types was unavailable. The lack of such a catalogue also prevented a better understanding of how the airway epithelium is restored after injury, a mechanism which could be harnessed for treatment of airway pathologies. Therefore, we performed scRNAseq on human and mouse upper airway epithelium to provide a detailed catalogue of cell types and their dynamics during airway recovery. Unexpectedly, we identified a previously unappreciated cell type: pulmonary ionocytes. We further characterized ionocytes, showing that they are the main source of CFTR activity and gaining insight into their specification.

The discovery of pulmonary ionocytes can also be framed in the context of cystic fibrosis research. At the protein level, compromised function of the ion channel CFTR has been known to be the cause of cystic fibrosis for two decades. At the physiological level, the airway is one of the most heavily affected organs.

However, despite extensive research in the field, the cell type expressing CFTR was not established. Therefore, from the point of view of cystic fibrosis research, our study provided are missing level of resolution: at the cellular level, the CFTR-expressing cell type was identified.

Story III: Comparison of human and mouse lung tumor-infiltrating immune populations. The success of current immunotherapeutic approaches justified the promise that modulating the immune system can be used as a cancer treatment strategy. Major efforts are being made to further elucidate how tumor and immune cells interact and create new treatment strategies. Both mechanistic studies and drug research involve experiments in mouse models and rely on the conservation of human and mouse immune systems. However, some aspects of the immune system are species-specific, preventing discoveries to be translated from mouse to human. It is reasonable to anticipate that a drug targeting a cell population present in mouse but not human would fail to translate. Therefore, we sought to identify conserved and divergent immune cell types in a systematic way. A prerequisite for such a comparison is an unbiased characterization of the immune cell populations in each of the species, which we achieved by scRNAseq. We chose to focus on the less well characterized myeloid cells, the diversity of which is highly context dependent. Our study serves as a starting point for further research focusing on those cell types and pathways that are conserved in human and mouse. It is also a resource which can be used, for example, to verify if a gene target being studied in mice is human-relevant and is expressed on the expected cell type.

5. GOALS AND OBJECTIVES

The knowledge gaps discussed in the previous Chapter were addressed by defining the following goals and objectives.

Goals:

I. To facilitate access to the inDrops technology (Story I).

II. To provide an unbiased view of the cellular composition of the airway epithelium under homeostasis and during recovery (Story II).

III. To identify conserved and divergent lung tumor-infiltrating myeloid cell populations in human and mouse (Story III).

Objectives:

Story 1

i) To identify and correct the source of variability between inDrops experiments.
ii) To prepare a detailed step-by-step protocol describing inDrops.

Story 2

i) To provide an unbiased atlas of cell types in airway epithelium using scRNAseq.
ii) To describe the dynamics of the cellular composition of airway epithelium during recovery.

A third objective was raised during the course of the study after observing pulmonary ionocytes as a novel cell type, and that they are the main source of *CFTR* transcripts in the airway epithelium:

iii) To further characterize pulmonary ionocytes in terms of localization, specification, and function of the CFTR ion channel.

Story 3

i) To provide an unbiased atlas of non-small cell lung tumor-infiltrating immune cells in human and mouse.
ii) To identify myeloid cell population conserved in human and mouse.

iii) To provide a resource that could be readily accessed and used by a broader community.

6. RESULTS

6.1. Story I: availability and ease of use of inDrops

InDrops was not readily accessible following its original development

As I was about to start my PhD in summer 2014, dr. Linas Mažutis taught me about an exciting method that he had developed with dr. Allon Klein, and that was being prepared for publication. The method, called inDrops but nameless at the time, was a solution for single cell RNA sequencing (scRNAseq). Back then, the idea alone of working with the minute amounts of mRNA from single cells fascinated me, as my background in biochemistry allowed me to appreciate the challenge that this involved. Upon learning more about the state of the field and previous success in sequencing transcriptomes of individual cells (4, 5, 7, 15), I realized that the real strength of inDrops was in the throughput that it offered, allowing to barcode tens of thousands of cells in a single experiment. Mastering all the steps of the method, attempting improvements, and making it widely accessible completed my acquaintance with inDrops, as described in this Chapter. It took me another 5 years to appreciate first-hand the power of the technology once put in action (Chapters 6.2 and 6.3).

By the time inDrops was published in 2015 (1), interest in the technology was already apparent, and only grew with the release of the publication. It soon became clear that a detailed public protocol would facilitate the spread of the method. Moreover, refinements to the original protocol were already being made and deserved to be published. While minor updates were numerous, I decided to highlight the following problems we had to solve:

a) The efficiency of the primer extension step during combinatorial barcode synthesis on hydrogels (Section 3.2.2. and **Fig. 2c of Paper I**) is ~60%, over two steps yielding 35-40% of full-length barcoded primers. Apart from reagent loss, this also lead to the presence of incomplete primers in the single cell transcriptome barcoding reaction. These by-products did not cause noticeable problem with mouse and human cells. However, in a study using Xenopus frog cells (83), the incomplete primers primed rRNA, which is in high excess to poly-A RNA (84), preventing the use of the technology as is.
b) We observed a lack of reproducibility in the early steps of inDrops with reverse transcription failing to produce cDNA.

Exonuclease I treatment removed side products of barcode synthesis

An ideal solution for the first problem would have been to increase the efficiency of primer extension (**Fig. 2c of Paper I**). However, unless the efficiency was close to 100%, side products would still present a problem. We therefore focused our efforts on cleaning up barcoding hydrogel beads (BHBs) from incomplete primers. We made use of the fact that all full-length primers had a poly-T tail at their 3'end (**Fig. 2c of Paper I**). After annealing a poly-A protective oligo, BHBs were treated with Exonuclease I (ExoI), which digests single-stranded DNA in the 3'->5' direction but has no activity on double stranded DNA. Success removal of incomplete primers was confirmed by capillary electrophoresis (**Fig. 6A,B of Paper I**). The clean-up approach was suggested by Allon Klein, first applied by James Briggs on a small batch of BHBs, and I then incorporated it in the standard barcode synthesis procedure.

Improving the efficiency of primer extension in barcode synthesis remains to be addressed. Possible directions to explore include adjusting the pore size of the hydrogel mesh and using different polymerases.

Adjusting magnesium concentration lead to reproducible and more efficient reverse transcription

The second problem causing reverse transcription (RT) to occasionally fail turned out to be easily addressable by adjusting the reaction composition. Using the exact conditions described in (1), the final concentration of magnesium ions in the barcoding droplet was <1.5 mM, while dNTPs which could chelate them were in 1.6x excess. As shown in **Fig. B**, the optimal Mg^{2+} concentration was determined to be 5 mM. The problem of insufficient magnesium was more severe when preparing and freezing a pre-mix of dNTPs and RT buffer (source of magnesium) in advance as opposed to mixing reagents fresh.

A detailed protocol was published

Finally, the lack of a detailed public protocol was addressed by describing the procedure in a Nature Protocols publication (Paper I, (12)).

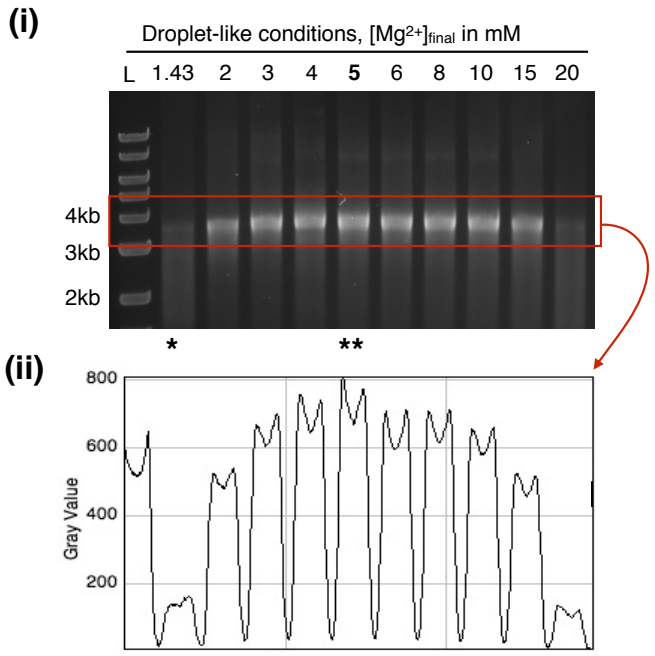

Figure B. $[Mg^{2+}]_{final}$ of 5 mM is optimal in a bulk reaction mimicking conditions in droplets. (i) Products of a RT reactions using a 3.4kb polyadenylated RNA as input and different concentrations of Mg^{2+}. 1xTAE agarose gel. L – ladder (NEB N3200S). (ii) Plot showing the cumulative white color intensity along the y axis in the red rectangle. *Original conditions described in (1). **Optimal $[Mg^{2+}]$ reported in Paper I.

6.2. Story II: discovery and characterization of pulmonary ionocytes

The understanding of the cellular composition of airway epithelium could have been incomplete

Despite extensive prior knowledge on the biology of the airway, restoring its function in pathological conditions remains a challenge (see also Chapter 3.3). Progress could be hindered by failure to appreciate the full diversity of cell types in the airway epithelium. The advent of scRNAseq motivated us to revisit the cellular composition of the airway epithelium in a comprehensive and unbiased manner.

To this end, we performed single cell RNA sequencing of 2,970 primary human bronchial epithelial cells (HBECs) from 3 donors. Prior to scRNAseq, cells were expanded and differentiated *in vitro* in an air-liquid interface (ALI) culture. We also profiled 7,662 freshly dissociated mouse tracheal epithelial cells from mice under normal homeostasis (n=3) as well as in a recovery time course following injury by polidocanol inhalation (6 mice, 4 timepoints). This perturbation eliminates all but basal cells in the airway epithelium (85). This choice of experimental design enabled mechanistic perturbations in mice while retaining the focus on findings consistent across species.

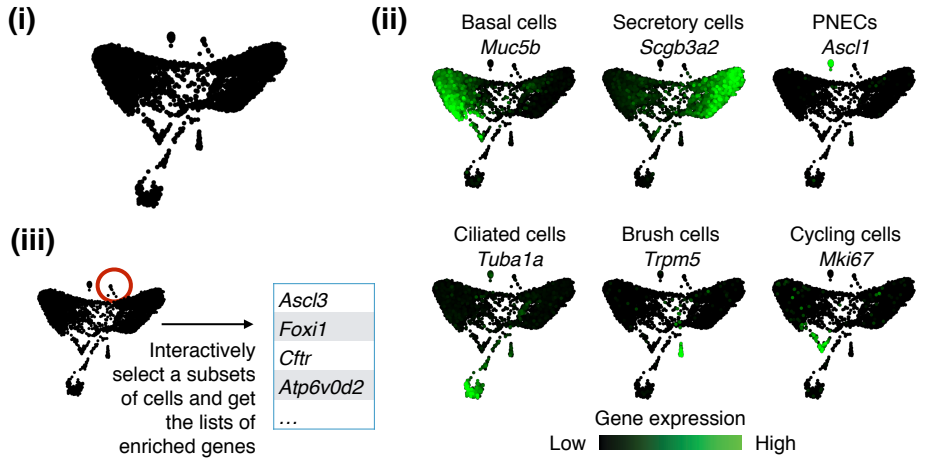

Figure C. Interactive exploration of single cell transcriptomes. (i) 2D representation of the multidimensional gene expression data using SPRING (86). Each dot is a cell. Cells cluster together by gene expression similarity. (ii) Knowledge-driven exploration: SPRING plot colored by canonical marker gene expression for previously described cell types. Literature references for marker genes: Scgb3a2 (38), Krt5 (21), Tuba1a (87), Trpm5 (43), Ascl1 (41). (iii) Example of data-driven exploration: lists of enriched genes in a selected group of cells can readily be accessed. A small satellite cluster in the data was enriched for genes previously reported in ionocytes in Xenopus frog and mouse kidney (88). PNEC – pulmonary neuroendocrine cell.

An interactive representation of the scRNAseq data helped build intuition

After initial cleanup to remove low quality transcriptomes, the high dimensional single cell gene expression data (10,000s genes x 1,000s cells) was visualized in a 2D representation, allowing an intuitive interpretation of the populations formed by single cell transcriptomes. Here, we used a method called SPRING (86), which involves calculating a k-nearest graph (k=10) of cells and projecting it in 2D using a force layout. Both in mouse and human data, the majority of cells formed a continuum with a few small satellite populations (**Fig. C(i)**). The next step was to define populations we observe. It started with an interactive exploration of the data, enabled by a browser-based interface available as part of the SPRING toolkit. **Fig. C(ii)** shows how exploring the expression of previously known marker genes reveals the identity of populations observed in the data. Exploration of the initial map was also performed in a data-driven manner, by finding genes enriched in selected cell groups relatively to the rest of the data (**Fig. C(iii)**), as a way of asking cell who they are. I believe that his exploration is a critical step of the scRNAseq data analysis, and my experience has been that the most salient features of a dataset are already appreciated at this stage. The reader of this thesis is invited to explore the interactive online version of the plot[1].

Having built an intuition for the population structure, we drew boundaries between cell states by spectral clustering. This is a graph-based partitioning method where the number of clusters is user defined. Therefore, spectral clustering alone cannot be used to find the "real" number of populations. Rather, we regard clusters as a convenient way to communicate observations in the data.

Two transient cell populations appear during airway recovery

Having defined the population structure under homeostasis, we asked how it changes during recovery following injury by polidocanol. We started by combining scRNAseq data from uninjured and recovering mice in a single 2D representation, which revealed two recovery-specific states: 1) recovery-specific cycling basal cells, which made up 84% of all epithelial cells 1 day post-injury (dpi), and 2) an alternative ciliated cell differentiation trajectory, that appeared 2 and 3 dpi (**Fig. 2b-d of Paper II**). Both states could be related to previous observations, and our data provides detailed molecular descriptions of these states.

---

1 kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?datasets/uninjured_MTECs/uninjured_MTECs.

Under homeostasis, Krt14 expression is restricted to 20% of airway basal cells while most express Krt5. However, with injury Krt14 expression is upregulated (35, 36). We recapitulated this observation in our data and extended it to other cytokeratins: recovery-specific cycling basal cells showed upregulation of Krt14, 8, 4, 13, which were co-expressed with Krt5 (**Fig. 2d of Paper II**). Krt8 is considered a luminal cell marker (89), and Krt4 and 13 were enriched in a subset of homeostatic basal cells which we observed in our data but had not been described before.

The second recovery-specific state we observed was a group of cells transiting directly from basal to ciliated cells, referred to as pre-ciliated. By contrast, under normal homeostasis cells pass through a secretory intermediate. The presence of two trajectories can be readily appreciated in the SPRING plot combining data from uninjured mice and those undergoing recovery (**Fig. 2b of Paper II)**. The direct conversion between basal and ciliated cells has been reported before in the context of injury (36). Here, with transcriptome-wide gene expression data in hand, we explored transcriptional changes along this trajectory. The analysis involved designing a procedure for identifying genes variables along the trajectory. In the following three paragraphs, I chose to provide more details on this particular analysis.

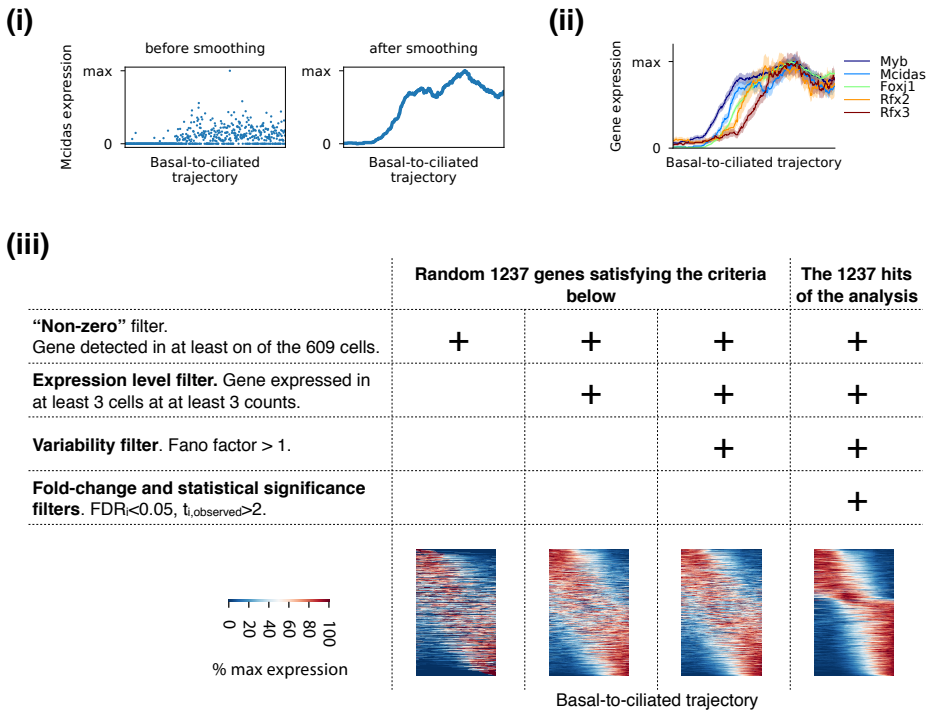

Figure D. Analysis of the basal-to-ciliated trajectory observed during recovery. (i) *Mcidas* (involved in ciliated cell specification) expression in single cells ordered by basal-to-ciliated trajectory before and after smoothing. (ii) Table showing the effect of individual filtering steps on the selection of genes variable along the basal-to-ciliated trajectory. Rightmost heatmap is the same as shown in Fig. S6b of Paper II. (iii) Smoothed expression of selected ciliated cell specification factors along the basal-to-ciliated trajectory.

Genes variable along the pre-ciliated trajectory are candidate ciliated fate specification factors

The principle of the analysis is detailed in **Fig. S6 of Paper II**, and the main steps, except variable gene selection, had been described before (49). It starts with ordering of cells along the trajectory. We used the ordering approach which is part of Population Balance Analysis (90), but other methods (91) are expected to give similar results, especially for a simple trajectory with no branching points. Next, the expression of each gene was smoothed by averaging over a sliding window of a chosen size (100 cells in this study). **Fig. D(i)** shows how smoothing mitigates the noisy nature of scRNAseq data. At this point, the expression of genes of interest could be easily explored along the trajectory. For example, Foxj1, Myb, Mcidas, Rfx2 and Rfx3 are established ciliated cell specification factors (**Fig. D(ii))** (87). Our analysis enabled us to appreciate the order in which their transcripts appear during differentiation, allowing to hypothesize on the regulation of ciliated cell specification.

We chose to limit the analysis of ciliated cell differentiation to providing a resource: the list of genes which are variable in a non-trivial way along the basal-to-ciliated trajectory (**Table S4 of Paper II**). Such genes could potentially play a role in ciliated cell specification. First, the full gene list was pre-filtered on expression level and variability but without taking into account their ordering. Next, for each surviving gene, we asked the question whether the cell ordering along the trajectory is special in any way. More specifically, for each gene *i* we calculate a statistic $t_{i,observed} = \frac{m_{i,max}+100CPM}{m_{i,min}+100CPM}$, where m_i is the expression vector of gene *i* in smoothed cells along the trajectory (CPM – counts per million). A distribution of $t_{i,random}$ was obtained by repeating the procedure on cell ordered in a random way for multiple permutations, allowing the calculation an empirical one-sided p-value p_i, defined as the fraction of times $t_{i,random} > t_{i,observed}$. To account for multiple hypothesis testing, the Benjamini-Hochberg procedure was used to obtain false discovery rates (FDR). Genes with $t_{i,observed}>2$ and $FDR_i<0.05$ were considered as differentially expressed along the trajectory.

The expression of 1,237 differentially expressed genes we identified is shown in **Fig. S6b of Paper II** and the rightmost plot in **Fig. D(iii)**. Genes are ordered by the expression-weighted mean position along the trajectory. From a data representation point of view, such gene ordering results in gene expression peaking along the diagonal of the heatmap. While being convenient in providing a neat representation, this ordering approach will always result in a diagonal, even for randomly selected genes. **Fig. D(iii)** shows how individual filtering steps used for differentially expressed gene selection affect the final heatmap, support the idea that observing a diagonal alone does not necessarily mean that gene shown are statistically significantly variable. We hope that the list of 1,237 genes variable along the basal-to-ciliated trajectory will serve as a resource in research of ciliated cell specification.

The homeostatic airway composition is restored within a week

Injury by polidocanol caused a major disruption of the tissue composition. As expected, 1 day post-injury (dpi) almost no luminal cell types were observed. Basal cells accounted for 97% of all epithelial cells instead of the homeostatic 44%. However, by day 7, relative population abundances had returned to those observed under homeostasis (**Fig. 2c of Paper II**). Recovery specific cycling basal and pre-ciliated cells peaked at 1 and 3 dpi, respectively, but were almost completely absent by day 7 (**Fig. S5b of Paper II**). The number of secretory cells grew gradually with ¼ of their homeostatic abundance restored at 2 dpi, while

ciliated cells were still virtually absent at 3 dpi. By 3 dpi, the rare PNECs, brush cells, and ionocytes (discussed below) were all observed.

We observed a new cell types and called it pulmonary ionocytes

Observing the expected basal, ciliated, secretory, neuroendocrine and brush cells built confidence in our results. However, the data also revealed an unexpected new cell type, observed in both species, and appearing in the data as a small detached cluster (0.27% and 1.5% of epithelial cells in mouse and human, respectively). It was characterized by the specific expression of the transcription factors FOXI as well as genes involves in ion transports. This expression profile was reminiscent of ionocytes previously described in other contexts, such as mouse kidney and inner ear (92) or Xenopus frog larvae skin (88), where they are responsible for pH and osmoregulation. We called these cells pulmonary ionocytes. Further characterization of pulmonary ionocytes was an additional aim we raised during the study.

Additional experiments validated hypotheses on ionocytes specification and function

When compared to other airway epithelial cells in our data, one of the most enriched genes in pulmonary ionocytes was *CFTR*, the gene mutated in cystic fibrosis (see also Chapter 3.3). Ionocytes, which were <0.3% of all epithelial cells in our mouse data, accounted for >50% of CFTR transcripts. The airway is one of the most affected tissues in cystic fibrosis, and the genetic cause of this life-limiting disease has been linked to CFTR mutations already back in 1989. Our data suggested that the activity of CFTR is localized to a rare specialized cell type in the airway. This provides cystic fibrosis research with the cellular level of resolution which was missing before.

Toward further characterizing pulmonary ionocytes, we fulfilled the following tasks:

1) Provided lists of enriched genes in human and mouse, which can be used as ionocyte markers or candidates for perturbations in further studying this cell type (**Figs. 1 c,d** and **Table S1,2 of Paper II**).
2) Localized ionocytes *in situ* by staining for FOXI1, which was specific for ionocytes as suggested by scRNAseq data. FOXI1 expressing cell were more concentrated in bronchial gland ducts than in surface epithelium as shown in **Fig. S7d.**
3) Demonstrated that Notch1 signaling is necessary for ionocyte specification.
4) Demonstrated that overexpression of the transcription factor FOXI1 is sufficient to drive FOXI1 production in HBECs cultured in ALI.

5) Showed that ionocytes are the functional source of CFTR activity in the airway epithelium.

For tasks 3) and 4), we hypothesized that Notch1 signaling and expression of FOXI1 would play a role in ionocyte specification in the airway as this had been to be the case of ionocytes described in other systems (88, 92).

For task 5), we sought to address the concern that the presence of mRNA does not necessarily mean protein function. Although highly specific to ionocytes, *CFTR* transcripts were also observed in other cells at low levels. As other cells are much more abundance, they could still contribute to a significant fraction of the overall CFTR activity. Lindsey Plasschaert used an electrophysiological assay (Ussing chamber) to measure the CFTR-specific ion current through the airway epithelium. She demonstrated that this current a) correlates with the number of ionocytes when considering the natural variation in ionocytes between donors; b) is undetectable when HBECs are grown under conditions preventing CFTR specification.

Tasks 2), 3), and 5) were performed exclusively by Lindsey Plasschaert and the Jaffe group. For task 4), HBEC culture and FOXI1 overexpression experiments were performed by L. Plasschaert, while I was responsible for scRNAseq experiment and data analysis. We observed, that FOXI1 overexpression by lentiviral transduction lead to a 23-fold increase in the number of ionocytes compared to the transduction by GFP, supporting the prediction that FOXI1 is sufficient for ionocyte specification.

Discussion

Providing a catalogue of airway epithelial cells and identifying cell population changes during recovery fell within our initial goal of the study. While the scRNAseq dataset provided a clearer picture of the airway cellular composition, follow up experiments are required to gain further mechanistic insight, which could eventually be used to manipulate the airway composition. In our study, we chose to focus on the newly identify pulmonary ionocytes, which were intriguing both for their novelty and potential relevance of cystic fibrosis. Pulmonary ionocytes were also identified by the Rajagopal and Regev groups, and reported in a back-to-back publication in the same issue of Nature as Paper II (93). Neither study suggests direct therapeutic solutions for cystic fibrosis. However, it is reasonable to anticipate that knowing which cell type is involved will help guide efforts in CF research.

6.3. Story III: identifying tumor-infiltrating myeloid subsets conserved in human and mouse

Myeloid cells are candidate immunotherapy targets but are poorly characterized in tumors

In this study, we performed scRNAseq to interrogate the unbiased population structure of myeloid cells in human and mouse lung tumors, identified conserved and diverging cell subsets between the two species, and made the data accessible at multiple level of processing, creating a resource for the broader community.

To better understand the motivation of this project the following questions have to be answered: Why did we focus on myeloid cells specifically? Why did we compare mouse to human? Why scRNAseq? Why lung tumors?

As direct killers of tumor cells, cytotoxic T cells play a central role in controlling tumors (94). Immune checkpoint-blockade (ICB) immunotherapies targeting inhibitory pathways on T cells resulted in durable tumor control but only benefit a fraction of patients. For example, tumor lacking T cell infiltration are resistant to ICB (95). Despite the focus on T cells, the tumor environment houses a wide diversity of immune cells (45, 95). Among them, tumor infiltrating myeloid cells (TIMs) represent an untapped pool of potential immunotherapeutic targets but remain relatively understudied compared to T cells. TIMs dominate the immune environment of multiple tumors, are key players in T cell activation and inhibition, and their presence has been associated with both pro- and anti-tumor activities (45, 77). A detailed description of the diversity of myeloid cells in cancer is instrumental to advance our understanding of their function in cancer.

At the time we were entering the project in fall 2015, immune cell diversity was mainly characterized using flow cytometry and histology, both of which are biased in the sense that they rely on the knowledge of discriminative molecular markers and the availability of antigens against them. It easy to speculate that aspect of cell diversity not reflected in known marker expression would be missed. Moreover, even distinguishing the major lineages (monocytes, macrophages, neutrophils) can fail in tumors, as the expression of conventional molecular markers can become localized to different cell lineage compared to normal homeostasis (96). The development of high-throughput scRNAseq enable us to interrogate the myeloid cell diversity at an unprecedented degree of molecular resolution in a truly marker-free way.

Apart from enabling the identification of transcriptionally distinct states, the information on gene activity provided by scRNAseq creates a powerful starting point for hypothesizing on the function of these states. The definitive

establishment of function and role in disease requires manipulations and mechanistic perturbations of cells, with the model organism for performing such experiment being mice. They are also instrumental in developing therapeutic approaches.

The conservation of the immune system between human and mouse is a major reason why this rodent is a model organism in the first place. Yet, differences are also appreciated. For example, the presence of arginase 1 in tumor stroma is known to play a tumor-promoting function, and could potentially present a therapeutic target. However, it is expressed by neutrophils and macrophages in human and mouse, respectively (97). Failure to appreciate such a discrepancy between species is likely to prevent the translation of approaches targeting specific cell types.

Finally, our choice to use lung cancer as a study case is motivated by the prevalence of the disease and its frequent resistance to existing treatment, making it the primary cause of cancer deaths (98). For the comparison to mouse, we chose an genetic orthotopic *Kras*$^{G12D/+}$;*Trp53*$^{-/-}$ (KP) non-small cell lung cancer model, which enables studying the progression of the disease in an uncompromised immune environment (see also Section 3.4.3).

ScRNAseq analysis of lung tumor-infiltrating immune cells in human and mouse was used to address myeloid cell diversity

Towards to goal of characterizing TIM diversity and their conservation between human and mouse, we:

i) Performed scRNAseq on freshly excised whole tumor biopsies from 7 NSCLC patients undergoing lung surgery.
ii) Performed scRNAseq on immune cells from lungs of healthy mice (n=2) as well as tumor bearing lungs from an orthotopic KP mouse tumor model (n=2).
iii) Defined cell states observed in scRNAseq data and relate them to previous knowledge.
iv) Identified which myeloid states had counterparts between the two species.
v) Made our data available as a resource for the broader community.

In the rest of this chapter, I briefly comment on the structure of Paper III, which entails a thorough description of the key observation we made by performing i)-iv). Next, I chose to highlight selected results, some of them only briefly mentioned in the publication. Special attention is given to neutrophils, which are also the focus of the study described in Paper IV. Finally, for our study to serve as a basis for further investigation of TIMs, both by us and the broader

community, the resource aspect (Task v)) of paper needed to be fulfilled, and I therefore chose to list the resource made available during this study.

In terms of Paper III structure, Figures 1 and 2 cover the experimental design, and, at the level of major lineages, help us build confidence in our ability to identify cell types and compare them between species. We observe that at our analytical framework correctly identifies the expected cell type conservation at the major lineage level. Figures 3-5 focus on the previously known and novel states identified within neutrophils, dendritic cells, and monocytes/macrophage, respectively. Our analysis revealed a clear interspecies correspondence between 4 DC and 4 monocyte subsets; a partial correspondence within neutrophils; and species differences among macrophages. Figure 6 demonstrates the utility of our data in finding uniquely enriched genes within subsets, the expression of which can be related to prognostic score by harnessing the available databases on patient survival. Such genes can also be used as novel markers for identification of populations *in situ*, as demonstrated with the gene PI3 (**Fig. 6G-J of Paper III**). Finally, Figure 7 introduces matched blood scRNAseq data from the same patients, revealing considerable differences between myeloid cell subsets across tissues. Without ruling out that liquid biopsies could be used to predict disease outcome, we conclude that there is a poor relationship between blood and tumor myeloid states. Our data shows that, at least with a median of 5283 tumor and 913 blood cells per patient, blood does not reflect the diversity of TIMs.

Results I chose to highlight are the observation of tumor-specific neutrophils with tumor-promoting properties and our attempts to relate the myeloid cell states we define to previous studies.

Tumors cause a radical chance in lung neutrophil population structure

At 2 pm on November 18, 2015, I started barcoding transcriptomes of the first set of samples from healthy and tumor bearing mice. These were immune cells, which, at a specific hour of the day to control for circadian gene expression fluctuation, had been quickly isolated from freshly excised mouse lungs by Camilla Engblom and Christina Pfirschke. By December, Virginia Savova had already prepared the data for exploration and we were having the first look at the unbiased structure of immune populations revealed by scRNAseq. It took us over a year before we finalized our analyses, including the detailed population structure, to the format presented in Paper III. However, already at the first look at the data, a comparison of healthy and tumor-bearing lung revealed a feature which immediately caught our attention.

We observed that the neutrophil population structure undergoes a radical change in tumor, including the emergence of a tumor-specific neutrophil

population (**Fig. E(i-ii)**). This observation was even more intriguing in the context of another, mechanistic, study, which at the time was being completed by the Pittet group and lead by C. Engblom and C. Pfirschke (described in Paper IV).

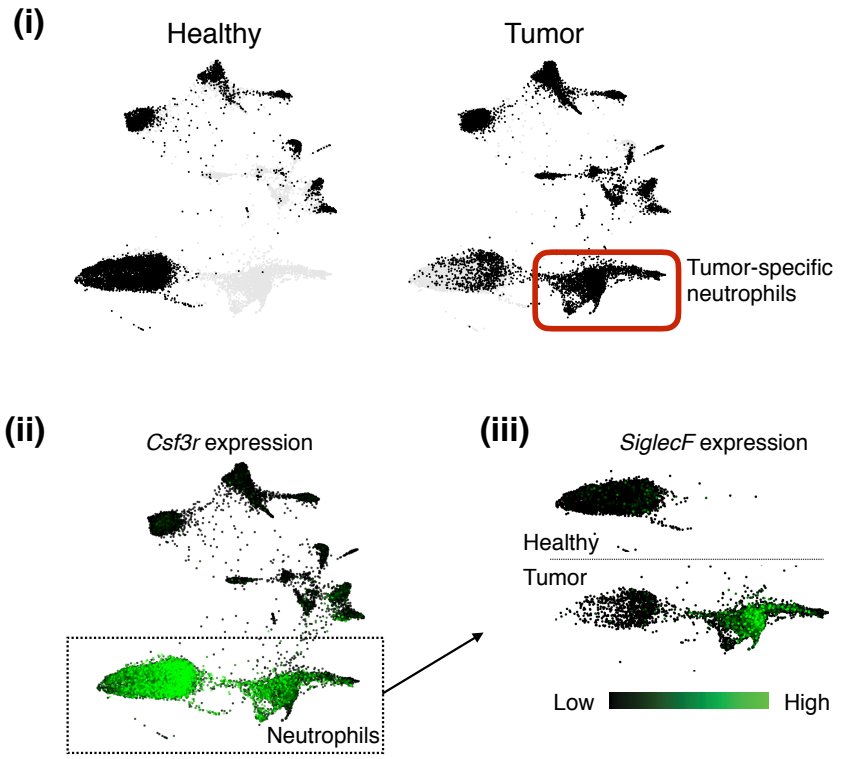

Figure E. Tumor-specific neutrophils are observed in tumors. (i) SPRING plot of all mouse data with healthy (left) and tumor (right) samples highlighted in black. (ii) Neutrophils are recognizable by the expression of the marker Csf3r. (iii) *SiglecF* expression in neutrophils from healthy and tumor-bearing lungs.

Osteoblast promote tumor growth via neutrophils

Paper IV reports on a systemic tumor-promoting positive feedback loop, where presence of lung tumors lead to increase osteoblast activity in human and mouse bones. These cells in turn remotely supply tumor-infiltrating neutrophils with cancer-promoting properties. The cross-talk also involves soluble osteoblast activity-promoting factors found in blood of tumor-bearing animals, this way closing the feedback loop. This study is supported by a fascinating diversity of experimental approaches, including but not limited to clinical human patient data analysis, the derivation of multiple mouse models, use of parabiosis, *in vivo* cell ablation, and manipulation of primary cells *in vitro*. None of these mechanistic experiments, which I mention for context, should be considered as part of this thesis as I did not contribute to them. My specific contribution was the

characterization of tumor-promoting neutrophils at the whole transcriptome level. I also related the gene expression of this cell population to human patient survival.

The tumor-promoting function is restricted to a SiglecF^{high} neutrophil subset

Around the same time that we observed a striking change in neutrophil population structure in scRNAseq data, the Pittet lab independently identified SiglecF as a cell surface marker discriminating the neutrophil subset specifically induced by osteoblasts. Under homeostatic conditions, SiglecF is considered to mark eosinophils and tissue resident macrophages (52, 99). Being a conventional marker, it was included in the flow cytometry analysis used to dissect the tumor-infiltrating immune cell types. However, in the presence of tumor, SiglecF was expressed on a subset of neutrophils. This is also reflected in scRNAseq data (**Fig. E(iii)**). In paper IV, Engblom *et al.* demonstrated that SiglecF^{high} neutrophil, and not those negative for the marker, were specifically responsible for the tumor-promoting function. My task was to validate the hypothesis that the tumor-promoting role would be reflected in the transcriptomes of SiglecF^{high} neutrophils.

The tumor-promoting function of SiglecF^{high} neutrophils is reflected in their gene expression

To this end, we first divided single neutrophil transcriptomes into SiglecF^{high} and SiglecF^{low}. Due to the noisy nature of scRNAseq data, *SiglecF* expression alone could not be used to partition cells. To compensate for that, we defined a gene expression signature composed of tens of genes positively and negatively correlated to *SiglecF*, and used it to partition neutrophils from healthy (H) and tumor-bearing (T) mouse lungs (**Fig. S19A of Paper IV**) into SiglecF^{high} vs SiglecF^{low}. Consistently with flow cytometry data, all neutrophils in healthy mice were SiglecF^{low} (designated as H-SiglecF^{low}) while two population were observed in tumor, designated as T-SiglecF^{low} and T-SiglecF^{high}. A comparison of the three groups revealed that T-SiglecF^{low} and T-SiglecF^{high} differed in the expression of over 1700 genes, while 123 genes were identified as differentially expressed between H-SiglecF^{low} and T-SiglecF^{low}. This shows that a) the gene expression differences between SiglecF-low and -high neutrophils are substantial and by far not limited to this one marker gene, and b) tumor does not affect the gene expression of SiglecF^{low} neutrophils and they represent the same cell states in both conditions. Next, we asked how the expression of genes known to be involved in angiogenesis, myeloid cell recruitment, tumor proliferation, extracellular matrix remodeling, and immunosuppression was distributed among the 3 groups of cells. As suspected, the expression of genes involves in these tumor-promoting functions was upregulated in SiglecF^{high} neutrophils (**Fig. 5B of Paper IV**). By contrast, these cells downregulated the expression of genes involved in cytotoxicity, which is an anti-tumor function.

The SiglecF neutrophil gene signature associates with shorter patient survival

The knowledge of genes enriched in T-SiglecF^{high} neutrophils enabled us to interrogate available databases linking bulk tumor gene expression profiles with patient survival. An analysis including 568 patients across 5 databases revealed that the expression of a T-SiglecF^{high} neutrophil gene signature in bulk tumor gene expression data significantly associated with shorter survival (**Fig. 5G of Paper IV**). This observation reinforces the notion of SiglecF^{high} neutrophils having a tumor-promoting function.

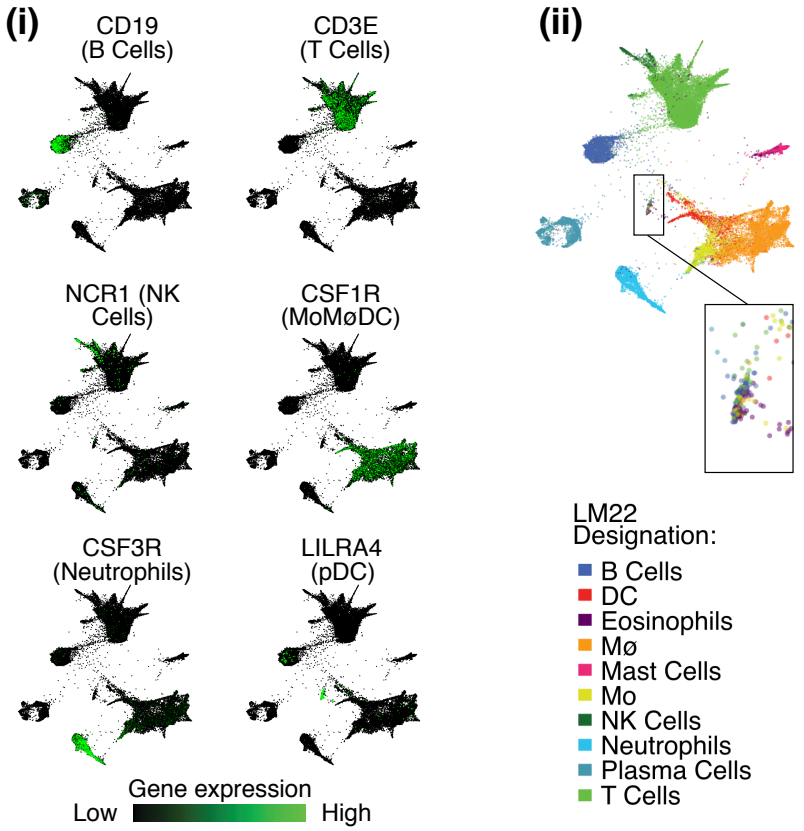

Figure F. Major lineage identification in human tumor-infiltrating immune cell data. (i) Exploration of major immune lineage marker gene expression. (ii) Results of classifying scRNAseq data by 10 expression profiles obtained from bulk RNAseq ("LM22" dataset, (100)). Each cell is given the name of the closest bulk expression profile. The inset shows the pDC cluster, which failed to properly classify because there were no pDCs in the reference dataset. Mø – macrophage.

Our findings recapitulate previous knowledge on DC and monocyte diversity, and suggest neutrophils subsets are largely novel

In the process of defining myeloid cell populations in human and mouse scRNAseq data, we naturally were relating our observations to previous knowledge to orient ourselves, and identify potentially novel populations. One simple but powerful approach was to explore the expression of known marker genes (**Fig. F(i)**). The second approach was to classify each cell by the most similar expression profiles from a reference dataset (see "Bayesian classifier" in Methods of Paper III for details). Rather than relying on individual marker genes, this approach made use of gene expression information at the whole transcriptome level. It also offered the convenience of being automated and assigning a label to every cell. In using the classifier, the most important consideration was the choice of references profiles. **Fig. F(ii)** shows classification results using a reference dataset of human immune cells, referred to as LM22, obtained by bulk RNAseq of immune populations isolated by FACS (100). A good agreement is observed with the expression of unique markers for major lineages (**Fig. F(i)**). **Fig. F(ii)** also reveals how the classifier can be misused: it always assigns one of the reference profiles, even if the reference dataset does not include the cell type of the cell being classified. As the LM22 dataset did not include pDCs, recognizable by the expression of LILRA4 (101), they were classified as a mixture of NK, B, plasma cells and eosinophils (**Fig. F(ii)** inset).

Assigning major immune cell lineages was a necessary intermediate step of the analysis. Yet, the main focus was on the minor subsets we defined, and we sought to relate them to previous knowledge. The 5 human and 6 mouse neutrophil subsets were largely novel. As covered earlier in this Chapter, in mouse, we could relate the subsets to SiglecF-high and -low neutrophils we had identified in the study described in Paper IV. By contrast, the 4 dendritic cells subsets matched well previously defined populations (see also Section 3.4.1.). Both by marker gene expression and using our classifier, 3 of the subsets we observed were identified as cDC1s, cDC2, and pDCs. The fourth subset, which we called DC3, showed an "activated" phenotype, as supported by their classifier-based similarity to DCs activate *in vitro* by LPS (102) (**Fig. 4G,O of Paper III**). The designation as "activated" is not based on ontogeny but rather reflect context-dependent functional state. We could not unambiguously identify the "resting" counterpart of DC3s in our data. The DC3 cluster could be derived from several resting populations activate at the tumor site or represent a population that migrates to the tumor site only after activation.

Similarly as with DCs, we could relate two monocyte subset in our data, Mono1 and Mono2, to well-established classical and non-classical monocytes,

respectively (see Section 3.4.1). The third monocyte subset we identified was characterized by neutrophil-associated gene expression. Although such a subset has been reported before in the blood (103), its function and ontogeny are still to be clarified.

We also observed a myeloid population with an expression profile similar to both monocytes and dendritic cells (**Fig. 5I of Paper III**), and we therefore refer to this subset as MonoDC. Monocytes are known to be able to differentiate into dendritic after migrating into tissue from the vasculature (104). It is tempting to speculate that our MonoDC subset captures this process. However, this needs further investigation.

The M1/M2 polarization paradigm (see Section 3.4.1.) places M1-like and M2-like macrophages as anti- and pro-tumorigenic (55, 105). Consistently with this view, no macrophage subset associated with an M1-like phenotype in our data.

Availability of our data at multiple levels of processing fulfilled the resource aspect of the study

A key goal of this project was to provide a resource that would help guide efforts in tumor-infiltrating myeloid cell research. Therefore, I chose to emphasize resources made available through this study:

1) Minimally processed single cell gene expression data of freshly excised whole tumor samples from 7 non-small cell lung cancer patients is accessible on the Gene Expression Omnibus (GEO), accession number GSE127465. These raw expression matrices are prior to filtering low quality cells and cell multiplets, and represent the least processed form of the data for human. The raw sequence data (fastq files) could not be made publicly available due to patient privacy concerns.
2) Online interactive tools enabling the exploration of gene expression in the unbiased 2D representation of the data obtained after dimensionality reduction. Web links are provided in the **Key Resources Table of Paper III**. For example, the human data can be accessed here[2].
3) Candidate marker genes for populations defined in this study. In particular, **Fig. 6c** and **Table S3** present a stringent selection strategy which find genes specific to a single population among all immune and non-immune set subsets in our human whole tumor data.

[2] https://kleintools.hms.harvard.edu/tools/springViewer_1_6_dev.html?datasets/Zilionis2018/human/NSCLC_immune

4) Demographic data for the 7 patients, including gender, age, smoking history, mutation information (**Table S1 of Paper III**).
5) Resources equivalent to 1)-2) for immune cells isolated from lung of healthy and lung tumor-bearing mice, including the starting sequence data (fastq files) are available on GEO (accession GSE127465).
6) Lists of gene orthologues similarly enriched in human and mouse in counterpart cell populations (**Table S3 of Paper III**).

Limitations of this resource:

1) The number of patients is insufficient to link observations in scRNAseq data with the survival of the 7 specific patients studied, even after long-term follow-up.
2) The use of a single mouse model. A possible reason for the lack of correspondence in a number of subsets between human and mouse can be due to evolutionary divergence. On the other hand, discrepancies can be an indication that different diseases are being studied. The KP lung tumor model could be capturing only certain aspects of the disease observed in the 7 NSCLC patients. As a future prospect, it would be fascinating to expand the analysis to different mouse lung tumor models, paving the way towards more personalized mechanistic studies and treatment strategies.

   3) A resource that we always had in mind during the project but chose to leave as a future project was specific FACS strategies for isolating the subsets we identified. Such a resource would open doors towards functional *in vitro* studies.

7. CONCLUSIONS

The **first goal** of facilitating access to the inDrops technology was achieved by:

i) Optimizing the method, including an enzymatic cleanup of barcoding hydrogel beads and adjusted conditions of the reverse transcription in drops.
ii) Describing the detailed procedure in a Nature Protocols publication.

The **second goal** of this thesis was to provide an unbiased view of the cellular composition of the airway epithelium under homeostasis and in recovery conditions. It was accomplished by performing the following:

i) ScRNAseq was used to reveal the unbiased population structure of airway epithelium in mouse and human. The data revealed pulmonary ionocytes as a rare novel cell type.
ii) Extending the mouse experiment to a recovery time-course following airway injury revealed two recovery-specific transient cell populations.
iii) Pulmonary ionocytes were shown to be the major source of CFTR activity in the airway epithelium. Notch signaling was necessary and FOXI was sufficient to drive ionocyte specification *in vitro*.

The **third goal** of this thesis was to identify conserved and divergent lung tumor-infiltrating myeloid cell populations in human and mouse. It was accomplished by performing the following:

i) An unbiased atlas of mouse and human lung tumor-infiltrating immune cells was generated. A novel tumor-specific and -promoting neutrophil population was identified.
ii) The comparison of myeloid cell subsets between human and mouse revealed a 1-to-1 correspondence in monocytes and DCs, conserved neutrophil subsets, and species-specific differences in macrophages.
iii) The data was made publicly available in multiple formats, including online interactive explorers.

If I were to choose a single keyword for Story I, II, and III, it would be *accessibility*, *discovery*, and *resourcefulness*, respectively.

While Paper I provides a detailed description of the inDrops technology, it should be decoupled from the development of the method, which was reported in 2015. Improvements of the method described in Section 6.1 are relatively minor changes when compared to the efforts required for early development of inDrops. Therefore, I consider my main contribution to this technology through my efforts of making it more *accessible* to others.

Story II demonstrates how scRNAseq can lead to *discovery*. ScRNAseq by itself is considered to be a hypothesis generation tool, but not definitely proof. For example, the characterization of basal cells described in Section 6.2 is descriptive and raises more questions than answers. What is the functional role of cytokeratin overexpression during injury? Do Krt4/13^{+} basal cells correspond to the basal luminal precursors from a previously proposed model (89)? Each such question requires further experiments and possibly deserves a separate study. In the case of pulmonary ionocytes, after observing them in scRNAseq data as a potentially novel cell type and the major source of *CFTR* transcripts, we only convinced ourselves of their existence and relevance to cystic fibrosis after performing a number of specifically designed follow-up experiments.

Story III delivers a *resource* and therefore is just the beginning of a story. We created a reference map of tumor-infiltrating myeloid cell populations for further mechanistic studies by ourselves and others, which would in most cases be performed in mice. By identifying cell populations conserved in human and mouse we reduced the list of possible research directions to human-relevant ones.

8. METHODS

The central method of this thesis is inDrops single-cell RNA sequencing, a multi-step procedure requiring microfluidic fabrication and device operation, gentle handling of cell suspensions, and application of molecular biology techniques. The latter was also necessary in optimizing the technology, as described in Chapter 6.1.

While the application of inDrops was a prerequisite to obtain the data on gene expression in thousands of cells used in Papers II-IV, the main effort consisted of analyzing and turning data into insight. This is in line with common wisdom in the scRNAseq field saying that it takes days to gather the data and years to analyze it. Data analysis can be further divided in low-and high-level analyses. Low-level processing involves a well-defined bioinformatic task of turning the raw sequence data (fastq files) into a gene expression matrix (cells x genes). By contrast, high-level analysis encompasses a wide diversity of techniques, including quality filters, data dimensionality reduction and visualization, the definition of cell populations and their characteristic genes. High-level analysis often needed to be tailored around the specific questions asked. Care was taken to explain analytical procedures used in Methods sections of Paper II-IV. In addition, example python code for selected analyses of Papers II-III is available at github/rapolaszilionis/Sharir_et_al_2019[3].

InDrops single cell RNA sequencing

Paper I provides a detailed description of the procedure as applied during the thesis. All step including custom reagent synthesis (i.e. microfluidic chips and barcoding hydrogel beads) were performed in Klein lab at HMS. Microfluidic chip fabrication was performed at the HMS Microfluidic facility. Liquid handling was performed at the ICCB-Longwood Screening Facility.

Data analysis

Gene expression count matrices were obtained from raw sequence data (fastq files or bcl files) using the indrops.py python pipeline available at github.com/adrianveres/indrops. The pipeline sorts sequencing reads by cell barcode information, aligns them to a reference transcriptome, and quantifies transcripts detected per gene. Computation resources needed to process the data were obtained by running the pipeline on a Linux-based high-performance computer cluster at Harvard Medical School.

[3] One of the resources of Paper XII

High-level data analysis was performed in python mainly using standard libraries (numpy, pandas, matplotlib, scipy). While noteworthy packages with wrappers around published scRNAseq methods exists (e.g. Seurat in R and, more recently, Scanpy in python), most of the code for analyses shown in Papers II-IV was explicitly written using the standard libraries. Exceptions are SPRING, used for visualization and interactive exploration of scRNAseq data (86), and Scrublet, used for identification of cell multiplets, which occur when two or more cells share the same barcode (106).

9. ACKNOWLEDGEMENTS

First and foremost, thank you to my supervisors, who played a main role in making my PhD such a wonderful and enriching experience.

Linas, I am immensely grateful for your trust, support, and the possibilities that you opened to me. You are a role model to me for your choice to return to Lithuania and the scientifically motivating and collegial environment you created in your lab. For me, you are an example of how to work towards big ambitions.

Allon, you completely spoiled me. I cannot imagine a better place to do science than under your supervision. Being in your lab was completely transformative to me, both professionally and personally. Boosting the self-esteem of your trainees and helping them grow seems to be the top priority to you. You are always here to support us. Including late nights and weekends. At least this is the way I feel. Your generosity and integrity are unmatched.

Members of Klein lab, I cannot thank you enough for your friendship. You created a work environment like no other. Thank you for the sad faces as I was leaving. Each of you is an example to me that I want to follow. Sam, I felt that you understood me particularly well. Thank you for the competition of who is coming to lab first, for answering my endless questions on data analysis, and the feeling I could always learn on you and ask for your help. Hailey, James, Naren, Kalki, Katie, Dan, Kyogo, Virginia, Caleb, Alex, Mandovi, thank you for your high spirits, everything I learned from you, your advice, and lots of fun together. I miss you all a lot!

Virginia, thank you for embarking me on the tumor immunology project you initiated in the first place. Thank you for all the deep scientific discussions we had, and everything I learned from you. Data analysis is today the main thing I do, but I still remember you teaching me what “cd” and “ls” did in Bash.

Members of Mažutis lab, thank you for exciting scientific discussion and your advice. A special thanks to Juozas Nainys, Greta Stonytė and Vaidotas Kiseliovas for the work and fun together both in Lithuania and the US. Thanks to Karolis G. for discussions and experimental insights into inDrops improvements. And all others for being a wonderful team.

I feel privileged to have had a chance to work together with the groups of Aron Jaffe and Mikael Pittet on Stories II and III, respectively. Not only I could expand my knowledge in airway biology and tumor immunology while working with you. Working together was as truly delightful experience which I am

extremely grateful for. Aron and Mikael, thank you for your trust, encouragement, and all the professional advice you gave me.

Lindsey, thank you for the all the fun we had working on the ionocyte discovery, from discussing figure panels to the photo shoot for the popular scientific article advertising our discovery.

Camilla, Christina, and David, thank you the wonderful journey that working with you has been and still is. Thank you for walking me through the world of immunology. Thank you for your trust and friendship.

Camilla, thank you so much for the intense, productive, and enriching sessions we had working side-by-side to push the Immunity paper to completion. Thank you for all the scientific, career, and life discussions we had. I learned a lot from you.

Thank you to the groups of Elena Levantini and Ophir Klein for fruitful and enriching collaborations.

Thank you to Adrian Veres and Eran Hodis, who taught me RNAseq data analysis in their course at Harvard and invited me to join them as a teaching assistant the follow year. Your course was transformative to me as it marked my transition toward computational biology.

I am very grateful to the member of the defense panel, Gintaras Valinčius, Aurelija Žvirblienė, Skirmantas Kriaučionis, Daiva Dabkevičienė, and Algirdas Žiogas, for accepting to evaluate my doctoral thesis. A special thank you to Aurelija Žvirblienė and Skirmantas Kriaučionis for not only reviewing my thesis but also doing it in a short time.

Thank you to all the people who enabled me to focus on science by helping me deal with various duties, deadlines, paperwork, and administrative questions. In particular, at the Life Sciences Center, I would like to thank Rokas Abraitis, Asta Abraitienė, Janina Žiūkaitė, Danutė Noreikienė, Vytautas Smirnovas. Thank you to Daiva Baltriukienė, Asta Abraitienė and Virginija Uksienė for guidance through the process of writing the thesis and preparing it for defense. At the SysBio department at HMS, thank you to Maria Ferreira for guidance through the visa and related questions. Jennie Epp, Samantha Keough, Sybille Kelleher for answer to all my questions related to lab management.

I would also like to thank the broader communities at the Life Science Center of VU and the SysBio department at HMS for a wonderful professional environment.

An important part of my PhD experience was taking results from the lab and presenting them to the general public. Thank you to Gintaras Valinčius for the encouragement, and Agnė Grinevičiūtė and Gediminas Dubonikas for spreading the news.

My beloved family, my deepest gratitude goes to you all. Mom, the self-esteem that you nurtured for all my life reminds me that everything is possible.

My dear wife Vilija, words are simply not enough to express my gratitude to you. Your unconditional love, care, courage, support, dedication and encouragement were indispensable during my doctorate.

10. LIST OF ORAL PRESENTATIONS

Oral presentations at international conferences:

1) 2018 01 03 Vita Scientia, Vilnius University, “New cell type discovery by single cell RNA sequencing”

2) 2018 10 05 Baltic Biophysics Conference, Vytautas Didysis University “Revisiting the composition of airway epithelium by single cell RNA sequencing”

3) 2018 11 27 4Bio Summit Europe, Rotterdam, “Droplet microfluidics technology for single-cell biology research”

4) 2019 02 26 theCOINS, Vilnius University, keynote speaker, “Single cell RNAseq as a tool for cell type discovery”

11. CURRICULUM VITAE

12. REFERENCES

1. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187-201.
2. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc. 2018;13(4):599-604.
3. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell. 2015;58(4):610-20.
4. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377-82.
5. Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011;21(7):1160-7.
6. Grun D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nat Methods. 2014;11(6):637-40.
7. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014;343(6172):776-9.
8. Herman JS, Sagar, Grün D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat Methods. 2018;15(5):379-86.
9. Bose S, Wan Z, Carr A, Rizvi AH, Vieira G, Pe'er D, et al. Scalable microfluidics for single-cell RNA printing and sequencing. Genome Biol. 2015;16:120.
10. Fan HC, Fu GK, Fodor SPA. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. Science. 2015;347(6222):1258367.
11. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.
12. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, et al. Single-cell barcoding and sequencing using droplet microfluidics. Nat Protoc. 2017;12(1):44-73.
13. Bardin D, Kendall MR, Dayton PA, Lee AP. Parallel generation of uniform fine droplets at hundreds of kilohertz in a flow-focusing module. Biomicrofluidics. 2013;7(3):34112.
14. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016;17:77.
15. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012;2(3):666-73.
16. Bustamante-Marin XM, Ostrowski LE. Cilia and Mucociliary Clearance. Cold Spring Harb Perspect Biol. 2017;9(4).
17. Knowles MR, Zariwala M, Leigh M. Primary Ciliary Dyskinesia. Clin Chest Med. 2016;37(3):449-61.
18. Ramos FL, Krahnke JS, Kim V. Clinical issues of mucus accumulation in COPD. Int J Chron Obstruct Pulmon Dis. 2014;9:139-50.
19. Who. World health statistics 2008: World Health Organization; 2008.

20. Jeffery PK. Remodeling and inflammation of bronchi in asthma and chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2004;1(3):176-83.
21. Hogan BLM, Barkauskas CE, Chapman HA, Epstein JA, Jain R, Hsia CCW, et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. Cell Stem Cell. 2014;15(2):123-38.
22. Cutting GR. Cystic fibrosis genetics: from molecular understanding to clinical application. Nat Rev Genet. 2015;16(1):45-56.
23. Andersen DH. Cystic fibrosis of the pancreas and its relation to celiac disease: a clinical and pathologic study. American journal of Diseases of Children. 1938;56(2):344-99.
24. Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science. 1989;245(4922):1059-65.
25. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989;245(4922):1066-73.
26. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989;245(4922):1073-80.
27. Ratjen FA. Cystic fibrosis: pathogenesis and future treatment strategies. Respir Care. 2009;54(5):595-605.
28. Boucher RC. Evidence for airway surface dehydration as the initiating event in CF airway disease. J Intern Med. 2007;261(1):5-16.
29. Engelhardt JF, Yankaskas JR, Ernst SA, Yang Y, Marino CR, Boucher RC, et al. Submucosal glands are the predominant site of CFTR expression in the human bronchus. Nat Genet. 1992;2(3):240-8.
30. Kreda SM, Mall M, Mengos A, Rochelle L, Yankaskas J, Riordan JR, et al. Characterization of wild-type and deltaF508 cystic fibrosis transmembrane regulator in human respiratory epithelia. Mol Biol Cell. 2005;16(5):2154-67.
31. McCauley KB, Alysandratos K-D, Jacob A, Hawkins F, Caballero IS, Vedaie M, et al. Single-Cell Transcriptomic Profiling of Pluripotent Stem Cell-Derived SCGB3A2+ Airway Epithelium. Stem Cell Reports. 2018;10(5):1579-95.
32. Rock JR, Hogan BLM. Epithelial progenitor cells in lung development, maintenance, repair, and disease. Annu Rev Cell Dev Biol. 2011;27:493-512.
33. Whitsett JA. Airway Epithelial Differentiation and Mucociliary Clearance. Ann Am Thorac Soc. 2018;15(Supplement_3):S143-S8.
34. Hutton E, Paladini RD, Yu QC, Yen M, Coulombe PA, Fuchs E. Functional differences between keratins of stratified and simple epithelia. J Cell Biol. 1998;143(2):487-99.
35. Cole BB, Smith RW, Jenkins KM, Graham BB, Reynolds PR, Reynolds SD. Tracheal Basal cells: a facultative progenitor cell pool. Am J Pathol. 2010;177(1):362-76.
36. Ghosh M, Brechbuhl HM, Smith RW, Li B, Hicks DA, Titchner T, et al. Context-dependent differentiation of multipotential keratin 14-expressing tracheal basal cells. Am J Respir Cell Mol Biol. 2011;45(2):403-10.
37. Jeffery PK, Li D. Airway mucosa: secretory cells, mucus and mucin genes. Eur Respir J. 1997;10(7):1655-62.

38. Reynolds SD, Reynolds PR, Pryhuber GS, Finder JD, Stripp BR. Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways. American journal of respiratory and critical care medicine. 2002;166(11):1498-509.
39. Spassky N, Meunier A. The development and functions of multiciliated epithelia. Nat Rev Mol Cell Biol. 2017;18(7):423-36.
40. Rawlins EL, Okubo T, Xue Y, Brass DM, Auten RL, Hasegawa H, et al. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. Cell Stem Cell. 2009;4(6):525-34.
41. Branchfield K, Nantie L, Verheyden JM, Sui P, Wienhold MD, Sun X. Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. Science. 2016;351(6274):707-10.
42. Reid L, Meyrick B, Antony VB, Chang L-Y, Crapo JD, Reynolds HY. The mysterious pulmonary brush cell: a cell in search of a function. American journal of respiratory and critical care medicine. 2005;172(1):136-9.
43. Saunders CJ, Reynolds SD, Finger TE. Chemosensory brush cells of the trachea. A stable population in a dynamic epithelium. Am J Respir Cell Mol Biol. 2013;49(2):190-6.
44. Krasteva G, Canning BJ, Hartmann P, Veres TZ, Papadakis T, Mühlfeld C, et al. Cholinergic chemosensory cells in the trachea regulate breathing. Proc Natl Acad Sci U S A. 2011;108(23):9478-83.
45. Engblom C, Pfirschke C, Pittet MJ. The role of myeloid cells in cancer therapies. Nat Rev Cancer. 2016;16(7):447-62.
46. Kondo M. Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. Immunol Rev. 2010;238(1):37-46.
47. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. Cell. 2008;132(4):631-44.
48. Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science. 2016;351(6269):aab2116.
49. Tusi BK, Wolock SL, Weinreb C, Hwang Y, Hidalgo D, Zilionis R, et al. Population snapshots predict early haematopoietic and erythroid hierarchies. Nature. 2018;555(7694):54-60.
50. Monie TP. Section 2 - Immune Cells and the Process of Pattern Recognition. In: Monie TP, editor. The Innate Immune System: Academic Press; 2017. p. 41-82.
51. van Furth R, Cohn ZA. The origin and kinetics of mononuclear phagocytes. J Exp Med. 1968;128(3):415-35.
52. Davies LC, Jenkins SJ, Allen JE, Taylor PR. Tissue-resident macrophages. Nat Immunol. 2013;14(10):986-95.
53. Murray PJ, Allen JE, Biswas SK, Fisher EA, Gilroy DW, Goerdt S, et al. Macrophage activation and polarization: nomenclature and experimental guidelines. Immunity. 2014;41(1):14-20.
54. Ginhoux F, Schultze JL, Murray PJ, Ochando J, Biswas SK. New insights into the multidimensional concept of macrophage ontogeny, activation and function. Nat Immunol. 2016;17(1):34-40.
55. Biswas SK, Mantovani A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. Nat Immunol. 2010;11(10):889-96.

56. Carter Philip B. George Bellamy Mackaness. 20 August 1922 — 4 March 2007. Biogr Mem Fellows R Soc. 2014;60:277-98.
57. Van Epps HL. Macrophage activation unveiled. Journal of Experimental Medicine. 2005;202(7):884-.
58. Stein M, Keshav S, Harris N, Gordon S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. J Exp Med. 1992;176(1):287-92.
59. Mills CD, Kincaid K, Alt JM, Heilman MJ, Hill AM. M-1/M-2 macrophages and the Th1/Th2 paradigm. J Immunol. 2000;164(12):6166-73.
60. Berger A. Th1 and Th2 responses: what are they? BMJ. 2000;321(7258):424.
61. Allen JE, Sutherland TE. Host protective roles of type 2 immunity: parasite killing and tissue repair, flip sides of the same coin. Semin Immunol. 2014;26(4):329-40.
62. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. F1000Prime Rep. 2014;6:13.
63. Davis MJ, Tsang TM, Qiu Y, Dayrit JK, Freij JB, Huffnagle GB, et al. Macrophage M1/M2 polarization dynamically adapts to changes in cytokine microenvironments in Cryptococcus neoformans infection. MBio. 2013;4(3):e00264-13.
64. Quintar A, McArdle S, Wolf D, Marki A, Ehinger E, Vassallo M, et al. Endothelial protective monocyte patrolling in large arteries intensified by western diet and atherosclerosis. Circulation research. 2017;120(11):1789-99.
65. Merad M, Sathe P, Helft J, Miller J, Mortha A. The dendritic cell lineage: ontogeny and function of dendritic cells and their subsets in the steady state and the inflamed setting. Annu Rev Immunol. 2013;31:563-604.
66. Reizis B. Plasmacytoid dendritic cells: development, regulation, and function. Immunity. 2019;50(1):37-50.
67. Joffre OP, Segura E, Savina A, Amigorena S. Cross-presentation by dendritic cells. Nat Rev Immunol. 2012;12(8):557-69.
68. Dutertre C-A, Wang L-F, Ginhoux F. Aligning bona fide dendritic cell populations across species. Cell Immunol. 2014;291(1-2):3-10.
69. Lämmermann T, Afonso PV, Angermann BR, Wang JM, Kastenmüller W, Parent CA, et al. Neutrophil swarms require LTB4 and integrins at sites of cell death in vivo. Nature. 2013;498(7454):371-5.
70. Peiseler M, Kubes P. More friend than foe: the emerging role of neutrophils in tissue repair. J Clin Invest. 2019;130:2629-39.
71. Silva MT, Correia-Neves M. Neutrophils and macrophages: the main partners of phagocyte cell systems. Front Immunol. 2012;3:174.
72. Stone KD, Prussin C, Metcalfe DD. IgE, mast cells, basophils, and eosinophils. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S73-80.
73. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science. 2011;331(6024):1565-70.
74. Darvin P, Toor SM, Sasidharan Nair V, Elkord E. Immune checkpoint inhibitors: recent progress and potential biomarkers. Exp Mol Med. 2018;50(12):165.
75. Böttcher JP, Reis E Sousa C. The Role of Type 1 Conventional Dendritic Cells in Cancer Immunity. Trends Cancer Res. 2018;4(11):784-92.
76. Dvorak HF. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N Engl J Med. 1986;315(26):1650-9.

77. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21(8):938-45.
78. DuPage M, Dooley AL, Jacks T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat Protoc. 2009;4(7):1064-72.
79. Kersten K, de Visser KE, van Miltenburg MH, Jonkers J. Genetically engineered mouse models in oncology research and cancer medicine. EMBO Mol Med. 2017;9(2):137-53.
80. DuPage M, Jacks T. Genetically engineered mouse models of cancer reveal new insights about the antitumor immune response. Curr Opin Immunol. 2013;25(2):192-9.
81. Politi K, Pao W. How genetically engineered mouse tumor models provide insights into human cancers. Journal of Clinical Oncology. 2011;29(16):2273.
82. Pfirschke C, Engblom C, Rickelt S, Cortez-Retamozo V, Garris C, Pucci F, et al. Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy. Immunity. 2016;44(2):343-54.
83. Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science. 2018;360(6392).
84. Johnson L, Abelson H, Green H, Penman S. Changes in RNA in relation to growth of the fibroblast. I. Amounts of mRNA, rRNA, and tRNA in resting and growing cells. Cell. 1974;1(2):95-100.
85. Paul MK, Bisht B, Darmawan DO, Chiou R, Ha VL, Wallace WD, et al. Dynamic changes in intracellular ROS levels regulate airway basal stem cell homeostasis through Nrf2-dependent Notch signaling. Cell Stem Cell. 2014;15(2):199-214.
86. Weinreb C, Wolock S, Klein AM. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics. 2018;34(7):1246-8.
87. Choksi SP, Lauter G, Swoboda P, Roy S. Switching on cilia: transcriptional networks regulating ciliogenesis. Development. 2014;141(7):1427-41.
88. Quigley IK, Stubbs JL, Kintner C. Specification of ion transport cells in the Xenopus larval skin. Development. 2011;138(4):705-14.
89. Watson JK, Rulands S, Wilkinson AC, Wuidart A, Ousset M, Van Keymeulen A, et al. Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium. Cell Rep. 2015;12(1):90-101.
90. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci U S A. 2018;115(10):E2467-E76.
91. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nature biotechnology. 2019;37(5):547.
92. Vidarsson H, Westergren R, Heglind M, Blomqvist SR, Breton S, Enerbäck S. The forkhead transcription factor Foxi1 is a master regulator of vacuolar H-ATPase proton pump subunits in the inner ear, kidney and epididymis. PLoS One. 2009;4(2):e4471.
93. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket SE, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature. 2018;560(7718):319-24.
94. Sharma P, Allison JP. The future of immune checkpoint therapy. Science. 2015;348(6230):56-61.

95. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541-50.
96. Elliott LA, Doherty GA, Sheahan K, Ryan EJ. Human tumor-infiltrating myeloid cells: phenotypic and functional diversity. Frontiers in immunology. 2017;8:86.
97. Pillay J, Tak T, Kamp VM, Koenderman L. Immune suppression by neutrophils and granulocytic myeloid-derived suppressor cells: similarities and differences. Cellular and Molecular Life Sciences. 2013;70(20):3813-27.
98. Didkowska J, Wojciechowska U, Mańczuk M, Łobaszewski J. Lung cancer epidemiology: contemporary and future challenges worldwide. Annals of translational medicine. 2016;4(8).
99. Macauley MS, Crocker PR, Paulson JC. Siglec-mediated regulation of immune cell function in disease. Nat Rev Immunol. 2014;14(10):653-66.
100. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-7.
101. Cao W, Rosen DB, Ito T, Bover L, Bao M, Watanabe G, et al. Plasmacytoid dendritic cell–specific receptor ILT7–FcεRIγ inhibits Toll-like receptor–induced interferon production. Journal of Experimental Medicine. 2006;203(6):1399-405.
102. Abbas A, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. Genes and immunity. 2005;6(4):319.
103. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science. 2017;356(6335).
104. León B, López-Bravo M, Ardavín C. Monocyte-derived dendritic cells. Semin Immunol. 2005;17(4):313-8.
105. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. Nat Rev Clin Oncol. 2017;14(7):399-416.
106. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. bioRxiv. 2018.

13. SANTRAUKA

REZIUMĖ

2015 metai žymi didelio našumo pavienių ląstelių RNR sekoskaitos (angl. *single cell RNA sequencing*, scRNAseq) technologijų atsiradimą. Šios technologijos leidžia vieno eksperimento metu išmatuoti visų genų raiškos lygį kiekvienoje iš dešimčių tūkstančių ląstelių. Vienas šių metodų – „inDrops“ – kūrybiškai išnaudojo mikroskysčių technologijų suteikiamą didelį našumą ir netrukus susilaukė susidomėjimo tiek iš metodo kūrimą liudijusios akademinės aplinkos, tiek ir iš platesnės mokslinės bendruomenės. Vis dėlto, technologijos platesnį taikomumą apsunkino šio daugiažingsnio ir nestandartinių reagentų reikalaujančio metodo sudėtingumas. Šioje daktaro disertacijoje aprašomos pastangos palengvinti „inDrops“ technologijos prieinamumą. Taip pat pristatomas technologijos pritaikymas dviejuose kontekstuose: tiriant kvėpavimo takų epitelio ir plaučio navikus infiltruojančių imuninių ląstelių sudėtį.

Abiejuose tyrimuose scRNAseq buvo naudojama sudarant nešališkus, nuo eksperimentatoriaus pasirinkimų nepriklausomus ląstelių tipų katalogus. Kvėpavimo takų epitelį tyrėme tikėdamiesi geriau įvertinti bazinių ląstelių, kurios yra šio audinio kamieninės ląstelės, heterogeniškumą. Taip siekėme suprasti, kaip homeostatinė audinio sudėtis atstatoma po pažaidų. Nors tyrimo metu laikėmės savo pradinių tikslų, pagrindinę tyrimo kryptį pakeitė netikėtas atradimas: iki šiol neaprašytas ląstelių tipas, kurį pavadinome plaučių jonocitais. Papildomais eksperimentais parodėme plaučių jonocitų lokaciją audinyje, geriau supratome, kas lemia šio ląstelių tipo specifikaciją, ir nustatėme jų funkcinį ryšį su cistine fibroze. Nors jonocitų funkcija kvėpavimo takuose dar laukia papildomų tyrimų, šių ląstelių identifikavimas turėtų tiesiogiai pasitarnauti cistinės fibrozės tyrimams.

Antrame tyrime siekėme tiksliau aprašyti vėžį infiltruojančių imuninių ląstelių įvairovę, bei palyginti ląstelių tipus tarp žmogaus ir pelės. Pasirūpinome, kad gausūs scRNAseq duomenys būtų lengvai prieinami ir taptų atskaitos tašku tolimesniems tyrimams *in vitro* ir *in vivo*. Duomenų analizė atskleidė tikslų žmogaus-pelės atitikimą tarp dendritinių ląstelių ir monocitų potipių, dalinį neutrofilų įvairovės atitikimą ir ženklius tarprūšinius skirtumus tarp makrofagų. Mūsų žiniai, aptikti neutrofilų potipiai iki mūsų tyrimo nebuvo gerai charakterizuoti. Dalis neutrofilų buvo aptikti tik vėžiu sergančių, bet ne sveikų pelių plaučiuose. Šių ląstelių vėžį skatinantis poveikis buvo įrodytas atskirame projekte, prie kurio turėjau galimybę prisidėti. Tikimės, kad mūsų atliktas žmogaus ir pelės imuninių ląstelių palyginimas padės geriau įvertinti skirtumus tarp rūšių ir nukreipti pelėje atliekamus tyrimus žmogui aktualiomis kryptimis.

STRAIPSNIAI SUDARANTYS DISERTACIJOS PAGRINDĄ

I. **Zilionis R**, Nainys J, Veres A, Savova V, Zemmour D, Klein AM[#], Mazutis L[#]. Single-cell barcoding and sequencing using droplet microfluidics. *Nature Protocols*. 2017 Jan;12(1):44–73.

Atlikau metodo optimizavimo eksperimentus, paruošiau duomenis ir parašiau protokolą.

II. Plasschaert LW[*], **Zilionis** R[*], Choo-Wing R, Savova V, Knehr J, Roma G, Klein AM[#], Jaffe AB[#]. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*. 2018 Aug;560(7718):377–81.

Atlikau scRNAseq eksperimentus, išanalizavau gautus duomenis ir paruošiau galutinius grafikus, pavaizduotus straipsnyje. Dalyvavau straipsnio rašyme ir šiuo metu kuruoju duomenis, kuriuos pateikėme į viešas duomenų bazes.

III. **Zilionis** R[*], Engblom C[*], Pfirschke C*, Savova V*[#], David Zemmour, Saatcioglu HD, Krishnan I, Maroni G, Meyerovitz CV, Kerwin CM, Choi S, Richards WG, De Rienzo A, Tenen DG, Bueno R, Levantini E, Pittet MJ[#], Klein AM[#]. Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity*. 2019 May 21;50(5):1317– 34.e10.

Mano indėlis toks pats, kaip ir II-ame straipsnyje.

IV. Engblom C[*], Pfirschke C[*], **Zilionis R**, Da Silva Martins J, Bos SA, Courties G, Rickelt S, Severe N, Baryawno N, Faget J, Savova V, Zemmour D, Kline J, Siwicki M, Garris C, Pucci F, Liao HW, Lin YJ, Newton A, Yaghi OK, Iwamoto Y, Tricot B, Wojtkiewicz GR, Nahrendorf M, Cortez-Retamozo V, Meylan E, Hynes RO, Demay M, Klein K, Bredella MA, Scadden DT, Weissleder R, Pittet MJ[#]. Osteoblasts remotely supply lung tumors with cancer-promoting SiglecFhigh neutrophils. *Science*. 2017 Dec 1;358(6367).

Atlikau neutrofilų scRNAseq duomenų analizę ir pacientų išgyvenamumo analizę pateiktas pav. 5 ir S19, aprašiau atitinkamą metodinę dalį, redagavau rezultatų skyrių.

[*]Šių autorių indėlis laikomas vienodu.

[#]Autoriai korespondencijai.

KITI STRAIPSNIAI

Doktorantūros metu paskelbti straipsniai, nesudarantys disertacijos pagrindo:

V. Galinis R[*], Stonyte G[*], Kiseliovas V, **Zilionis R**, Studer S, Hilvert D, Janulaitis A, Mazutis L[#]. DNA Nanoparticles for Improved Protein Synthesis In Vitro. *Angewandtle Chemie*. 2016 Feb 24;55(9):3120–3.

VI. Derr A, Yang C[*], **Zilionis R**[*], Sergushichev A, Blodgett DM, Redick S, Bortell R, Luban J, Harlan DM, Kadener S, Greiner DL, Klein AM, Artyomov MN, Garber M[#]. End Sequence Analysis Toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. *Genome Research*. 2016 Oct;26(10):1397– 410.

VII. Hrvatin S[*], Hochbaum DR[*], Nagy MA[*], Cicconet M, Robertson K, Cheadle L, **Zilionis R**, Ratner A, Borges-Monroy R, Klein AM, Sabatini BL[#], Greenberg ME[#]. Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nature Neuroscience*. 2018 Jan;21(1):120–9.

VIII. Tusi BK[*], Wolock SL[*], Weinreb C, Hwang Y, Hidalgo D, **Zilionis R**, Waisman A, Huh JR, Klein AM[#], Socolovsky M[#]. Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature*. 2018 Mar 1;555(7694):54–60.

IX. Zemmour D, **Zilionis R**, Kiner E, Klein AM, Mathis D[#], Benoist C[#]. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nature Immunology.* 2018 Mar;19(3):291–301.

X. DiSpirito JR[*], Zemmour D[*], Ramanan D, Cho J, **Zilionis R**, Klein AM, Benoist C[#], Mathis D[#]. Molecular diversification of regulatory T cells in nonlymphoid tissues. *Science Immunology*. 2018 Sep 14;3(27).

XI. Jones KB, Furukawa S[*], Marangoni P[*], Ma H, Pinkard H, D'Urso R, **Zilionis R**, Klein AM, Klein OD. Quantitative Clonal Analysis and Single-Cell Transcriptomics Reveal Division Kinetics, Hierarchy, and Fate of Oral Epithelial Progenitor Cells. *Cell Stem Cell.* 2019 Jan 3;24(1):183–92.e8.

XII. Sharir A, Marangoni P[*], **Zilionis R**[*], Wan M, Wald T, Hu J, Kawaguchi K, Castillo-Azofeifa D, Epstein Leo, Harrington K, Pagella P, Mitsiadis T, Siebel CW, Klein AM[#] and Klein OD[#]. A large pool of actively cycling progenitors orchestrates self-renewal and injury repair of an ectodermal appendage. Accepted to Nature Cell Biology.

[*]Šių autorių indėlis laikomas vienodu.

[#]Autoriai korespondencijai.

V-ame straipsnyje atlikau pradinius eksperimentus rodančius, kad pirofosfato (PPi) pridėjimas į DNR tirpalą turintį Mg^{2+} jonų užtektinas DNA:PPi:Mg nuosėdoms susidaryti. VI-X straipsniuose atlikau arba visus (VI, IX ir X), arba dalį ląstelių transkriptomų indeksavimo eksperimentų. XI ir XII straipsniuose atlikau visą bioinformatinę scRNAseq duomenų analizę.

Šį darbą sudarantys 4 straipsniai suskirstyti į 3 temas, kurių kiekviena sprendžia atskiras problemas.

I-a tema: „inDrops“ technologijos prieinamumas ir optimizavimas. „inDrops“ – didelio našumo pavienių ląstelių RNR sekoskaitos metodas, leidžiantis išmatuoti visų genų raišką tūkstančiuose individualių ląstelių vieno eksperimento metu. Pagrindinis metodo taikymas: ląstelių populiacijų identifikavimas *de novo*. Kadangi matuojama visų, o ne eksperimentatoriaus pasirinktų genų raiška, duomenys atskleidžia pilną ląstelių transkripcinių būsenų įvairovę ir leidžia sudaryti išsamius ląstelių tipų katalogus. Po to, kai „inDrops“ metodas buvo publikuotas 2015 m. (1), jis iš karto susilaukė susidomėjimo iš norinčių jį taikyti laboratorijų. Platesnį taikomumą sunkino detalaus protokolo stoka. Taip pat buvo pastebėtas variabilumas tarp „inDrops“ eksperimentų, kurį reikėjo identifikuoti ir pašalinti. Galiausiai, nuo originalaus metodo publikavimo jau buvo sukaupta patobulinimų daugelyje metodo žingsnių, kuriais norėjome pasidalinti su platesne mokslininkų bendruomene. Savo doktorantūros metu atlikau „inDrops“ metodo optimizavimo eksperimentus, išsiaiškinau ir ištaisiau eksperimentų techninio variabilumo priežastį, bei aprašiau metodą griežtai struktūrizuotoje „Nature Protocols“ publikacijoje. Manau, kad šie darbai prisidėjo prie platesnio metodo taikymo bei sukūrė pagrindą tolimesniems jo tobulinimams.

II-a tema: Plaučių jonocitų atradimas ir charakterizavimas. Pradedant šį projektą, viršutinių kvėpavimo takų epitelis jau buvo gerai ištirtas. Buvo žinoma, kad jį sudaro bazinės ląstelės, kurios atlieka kamieninių ląstelių funkciją ir diferencijuoja į sekretorines (*secretory*), blakstienuotasias (*ciliated*), neuroendokrinines (*neuroendocrine*) ir šepetines (*brush*) ląsteles su specializuotomis funkcijomis (2). Kartu šios ląstelės atsakingos už svarbiausią kvėpavimo takų apsauginę funkciją – mukociliarinį transportą (*mucociliary,* MCC). Tai procesas, kurio metu blakstienuotųjų ląstelių blakstienėlės išstumia iš kvėpavimo takų gleives, kurios sugauna įkvepiamus patogenus ir daleles. Žinoma, kad MCC procesas ir epitelio ląstelinės sudėties balansas yra sutrikę daugelyje kvėpimo takų patologijų, tarp jų astmos ir cistinės fibrozės atvejais (3, 4). Pelių modeliai leidžia tirti epitelio atsistatymą po pažaidų. Supratus šį procesą galima būtų kurti terapines strategijas gydant ligas. Vis dėlto, nepaisant pažangos tiriant šį audinį, su kvėpavimo takais susijusios patologijos yra tarp pagrindinių mirčių priežasčių pasaulyje (5). Siekiant progreso ligų gydyme, svarbu geriau suprasti kvėpavimo takų epitelio sudėtį. ScRNAseq atsiradimas suteikė tam galimybę. Savo atliktame projekte ištyrėme kvėpavimo takų epitelio sudėtį homeostatinėmis sąlygomis ir išsiaiškinome, kaip sudėtis kinta gijimo po pažaidų metu. Be žinomų ląstelių tipų identifikavimo, įvertinome bazinių ląstelių heterogeniškumą ir

atskleidėme dvi tik gijimo metu atsirandančias ląstelių populiacijas. Svarbiausias atradimas – naujas, anksčiau necharakterizuotas retas ląstelių tipas, kurį pavadinome plaučių jonocitais. Kadangi taikant scRNAseq prarandama erdvinė informacija, atskirais eksperimentais parodėme jonocitų lokaciją audinyje. Naujojo ląstelių tipo pavadinimą pasirinkome pagal genų raiškos panašumą į jonocitus, aprašytus kituose biologiniuose kontekstuose, kaip *Xenopus* varlės buožgalvių oda ir pelių inkstai. Šiose biologinėse sistemose buvo nustatyta Notch signalinio kelio ir FOXI transkripcijos veiksnio svarba jonocitų specifikacijai (6, 7). Papildomais eksperimentais įrodėme, kad ir kvėpavimo takuose šie veiksniai svarbūs jonocitų diferenciacijai iš bazinių ląstelių.

Jonocitų atradimas svarbus ir kitame, cistinės fibrozės tyrimų kontekste. Cistinė fibrozė –reta genetinė recesyvinė autosominė liga, paveikianti daugelį organų ir ypač pasireiškianti kvėpavimo takuose. Ligai būdingas gleivių sutirštėjimas, dėl kurio jos sunkiai pašalinamos iš kvėpavimo takų. Be kvėpavimo takų kimšimosi, tai sukelia sąlygas komplikacijoms, tokioms kaip bakterinės infekcijos, kurios sveikiems individams nėra pavojingos (8, 9). Nuo tada, kai liga buvo pripažinta atskira patologija 1938-ais (10), buvo sukauptos gausios žinios apie jos pasireiškimą organų lygmenyje. 1989-ais trys tyrėjų grupės identifikavo genetinę ligos priežastį: daugeliu atveju, ją sukelia mutavęs genas, kurį mokslininkai pavadino CFTR (ang. *Cystic Fibrosis Transmembrane Conductance Regulator*) (11-13). Šis genas koduoja membranos jonų kanalą, kuris ligos atveju yra pažeistas ir neatlieka savo funkcijos. Sutrikus jonų judėjimui iš ląstelės, kartu pasikeičia osmosinis slėgis ir į kvėpavimo takus nebepatenka vanduo ir dėl to sutirštėja gleivės.

Visgi, ląsteliniame lygmenyje nebuvo žinoma, kuriose ląstelėse lokalizuotas CFTR. Savo duomenyse pastebėjome, kad šio geno transkriptų daugiausiai jonocituose. Papildomais eksperimentais įrodėme, kad ne tik transkriptomo lygmenyje, bet ir funkciškai jonocitai atsakingi už CFTR veiklą.

III-a tema: Plaučių navikus infiltruojančių imuninių ląstelių palyginimas tarp žmogaus ir pelės. Šiandieninių imunoterapijų sėkmė pateisino lūkestį, kad imuninės sistemos manipuliavimas gali būti naudojamas kaip gydymo strategija. Kadangi šios terapijos padeda tik daliai pacientų, toliau skiriamos intensyvios pastangos geriau suprasti imuninės sistemos ir vėžio sąveiką, bei kurti naujas gydymo strategijas (14). Tiek mechanizmų tyrimai, tiek vaistų paieška atliekama pelėse, o rezultatų aktualumas žmogui yra priklausomas nuo imuninės sistemos konservatyvumo tarp rūšių. Vis dėlto, nepaisant panašumų, dėl kurių pelės ir yra patrauklus modelinis organizmas, yra žinomi ir skirtumai. Nesunku nuspėti, kad pelėse veikiantis vaistas neveiks žmoguje, jei ląstelių tipas, į kurį vaistas nukreiptas, būdingas tik pelėms.

Savo projekte siekėme sistemingai identifikuoti žmogaus ir pelės panašumus ir skirtumus ląstelių populiacijų lygmenyje. Tokiam palyginimui būtina sąlyga buvo detalūs imuninių ląstelių tipų katalogai kiekviename iš organizmų. Kaip ir II-os temos atveju, scRNAseq suteikė galimybę pirmą kartą tokius katalogus sudaryti. Tyrimo metu susitelkėme į mieloidines ląsteles, kurios vėžio kontekste palyginti mažiau ištirtos už limfoidines (15). Didesnis dėmesys limfoidinėms, ypač T ląstelėms, suprantamas: citotoksinės T ląstelės specifiškai atpažįsta ir tiesiogiai naikina vėžines (16). Vis dėlto, mieloidinės ląstelės vėžio kontekste svarbios keliais aspektais: a) jos būtinos citotoksinių ląstelių aktyvavimui bei slopinimui; b) jos moduliuoja žaizdų gijimo procesus, kuriuos navikai gali panaudoti savo naudai; c) jos gausiai infiltruoja navikus ir nuo jų įvairovės navikuose priklauso pacientų išgyvenamumas, daugeliu atveju neigiamai (15).

Šiame projekte, pasinaudoję „inDrops“ technologija, sukūrėme detalius plaučių navikus infiltruojančių ląstelių katalogus ir palyginome juos tarp žmogaus ir pelės, atskleisdami panašumus ir skirtumus. Tikimės, kad šis darbas pasitarnaus tolimesniems tyrimams nukreiptiems į mechanizmų aiškinimąsi ir vaistų paiešką.

TIKSLAI IR UŽDAVINIAI

Tikslai:

I. Palengvinti „inDrops“ technologijos prieinamumą (I-a tema).

II. Sukurti išsamų kvėpavimo takų ląstelių tipų katalogą homeostatinėmis sąlygomis ir vykstant audinio regeneracijai (II-a tema).

III. Nustatyti, kurie plaučių navikus infiltruojančių mieloidinių ląstelių tipai tarp žmogaus ir pelės yra konservatyvūs (III-a tema).

Uždaviniai:

I-a tema

i) Identifikuoti ir ištaisyti priežastį, lemiančią techninį variabilumą tarp „inDrops“ eksperimentų.
ii) Paruošti detalų protokolą, aprašantį „inDrops“.

II-a tema

i) Sukurti išsamų kvėpavimo takų epitelio ląstelių tipų katalogą panaudojant scRNAseq.
ii) Nustatyti, kaip ląstelių tipų sudėtis kinta epiteliui atsistatant po pažaidų.

Trečias uždavinys buvo suformuluotas jau projekto metu, identifikavus naują ląstelių tipą – plaučių jonocitus – bei pastebėjus, kad jie yra pagrindinis *CFTR* transkriptų šaltinis:

iii) Nustatyti jonocitų padėtį audinyje, jų specifikacijos veiksnius ir CFTR jonų kanalo funkciją šiose ląstelėse.

III-a tema

i) Sukurti išsamius žmogaus ir pelės plaučių navikus infiltruojančių imuninių ląstelių katalogus.
ii) Identifikuoti mieloidinių ląstelių populiacijas, kurios konservatyvios tarp žmogaus ir pelės.
iii) Sukurti išteklį, kuris leistų platesnei mokslinei bendruomenei pasinaudoti surinktais duomenimis.

REZULTATAI IR IŠVADOS

I-as tikslas buvo pasiektas įgyvendinus šiuos **uždavinius**:

i) Optimizavus atvirkštinės transkripcijos žingsnį.
ii) Metodą detaliai aprašius griežtai struktūrizuotoje publikacijoje.

Kaip ir kitose scRNAseq technologijose, pirma fermentinė reakcija, taikant „inDrops“, yra atvirkštinė transkripcija (AT), kurios metu poliadenilintos RNR molekulės transkribuojamos į kopijinę DNR (kDNR). Svarbiausios optimizacijos, kurias atlikau, susijusios būtent su AT žingsniu. Pirma, buvo nustatyta, kad magnio jonų koncentracija reakcijos mišinyje nepakankama ir ji buvo padidinta iki 5 mM. Antra, AT pradmenys turėjo priemaišų, kurios lėmė pašalinių produktų susidarymą. Pradmenų sintezės protokolas buvo modifikuotas taip, kad būtų pašalintos priemaišos.

II-as tikslas buvo pasiektas įgyvendinus šiuos **uždavinius**:

i) Taikant scRNAseq, buvo sudaryti detalūs žmogaus ir pelės kvėpavimo takų epitelio ląstelių katalogai ir atskleistas naujas ląstelių tipas – plaučių jonocitai.
ii) ScRNAseq metodu tiriant pelės kvėpavimo takų atsistatymą po pažaidos, buvo nustatytos dvi papildomos ląstelių populiacijos, laikinai atsirandančios audinio regeneracijos metu, bet neaptinkamos homeostatinėmis sąlygomis.
iii) Buvo pademonstruota, kad plaučių jonocitai yra pagrindinis CFTR jonų kanalo funkcinis šaltinis kvėpavimo takuose. Taip pat buvo nustatyta, kad Notch signalinis kelias yra reikalingas jonocitų specifikacijai ir kad FOXI transkripcijos veiksnio raiška nukreipia bazines ląsteles jonocitų diferenciacijos keliu.

III-as tikslas buvo pasiektas įgyvendinus šiuos **uždavinius**:

i) Buvo sudaryti detalūs žmogaus ir pelės plaučių navikus infiltruojančių imuninių ląstelių katalogai. Buvo nustatyta nauja, tik vėžio atveju aptinkama, neutrofilų populiacija, turinti vėžį skatinantį fenotipą.
ii) Mieloidinių ląstelių palyginimas tarp rūšių atskleidė tikslų atitikimą tarp monocitų ir dendritinių ląstelių, dalinį atitikimą neutrofilų populiacijose ir ženklius skirtumus tarp makrofagų.
iii) Duomenys buvo paviešinti įvairiais formatais, tarp jų buvo pateiktos ir internete prieinamos priemonės duomenims naršyti interaktyviai.

METODAI

Pagrindinis šios disertacijos metu taikytas eksperimentinis metodas – „inDrops“ pavienių ląstelių RNR sekoskaita. Tai daugiažingsnis metodas reikalaujantis įgūdžių mikroskysčių technologijoje, dirbant su žinduolių ląstelėmis, bei molekulinės biologijos technikų valdymo, kurios buvo ypač svarbios atliekant „inDrops“ optimizavimo eksperimentus. „InDrops“ metodas prasideda disocijuotų ląstelių suspencijos enkapsuliacija į nanolitrų dydžio lašiukus, o viso proceso pabaigoje gaunama sekoskaitai paruošta DNR biblioteka. I-ą straipsnį sudaro detalus metodo aprašymas. Šio protokolo buvo laikomasi atliekant scRNAseq eksperimentus, aprašytus šioje disertacijoje.

ScRNAseq duomenų analizė buvo atliekama „Python“ programine kalba daugiausiai naudojant standartines bibliotekas, tokias kaip „numpy“, „pandas“, „matplotlib“, „scipy“. Analizės detaliai aprašytos II-IV straipsnių metodiniuose skyriuose. Programinio kodo pavyzdžiai pagrindinėms disertacijoje naudotoms analizėms prieinami adresu github/rapolaszilionis/Sharir_et_al_2019.

ŽODINIAI PRANEŠIMAI TARPTAUTINĖSE KONFERENCIJOSE

I. 2018 01 03 žodinis pranešimas tema „New cell type discovery by single cell RNA sequencing“ tarptautinėje konferencijoje Vita Scientia, Vilniaus universitetas.

II. 2018 10 05 žodinis pranešimas tema „Revisiting the composition of airway epithelium by single cell RNA sequencing” tarptautinėje konferencijoje Baltic Biophysics Conference, Vytauto Didžiojo universitetas.

III. 2018 11 27 žodinis pranešimas tema „Droplet microfluidics technology for single-cell biology research“ tarptautinėje konferencijoje 4Bio Summit Europe, Roterdamas.

IV. 2019 02 26 žodinis pranešimas tema „Single cell RNAseq as a tool for cell type discovery“ tarptautinėje konferencijoje theCOINS, Vilniaus universitetas. Kviestinis (*keynote*) pranešėjas.

SANTRAUKOS LITERATŪROS SĄRAŠAS

1. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187-201.
2. Hogan BLM, Barkauskas CE, Chapman HA, Epstein JA, Jain R, Hsia CCW, et al. Repair and regeneration of the respiratory system: complexity, plasticity, and mechanisms of lung stem cell function. Cell Stem Cell. 2014;15(2):123-38.
3. McCauley KB, Alysandratos K-D, Jacob A, Hawkins F, Caballero IS, Vedaie M, et al. Single-Cell Transcriptomic Profiling of Pluripotent Stem Cell-Derived SCGB3A2+ Airway Epithelium. Stem Cell Reports. 2018;10(5):1579-95.
4. Rock JR, Hogan BLM. Epithelial progenitor cells in lung development, maintenance, repair, and disease. Annu Rev Cell Dev Biol. 2011;27:493-512.
5. Who. World health statistics 2008: World Health Organization; 2008.
6. Vidarsson H, Westergren R, Heglind M, Blomqvist SR, Breton S, Enerbäck S. The forkhead transcription factor Foxi1 is a master regulator of vacuolar H-ATPase proton pump subunits in the inner ear, kidney and epididymis. PLoS One. 2009;4(2):e4471.
7. Quigley IK, Stubbs JL, Kintner C. Specification of ion transport cells in the Xenopus larval skin. Development. 2011;138(4):705-14.
8. Ratjen FA. Cystic fibrosis: pathogenesis and future treatment strategies. Respir Care. 2009;54(5):595-605.
9. Boucher RC. Evidence for airway surface dehydration as the initiating event in CF airway disease. J Intern Med. 2007;261(1):5-16.
10. Andersen DH. Cystic fibrosis of the pancreas and its relation to celiac disease: a clinical and pathologic study. American journal of Diseases of Children. 1938;56(2):344-99.
11. Rommens JM, Iannuzzi MC, Kerem B, Drumm ML, Melmer G, Dean M, et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science. 1989;245(4922):1059-65.
12. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science. 1989;245(4922):1066-73.
13. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, et al. Identification of the cystic fibrosis gene: genetic analysis. Science. 1989;245(4922):1073-80.
14. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24(5):541-50.
15. Engblom C, Pfirschke C, Pittet MJ. The role of myeloid cells in cancer therapies. Nat Rev Cancer. 2016;16(7):447-62.
16. Sharma P, Allison JP. The future of immune checkpoint therapy. Science. 2015;348(6230):56-61.

The electronic version of this thesis excludes copies of the 4 publications forming the basis of the thesis.

Elektroninėje disertacijos versijoje 4 publikacijų, sudarančių disertacijos pagrindą, kopijos neįtrauktos.

UŽRAŠAMS

Vilniaus universiteto leidykla
Universiteto g. 1, LT-01513 Vilnius
El. p. info@leidykla.vu.lt,
www.leidykla.vu.lt
Tiražas 11 egz.