SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Medicine and Surgery
PhD program in Translational and Molecular Medicine DIMET
Cycle XXXV

Tissue-based high-dimensional landscaping in inflammation, repair and transformation

Dr.ssa Bolognesi Maddalena Maria
Matr. No. 069860
Tutor: Prof. Giorgio Cattoretti
Co-tutor: Prof. Fabio Pagni
Coordinator: Prof. Andrea Biondi

ACADEMIC YEAR 2021/2022

To Matilde

Table of Contents

Chapter 1 pag. 6
General introduction
Scope of the thesis
Bibliography

Chapter 2 pag. 19
Title: **The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas**

Chapter 3 pag. 51
Title: **The normal and fibrotic mouse lung classified by spatial proteomic analysis**

Chapter 4 pag. 83
Title: **Next-Generation Pathology Using Multiplexed Immunohistochemistry: Mapping Tissue Architecture at Single-Cell Level**

Chapter 5 pag. 148
Title: **BRAQUE Bayesian Reduction for Amplified Quantization in Umap Embedding**

Chapter 6 pag. 197
Title: **Spatial single cell high-plex fluorescence imaging classifies normal lymphoid tissue.**

Chapter 7 pag. 231
Summary
Conclusions and Future Perspectives
Bibliography

List of publications pag. 236

Chapter 1

General Introduction

Single cell biology [1] has revolutionized the way we understand biological processes and a new era is started.

Before 2013 very few biological assays were informative on single cell biology (e.g. Flow Cytometry) and required disaggregation to single cells or homogenates for molecular studies. Typical of that era was the substantial progress in tumor biology obtained with cDNA microarrays technology. The entry of single cell RNA sequencing changed the paradigm.

In short, this technology can be explained as the detection of the same traits in every single cell in a sample, not just as the mean value of the bulk. This revolution has added a new level of resolution to what we can “see” and makes us better understand the complexity of a sample made of different entities: cells.

In medicine, in particular, the single cell approach is revealing still unknown mechanisms in health and pathologic tissues to improve, in the end, healthcare. Characterization of single cells has increased our understanding of cell phenotypes, the dynamics and trajectory of their development, and their communication network.

All of this has been possible thanks to the fast development of a large group of technologies: “single-cell multi omic technologies”.

Each of them provide a very specific kind of information for every cell analyzed based on genomic, transcriptomic,

proteomic, metabolomic signals, not to mention spatial approaches that associate to cell identity its x and y localization within the tissue. This, together with information related to neighboring cells, [2] is crucial because cell identity and role are determined by the context.

Integration of the data in the "single-cell multimodal omics"[3] [4] [5] is the subs step to deeply characterize cells and cell types.

The whole story can be recapitulated by citing the last "Nature Methods of the year" headlines: 2013 Single cell sequencing [6], 2019 Single cell multimodal-omics[3] , 2020 Spatially resolved transcriptomics [7], 2022 "among others to watch: Tracing cell relationship[8]".

All of this is impacting the way of doing research and new bioinformatic tools are indispensable to explore single cell data.

Single cell analysis was born as the discipline which has to deal these types of big data sets, with thousands of variables, to be managed with unsupervised/supervised approaches, clustering, integration, and which requires different expertises: biology, informatics and statistics and, eventually, astrophysics[9].

While these technologies are flourishing, many global research efforts have been promoted to collect and share an enormous amount of "single cell data" and to make them publicly available in order "To create comprehensive reference maps of all human cells—the fundamental units of life as a basis for both understanding human health and diagnosing, monitoring, and treating disease " (Human Cell Atlas, [HCA] Mission)

Some large consortia have been created and represent now the standard and the repository to which refers to. Among them the most important are the Human Cell Atlas (HCA)[10] , the BRAIN Initiative Cell Census Network (BICCN)[11] and the Human Biomolecular Atlas (HubMAP)[12] .

Among all of these novelties, high-plex spatial proteomic represents a small but crucial niche for many reasons: it has a single cell resolution associated with spatial localization, it evaluates the presence of protein and not RNA signal (bypassing post transcriptional modification), it analyses whole cells without losing any types of them[13] as often happens in tissue disaggregation [14] or other single cell technologies, it allows retrospective study on FFPE sample collections, it is easily integrable with other -omic technologies.

Thanks to these global initiatives and to the commercializations of these technologies, we have witnessed the need to approach these big data from a bioinformatic point of view, in order to standardize analysis and make them available.

In the last few years more than 1000 bioinformatic tools have been developed, mainly in the two world-wide used interpreted languages, R and Python [15]. As already said, computational method approach[16] is absolutely relevant.

My PhD thesis develops within this scenario. Briefly it concerns the analysis of human or mouse, healthy or pathologic tissues at single cell level in order to classify cell populations, quantify and localize them i*n situ* and recognize their specific interaction and characteristics.

-The MILAN technology

In 2017 we developed a method called **Multiple Iterative Labeling by Antibody Neodeposition (MILAN technology)** [17] "which implies multiple stainings of a tissues section with multiple antibodies on controlled antigen retrieval conditions (once, at the beginning), antigen retention over the staining cycles, six-color immunofluorescent staining, bringing the total amount of stainings into the dozens. It employs unconjugated primary antibodies, commercial secondary antibodies and IF microscopes and scanners."

Since its publication many efforts have been made to improve the technology and it has been applied in other labs (Leuven KUL University with Prof. Bosisio, Technical University of Athens with Prof. Alexopoulos,National, University of Parma, Veterinary Department, with prof. Ravanetti).

Individual labs created dedicated pipelines of high-dimensional analysis, applied to different types of tissue.

At the moment the state of the art of the MILAN technology in Milan Bicocca University stands as follows:

-standard protocol, published, now version 5 (https://doi.org/10.21203/rs.2.1646/v5)[18]

-list of antibodies (https://doi.org/10.21203/rs.2.1646/v5)[18]

-scanner (Hamamatsu S60 with these filter dataset) see suppl material in: [19]

-two high performance workstations for the data analysis and backup:

- HP Z8 G4 90 workstation, Intel Xeon 6130 2.1 2.6GHz 16 core x2, with memory 192 GB RAM (12x16 GB) NVIDIA Quadro P6000 24 GB RAM,

-HP Z640 Workstation, Intel Xeon E5-2667v3 3.2 2133 8C 1stCPU 64GB DDR4-2133 (4x16GB) 1CPU Registered RAM NVIDIA Quadro K620 2GB DL-DVI(I)+DP,

-three NAS Synology, two units, 4 disks 4 Tb each and one unit with up to 8 disks 14Tb each. Cloud storage is provided for data backup.

-softwares used (Imagej, Cell profiler, Cell pose, Matlab, R, Python)

This dissertation is organized as a selected collection of papers based on MILAN technology and its subsequent analysis implementation : Chapters 2-3 are 2 different papers which use MILAN to analyze 2 different type of tissue (human uterus, mouse lung) to evaluated respectively the immune-infiltrate components and the single cell classification of a mouse lung-fibrosis model, including the analysis of a drug transporter. Chapter 4 is a review of similar technologies with our major collaborator group. Chapter 5 is a proposed algorithm of analysis, Bayesian Reduction for Amplified Quantization in Umap Embedding (BRAQUE), tailored for high-plex spatial proteomic data from images. Chapter 6 is a preliminary classification of cell types in human lymph nodes. While chapters 2 and 3 belong to completed projects, and chapter 4 is

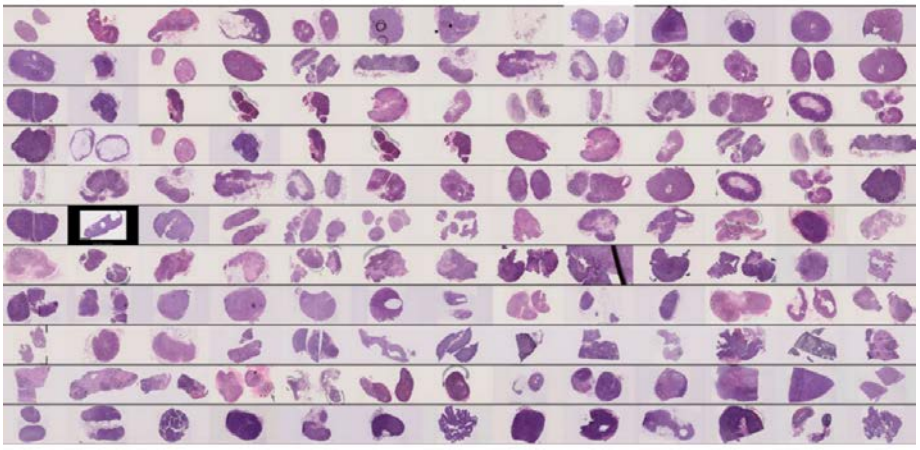
a review, the last part (chapters 5-6) is still ongoing and needs to be contextualized. After having analyzed with the MILAN technology different kinds of human and mouse tissue (lung, kidney, colon, uterus,) we focused the attention on lymphoid tissue for its key role in immune response. The goal was to classify phenotypically each cluster/cell population within the tissue and to localize it in lymph node architecture.

Lymph Node project

Lymph nodes have an intrinsic density (1mm3 of lymph node tissue approximately 1 to 2 × 10^6 cells)[20] and consent an easy disaggregation for single cell RNA sequencing for recirculating cells but not for sessile elements (stromal cells, some macrophages, dendritic cells etc.), making difficult the complete analysis of all cell types. For this reason they aren't the best candidate for classical single cell RNA sequencing. Differently, spatial technologies are suitable for them.

For the project more than 100 samples of lymph nodes were collected and digitally organized in a database (http://www.tiny.cc/LNproject) . They represent a broad spectrum of variability that can be found in humans, including viral, granulomatous and immune reactions, excluding cancer (Fig1). The majority of cases were non random sampled and organized in 2 cores of 2 mm of diameter (~40-50000 cells per core) in a TMA, 7 samples were analyzed on a full section (~1 millions cells each).

The whole project



12	LNS	sentinel lymph node (normal)
2	Normal	no pathology
13	Paracortical	LN hyperplasia, T cell zone (paracortex)
2	Progressive transf	Large atypical Germnal Centers in follicular hyperplasia
7	dermatopathic	cutaneous antigens-driven LN dendritic cell hyperplasia
2	Sinus hist	enlargement of LN sinuses, hyperplastic
2	autoimm	autoimmune disease
2	crohn	chronic autoimmune disease
4	IgG4	autoimmune disease
4	Inflpsdt	mostly stromal proliferation, autoimmune
6	cat scratch	viral-induced granulomatous reaction
7	Granulomat	Monocyte-macrophage proliferation with granulomas
3	Kikuchi	self-healing monocyte-macrophage disease
3	Sarcoid	autoimmune-driven granulomatous disease, TB-like
11	TB	tuberculosis
1	LCH	Langerhans cells (dendritic) neoplasm
13	castleman	HHV8 driven stromal and plasmacell proliferation
4	EBV	EBV infection
8	EBV (Ptx)	Post Transplant (mostly EBV driven)
7	EBV (HD)	Hodgkin's disease (lymphoid neoplasm)
11	EBV/HHV8	mixed EBV and/or HHV8 infection
6	HIV	HIV infection
2	Viral	mixed viral infections

Fig. 1

The major effort of this work was put on the standardization and optimization of image analysis strategies, thanks to the collaboration with other research groups, in particular for images management Dr. Mario Faretta (IEO institute, Milan), Prof. Simone Borghesi (University of Milan Bicocca) for the segmentation part , the group of prof. Gastone Castellani group (University of Bologna) for the bioinformatic tools of analysis.
In the end of the thesis there will be some considerations about single cell classification and future perspective of this research.

Scope of the thesis

Chapter 1
Title: **General introduction**

Chapter 2
Title: **The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas**
Paper published whose focus is the phenotypic classification at single cell level of innate and adaptive inflammatory infiltrates in uterine leiomyosarcoma. [21]

Chapter 3
Title: **The normal and fibrotic mouse lung classified by spatial proteomic analysis**
Paper published whose focus is the analysis of in situ spatial proteomics by multiplex antibody staining to routinely

processed mouse lung, healthy and during a fibrosis model. It shows intrinsic cell types in the progress of lung fibrosis over a 4 weeks course, the changing landscape and the cell-specific quantitative variation of a multidrug transporter. [22]

Chapter 4

Title: **Next-Generation Pathology Using Multiplexed Immunohistochemistry: Mapping Tissue Architecture at Single-Cell Level**

Review published concerning multiplex IHC staining methodologies aimed at single cell-spatial analysis and role of high dimensional multiplex in the next generation pathology.[23]

Chapter 5

Title: **BRAQUE: Bayesian Reduction for Amplified Quantization in Umap Embedding**

Manuscript in preparation. It describes a new and tailored approach of bioinformatic analysis to explore lymphoid tissue populations as single cells.

Chapter 6

Title: **Spatial single cell high-plex fluorescence imaging classifies normal lymphoid tissue**

Manuscript in preparation. It provides single cell analysis of a large database of human “normal” lymph nodes with an expanded catalog of more than 80 antibodies. It contains a granular classification of cell population, with new cell types,

their spatial distribution and a new segmentation algorithm tailored to lymphoid tissue.

Chapter 7

Summary

Conclusion and Future perspectives

Publications

References

1. Schier AF. Single-cell biology: beyond the sum of its parts. *Nat Methods*. 2020;17(1):17-20. doi:10.1038/s41592-019-0693-3
2. Cho CS, Xi J, Si Y, et al. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell*. 2021;184(13):3559-3572.e22. doi:10.1016/j.cell.2021.05.010
3. Zhu C, Preissl S, Ren B. Single-cell multimodal omics: the power of many. *Nat Methods*. 2020;17(1):11-14. doi:10.1038/s41592-019-0691-5
4. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29. doi:10.1016/j.cell.2021.04.048
5. Jain MS, Polanski K, Conde CD, et al. MultiMAP: dimensionality reduction and integration of multimodal data. *Genome Biol*. 2021;22(1):346. doi:10.1186/s13059-021-02565-y
6. Editorial. Method of the Year 2013. *Nat Methods*. 2014;11(1):1-1. doi:10.1038/nmeth.2801
7. Marx V. Method of the Year: spatially resolved transcriptomics. *Nat Methods*. 2021;18(1):9-14. doi:10.1038/s41592-020-01033-y
8. Mukhopadhyay M. Tracing cell relationships. *Nat Methods*.

2022;19(1):27-27. doi:10.1038/s41592-021-01370-6

9. Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. doi:10.1126/science.aba2609
10. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The Human Cell Atlas: from vision to reality. *Nature*. 2017;550(7677):451-453. doi:10.1038/550451a
11. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*. 2021;598(7879):86-102. doi:10.1038/s41586-021-03950-0
12. HuBMAP Consortium, Writing Group, Snyder MP, et al. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature*. 2019;574(7777):187-192. doi:10.1038/s41586-019-1629-x
13. Hor JL, Germain RN. Intravital and high-content multiplex imaging of the immune system. *Trends Cell Biol*. 2022;32(5):406-420. doi:10.1016/j.tcb.2021.11.007
14. Denisenko E, Guo BB, Jones M, et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol*. 2020;21(1):130. doi:10.1186/s13059-020-02048-6
15. Zappia L, Theis FJ. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biol*. 2021;22(1):301. doi:10.1186/s13059-021-02519-4
16. Efremova M, Teichmann SA. Computational methods for single-cell omics across modalities. *Nat Methods*. 2020;17(1):14-17. doi:10.1038/s41592-019-0692-4
17. Bolognesi MM, Manzoni M, Scalia CR, et al. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *J Histochem Cytochem*. 2017;65(8):431-444. doi:10.1369/0022155417719419

18. Cattoretti B. Multiple Iterative Labeling by Antibody Neodeposition (MILAN). *Protoc Exch*. doi:https://doi.org/10.21203/rs.2.1646/v5
19. Mascadri F, Bolognesi MM, Pilla D, Cattoretti G. Rejuvenated Vintage Tissue Sections Highlight Individual Antigen Fate During Processing and Long-term Storage. *J Histochem Cytochem*. 2021;69(10):659-667. doi:10.1369/00221554211047287
20. Cupedo T, Stroock A, Coles M. Application of tissue engineering to the immune system: development of artificial lymph nodes. *Front Immunol*. 2012;3. doi:10.3389/fimmu.2012.00343
21. Manzoni M, Bolognesi MM, Antoranz A, et al. The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas. *Sci Rep*. 2020;10(1):702. doi:10.1038/s41598-020-57627-1
22. Ciccimarra R, Bolognesi MM, Zoboli M, Cattoretti G, Stellari FF, Ravanetti F. The normal and fibrotic mouse lung classified by spatial proteomic analysis. *Sci Rep*. 2022;12(1):8742. doi:10.1038/s41598-022-12738-9
23. Bosisio FM, Van Herck Y, Messiaen J, et al. Next-Generation Pathology Using Multiplexed Immunohistochemistry: Mapping Tissue Architecture at Single-Cell Level. *Front Oncol*. 2022;12:918900. doi:10.3389/fonc.2022.918900

Chapter 2

The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas

Marco Manzoni 1,6, Maddalena M. Bolognesi 1,6, Asier Antoranz 2, Rosanna Mancari 3, Silvestro Carinelli 3, Mario Faretta 4, Francesca M. Bosisio 5 & Giorgio Cattoretti 1*

1 Pathology, Department of Medicine and Surgery, Università di Milano-Bicocca, Via Cadore 48, Monza, (MI), Italy. 2 National Technical University of Athens, Zografou Campus, 9 Iroon Polytechniou str., 15780, Zografou, Athens, Greece. 3 Division of Gynecologic Oncology and Pathology, European Institute of Oncology, Via Ripamonti 435, 20141, Milan, Italy. 4 Department of Experimental Oncology, European Institute of Oncology, via Adamello 16, 20139, Milan, Italy. 5 Laboratory of Translational Cell andTissue Research, KU Leuven, Herestraat 49, 3000, Leuven, Belgium. 6 These authors contributed equally: Marco Manzoni and Maddalena M. Bolognesi. *email: giorgio.cattoretti@ unimib.it

Scientific RepoRtS | (2020) 10:702
https://doi.org/10.1038/s41598-020-57627-1

Reactivation of the anti-tumor response has shown substantial progress in aggressive tumors such as melanoma and lung cancer. Data on less common histotypes are scanty. Immune checkpoint inhibitor therapy has been applied to few cases of uterine leiomyosarcomas, of which the immune cell composition

was not examined in detail. We analyzed the inflammatory infiltrate of 21 such cases in high-dimensional, single cell phenotyping on routinely processed tissue. T-lymphoid cells displayed a composite phenotype common to all tumors, suggestive of antigen-exposure, acute and chronic exhaustion. To the contrary, myelomonocytic cells had case-specific individual combinations of phenotypes and subsets. We identified five distinct monocyte-macrophage cell types, some not described before, bearing immunosuppressive molecules (TIM3, B7H3, VISTA, PD1, PDL1). Detailed in situ analysis of routinely processed tissue yields comprehensive information about the immune status of sarcomas. The method employed provides equivalent information to extractive single-cell technology, with spatial contexture and a modest investment.

The adaptive immune system has evolved into a very refined and complex coordination of multiple actors (T, B, dendritic, NK cells etc.) devoted to the control of exogenous attackers, while avoiding the collateral damage of the self1. The onset of cancer affects both the adaptive and the innate immune system via multiple mechanisms: by increasing the cell turnover and mass, by creating new vascularized tissue, by presenting new antigens in a previously tolerant scenario and by recruiting heterotopic inflammatory cells2. How the adaptive immune system deals with cancer has recently gained attention because of the promising clinical results with personalized medicine targeting the T-cell response against tumors3,4. The immune

reaction against the tumor is placed in check by the combined action of tumor escape and naturally occurring mechanisms which dampen the immune response3–8. These mechanism may co-exist and function independently9 and the result is a cancer immunogram in which each tumor has a combination of the various components 4,7. Inhibition of blocking immune checkpoints via therapeutic antibodies restores a pre-existing anti-tumor T-cell response and results in prolonged remission or cure of otherwise lethal cancers 10–12. Not all cancer patients respond to a checkpoint inhibitor therapy. The tumor mutational burden, i.e. the ability to present neoantigens to the adaptive immune system, has been identified as one major biomarker predictive of response 10,12. Hypermutated tumors are good candidates for this therapy13. An altered DNA copy number or ongoing DNA damage repair mechanisms 14,15 may also recruit intratumoral lymphocytes (TILs). Distinctive modules of inflammation shared by diverse cancer histotypes have been revealed by pan-cancer analysis of deposited gene-expression databases 16,17: these studies have shown that tumor mutation burden and CD8 TIL infiltration have an impact on the prognosis, but a macrophage signature may also affect the outcome17. Some studies have identified a macrophage signature independent of tumor type18, yet the awareness of the complex regulation of macrophage biology may suggest otherwise19.

Leiomyosarcomas originating from myometrium (ULMs), have not been intensively investigated because of their low incidence

and because they are perceived as a minor target for immune intervention, given the few tumor TILs on H&E sections and the low mutational burden20. As few as two dozens ULMs have been treated with checkpoint inhibitors, with dismal results 21–23. However, with the exception of one responsive case22, in none of them the TILs or the macrophages have been thoroughly studied. Here we present a comprehensive high-dimensional analysis of the inflammatory infiltrate in 21 sarcoma cases with a panel of 40 markers including lineage specific leukocyte proteins, activation markers and component of the immunological synapse, by using a novel robust method, effective on routinely processed materials, but capable of highly detailed single cell analysis.

Materials and Methods

Patients and case selection.

21 cases of leiomyosarcoma were selected out of 77, based on both full clinical history and tissue block availability. The female patients were aged 51.2 ± 11.8 years (34–69), 52% post-menopausal, 11 FIGO stage IB, 1 stage IIB, 5 stage IIIB/C, 3 IVB, one unknown. 13 were classified TNM pT1, 1 pT2, 6 pT3. The biological parameters are reported in Table 1. The study has been approved by the Institutional Review Board Comitato Etico Brianza (https://www.asst-monza.it/comitato-etico), N. 3204, “High-dimensional single cell classification of pathology (HDSSCP)”, October 2019. Patients consent was obtained or waived according to article 89

of the EU general data protection regulation 2016/679 (GDPR) and decree N. 515, 12/19/2018 of the Italian Privacy Authority. Two random 2 mm cores of non-necrotic, non-sclerotic, full tumor tissue for a total of 6.28 mm2 per case (equivalent to 40 HPF) were placed in a Tissue Microarray block (TMA model CK4500, Integrates Systems Engineering srl, Milan, Italy).

Multiple iterative labeling by antibody neodeposition (MILAN). Dewaxed, antigen retrieved 3 µm sections were processed for indirect IHC or multiple labeling as previously described in detail24,25 and Supplementary Data. Briefly, the sections were incubated overnight with optimally diluted primary antibodies in combination of four, washed and counterstained with four distinct fluorochrome-tagged secondary antibodies25. The slides, counterstained with DAPI and mounted, were scanned on an S60 Hamamatsu scanner (Nikon, Italia) at 20x magnification, after which the stains were removed with a beta-mercaptoethanol and sodium dodecyl sulphate mix, extensively washed and re-staind for the subsequent markers25. Re-staining a sample of antigens after about 30 cycles showed no consistent antigen loss and occasional increased antigenicity (Supplementary Fig. 1). The list of primary and secondary antibodies is available in the Supplementary Materials.

Preparation of immunofluorescent images for image analysis. Single.ndpi images for each case were registered via alignment of DAPI nuclear stained images with Fiji, saved as.tiff files and autofluorescence was subtracted 25. Two single cell

masks were produced for each case with Cell Profiler (2.2.0)26 by segmentation of thresholded images: a DAPI mask encompassing all the nuclei and a mask obtained by the combined CD43, CD45, CD68 and CD163, henceforth named “targeted mask”. Regions of interest composed of small portions of dendritic-shaped cytoplasm were excluded digitally (see Supplementary Methods). Comparison of the total cell yield and detailed high-dimensional phenotypic content obtained with the DAPI vs the targeted masks showed variable but constantly inferior cell number representation with the former (36%, 65%, 95%of the targeted mask-derived cells) and loss of minor phenotypic subsets by tSNE and Phenograph clustering (not shown). This because macrophages have reduced chromatin DAPI staining and do not provide enough contrast to be detected by threshold algorithms.

High-dimensional analysis.

Individual masks and .tiff files from all cases were loaded in HistoCAT27, data obtained from the image analysis were processed by dimensional reduction and unsupervised clusterization algorithms, t-SNE and Phenograph plots were generated. Image analysis data were subsequently exported as .csv files. In order to analyze the infiltrate composition of individual cases, we developed an R script (Supplementary Methods) to systematically process all cells of all cases identified by image analysis. Phenoclusters were plotted as heatmaps (Supplementary Fig. 2) with a custom R script, to allow the identification of cell composition. Each heatmap was

inspected for specific lineage- or function-associated markers, with the requirement that each defining marker needed to be expressed at high levels (clearly visualizable by a blue-red divergent color palette) in a third or more of the cells. A nine cell-type classification of inflammatory infiltrate was obtained by the inspection of all generated heatmaps and was based on criteria listed in Supplementary Table 1. The cell content of each phenogroup was used as the numerator to quantify the percentage of a given marker or phenotype.

Neighborhood analysis. An unbiased quantitative analysis of cell-cell interactions was performed using an adaptation of the algorithm described in27 for neighborhood analysis to systematically identify social networks of cells and to better understand the tissue microenvironment (Supplementary Methods).

PTEN status by FISH. In situ hybridization for PTEN was performed with the ZytoLight SPEC PTEN/ CEN 10 dual color probe (ZytoVision GmbH, Germany) for the centromeric and the gene-specific regions of chromosome 10.

Results

The clinicopathologic data of the 21 sarcomas are reported in Table 1.

The inflammatory infiltrate. High-dimensional analysis of all 21 cases showed a majority of independent, non-overlapping clusters of myeloid phenotype, one or two per case, and smaller overlapping clusters, comprising T-cells and endothelial cells

(Fig. 1). Only in four instances (cases N. 17,18, 20, 21) myeloid phenoclusters from separate cases did overlap (Fig. 1). Thus, the majority of the infiltrating inflammatory cells in each case is composed of macrophages whose phenotype reflects the unique biology of each tumor (Fig. 2), and a minor population of T-cells. In order to understand the composition of the inflammatory infiltrate, each sarcoma case was analyzed separately in high-dimension (Supplementary Figs. 2 and 3).

Lymphoid cells

TILs, almost exclusively T-cells and NK-cells, represents 3%-29% of the inflammatory infiltrate (0.3%-15.3% of the total sample cellularity), the rest being myelomonocytic cells (Fig. 2, Supplementary Table 2 and Supplementary Data). A few B cells in one case and no plasma cells were identified. TILS were composed of 30% ± 22% CD4+, 62% ± 23% CD8+ and 9% ± 8% NK-cells. CD4+ T-cells were 68% ± 36% FOXP3+, largely negative for activation markers (OX40, CD69,CD32). (Fig. 2, Supplementary Table 2 and Supplementary Data). CD8+ T-cells were identified as distinct phenoclusters in about half of the cases, whenever a sufficient number of TILS was present. In those cases, often multiple phenotypically distinct phenoclusters were detected per case, displaying evidence of activation (CD69) and exhaustion (PD1, TIM3, VISTA, CD39). VISTA+ T-cells were observed in 8 cases, largely CD8+ TCF7−. TCF7, a transcription factor linked to resident memory phenotype and reactivation, was contained in 42% ± 18% of CD8+ cells, in an inverse relationship with PD1 (Figs. 3 and 4).

Lymphocytes expressing GranzymeB and Granulysin, partly CD8+, were identified as a separate phenocluster in some cases or as part of a single cluster of cells in samples with fewer TILs. NK cells, defined as CD45+ CD3− and expressing GranzymeB and Granulysin, were a minority of lymphoid cells in abundant infiltrates only.

Myelomonocytic cells

Myelomonocytic cells represents 64% ± 13% of the infiltrate (Fig. 2, Supplementary Table 2 and Supplementary Data) and contained the most diversified cell type, often represented by multiple distinct phenogroups within each case (Fig. 5). The most distinctive group, named Tumor Associated Macrophages (TAM; 38% ± 13%) expressed CD16 and restricted lineage markers (CD68, CD163) and was consistently present in all tumor cases, but only occasionally in the non-neoplastic samples (Case #8 and 3 normal myometria, not shown). TIM3 expression among myelomonocytic cells was restricted to this cell type and to the inflammatory monocytes (see below). The sarcomas contained CD16-negative histiocytes (20% ± 17%) and phagocytes (CD68+ CD163−; 5% ± 10%), these latter in 7/21 cases. Phagocytes and Undefined Monocyte-Macrophagic cells (9% ± 12%) were characterized by the absence or spotty presence of several monocytic lineage markers (CD14, CD64, lysozyme/LYZ, CD163), PD1, PDL1, VISTA, CD83, B7H3. All cases but five contained a small but very distinct population of small monocytic cells, often intravascular and proliferating, with a distinctive LYZ+ VISTA+ phenotype, which we defined

“inflammatory” because of the association with the inflammatory infiltrate. These cells were often CD14+, CD16low or negative, sometimes TIM3+, but devoid of other myelomonocytic markers.

Spatial relationships between lymphoid and myelomonocytic cells. A neighborhood analysis between all phenotypic subsets revealed a substantial mutual avoidance across the phenotypes and clustering together of similar cell types, both at submicroscopic (<100 μm range) and at microscopic range (>100 μm range) (Supplementary Fig. 4). TILs avoid close contact with checkpoint-bearing macrophages (Supplementary Fig. 4 and Methods).

Tumor cells. Tumor cells had histocompatibility antigens staining patterns which could be defined as constitutive and induced; in the first case tumor cells were all positive or negative throughout the sections (Supplementary Fig. 5). In the second case, tumor cells admixed with the inflammatory infiltrate showed increased staining, compared to non-inflamed portions (Supplementary Fig. 5). No HLA-DR was observed on tumor cells. Tumor cells did not expressed most of the markers tested, including PDL1, with the notable exception of MCM5, IDO, B7H3 and Axl (see Table 1 and Supplementary Fig. 5). Occasional PD1+ tumor cells were observed in some cases (Fig. 5B and Supplementary Data). CD34+ endothelial cells stained for VISTA and B7H3 in about half of the cases, often in a subset28 (Supplementary Fig. 5). The vast majority of the cases tested were either diploid or had loss of one copy of the

PTEN gene in 3.1%22.1% of the cells (Table 1), with only one case (case #7) carrying homozygous PTEN deletion in 13.2% of the cells.

Discussion

Data about the inflammatory infiltrate in untreated primary sarcomas are scarce. We detail here the composition of the innate and adaptive arm of the response of the host to a soft tissue tumor. We found a variety of individual phenotypic profiles, not unlike other reports29,30. In 11/21 cases we found a T-cell phenotype consistent with antigen exposure and acute or chronic stimulation, leading to exhaustion, mostly in tumors HLA Class I+. The remaining 10 cases that we classified as “deserted” had usually less than 60 lymphocytes per mm2 and 0.3%-4.2% of total cellularity, with a phenotype suggestive of passer-by. These tumors were all negative or weakly expressing HLA-A,B,C, with two exceptions. Each of the cases hosting antigen-experienced T-cell phenotypes displayed an unique combination of cell markers, which could be summarized as acutely exhausted CD8+ T cells, chronically exhausted CD8+ and FOXP3+ CD4+ T cells. These phenotypes were reminiscent of T-cells acutely exposed to a persistent antigen such as a virus or a neoantigen, bearing PD1, CD69, CD3931,32 and TIM333, and induced to anergy (exhaustion)3. Others expressed TIM3, lower levels of PD1 and TCF7 to a variable amount (17–73%), resembling exhausted CD8+ T-cells chronically exposed to an antigen, but in a resident memory

state, susceptible of reactivation upon re-stimulation or therapeutic checkpoint inhibition reversal33,34. A mixture of these two cell types was sometimes present. It has been reported that up to a third of the cases displaying a CD8+ TCF7+ phenotype may benefit from a checkpoint blockade therapy33. CD4+ T-cells had a FOXP3+ regulatory phenotype, occasionally as the minor population, as shown in other cancer models35. The variability of T-cell phenotypes found between and within cases is suggestive of an ongoing editing of the adaptive immune response36, on a case-by-case basis. The tumors we have examined are full blown malignant tumors of low mutation rate, yet, as shown in a single successful immunotherapy case22, this may induce an anti-tumor T-cell response. ULMs are also conspicuously devoid of HLA-DR+ CD14− dendritic cells and tertiary lymphoid structures, thus an adaptive response may occur in tumor-draining lymph nodes. The failure of single-agent checkpoint inhibitor therapy in several published cases21–23 is at odd with our findings. One hypothesis is that the antigen to which the adaptive immune system responds is not a tumor antigen37. In the single ULM case studied22, T-cells were tumor antigen-specific and responded. None of our cases except one (Table 1) had homozygous deletion of PTEN, associated with an immunosuppressive tumor phenotype22.

Alternatively, failure to elicit an anti-tumor response may be caused by an independent immunosuppressive effect brought by the inflammatory infiltrate, chiefly the infiltrating

macrophages. Macrophages, and in particular tumor-associated macrophages (TAMs) have been shown to mediate the suppression of an anti tumor response38,39. A dichotomic view of alternatively polarized macrophages has let to a more nuanced picture40, where there is a dynamic equilibrium between various defined stages of macrophage polarization. As published by others16,41, we found a range of leukocytes (CD45+, CD43+) with myelomonocytic differentiation. Differently from lymphocytes, whose aggregate phenotype was conserved across all cases, myelomonocytic cells had unique, case-specific populations. A tissue-restricted secretome influencing macrophages has been described19; a similar effect may occur via monoclonal sarcoma cells, yielding a highly diversified, stimulus-driven differentiation reflecting individual tumor-specific microenvironment30,42,43. As shown by others30,38,44, the tumor harbors multiple subsets, often not found in the normal counterpart neither described before. Furthermore, our neighborhood analysis of the macrophage subsets show both a submicroscopic and a microscopic local differentiation, producing a checkered pattern, which has to be kept in mind in order to sample the tissue adequately, as we did with larger (2 mm), multiple TMA cores. Despite the large variety of individual phenotypes, by analyzing single cases with multiple markers we could identify five rather consistent groups present in all cases: histiocytes, phagocytes, TAM, inflammatory monocytes and Monocytes/Macrophages, undefined. With the exception of TIM3, largely restricted to CD16+ TAMs, all the

immuno-modulating markers were expressed in a minority of cells and were distributed across the five subsets. We observed PD1, previously reported in circulating monocytes in HIV and in M2-type TAMs45 expressed on macrophages in 18/21 cases, occasionally with activation markers such as HLA-DR, OX40 and/or CD83. Interestingly, non-lymphoid PD1 expression in FFPE material could only be detected with UMAB197; this antibody has a broader reactivity than other PD1 antibodies and detects the molecule on B cells, monocytes and tumor cells, as reported by others45,46. Subtle subcellular variations in membrane staining by each antibody we tested (Supplementary Fig. 6) may has to do with the recognition of glycosylation-dependent, thus cell type restricted epitopes on PD1. Consistent with the mouse data45, PD1 was expressed on 27 phenogroups, only two of them HLA-DR+ and CD83+. Markers of activation (HLA-DR, CD83, OX40) were found on macrophages in almost all cases in all subsets, except on the inflammatory monocytes. Often activation was matched with proliferation (MCM5+). More complex to understand was the expression of members of the immunological synapsis (PD1, PDL1, B7H3, VISTA), which, with the exception of TIM3, restricted to TAMs, were displayed by several subsets. PDL1, expressed by myelomonocytic cells in 15/21 cases, is a ligand for PD1 and CD80 on cognate T-lymphocytes6,7 and a constitutively negative signaling molecules on macrophages47. Engagement of PDL1 induces proliferation, survival and upregulation of MHC Class-II, CD86 and cytokine secretion,

promoting a proinflammatory phenotype47. In our PDL1-negative sarcoma cases, there is little relationship between activation and PDL1 expression in macrophages; of the 24 PDL1+ phenogroups, only seven express HLA-DR and eight CD83. These data hint at a lack of engagement of this molecule on macrophages. VISTA-expressing macrophages have been described in prostate cancer48, particularly after chemotherapy; we do see this phenotype largely restricted to endocapillary LYZ+ VISTA+ monocytes, occasionally proliferating, which could be the seeding population from the bloodstream. Human macrophages found in tumors may be able to present antigens to T-cells41, where the co-inhibitory molecules act to protect the macrophage from cytotoxicity during the encounter. We failed to demonstrate a statistically significative close interaction between T-cell subsets and checkpoint molecules-bearing macrophages, except for the suggestive images provided by tissue snapshots (Fig. 4). This may has to do with the highly motile properties of T-cells, whereby during an asynchronous activation of multiple immune synapsis, the challenged T-cells may distance themselves from the interactor by the time the phenotype has changed to reflect the effect of the challenge49. In summary, we have described a variety of innate and adaptive immune cell phenotypes in ULMs, suggestive of antigen experience, and exploitable for a targeted immune intervention, despite the variability in phenotypes, particularly on the macrophage side.

Received: 27 October 2019; Accepted: 2 January 2020; Published on line 20 January 2020

Acknowledgements This work has been supported by Departmental University of Milano-Bicocca funds. MMB is employed by the Department of Medicine and Surgery of the University of Milano-Bicocca within a GlaxoSmithKline clinical research project BEL114054 (HGS1006-C1121), under which the Hamamatsu S60 digital scanner was obtained. AA was supported by the European Union Horizon 2020 grant SyMBioSys MSCA-ITN-2015-ETN #675585. Denis Schapiro (Harvard University, Boston, USA) graciously provided a pre-release copy of HistoCAT. We are grateful to Nicoletta Colombo, M.D. and Viviana Stufano, MSci (IEO) for clinical advice and contribution with the FISH analyses for PTEN, respectively.

Author contributions G.C., M.M. and M.M.B. equally designed the experiments. M.M.B., M.F. and M.M. devised the image analysis algorithms. F.M.B., R.M. and S.C. provided essential reagents and tissues. F.M.B., A.A., M.M.B. and M.M. performed visual and digital image analysis as well as bioinformatic evaluation. M.F. provided a customized version of the AMICO software. M.M., M.M.B. and G.C. wrote the manuscript. All authors have read and approved the final manuscript. Competing interests The authors declare no competing interests.

Additional information Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-57627-1. Correspondence and requests for materials should be addressed to G.C.

Table 1. Clinicopathologic and phenotypic data. NOTE: Abbreviations: FIGO: International Federation of Gynecology and Obstetrics staging system; TMN: standard for classifying the extent of spread of cancer; HPF: high power fields; het: heterogeneous; Total cells/sq mm: total number of cells in the core, based on DAPI staining; Tex: T cell exhausted. The predominant T cell phenotype was defined based on PD1 vs TCF7 expression: High PD1 – low/absent TCF7: acute exhaustion. Low PD1 – TCF7 present: chronic exhaustion. (See supplementary data).

Figure 1. The lymphocyte and endothelial phenotypes are shared among the sarcoma cases but each one has an individual macrophage population. (A) tSNE plot of all 21 cases. Each case is color-coded and marked by the case number. On the right are enlarged portions highlighted on the plot. Note admixture of the cases in the boxed areas and in cases 17, 18, 20 and 21. Case 15, containing very few cells, is not marked. (B) Phenograph groups are plotted on the tSNE plot shown in A. Note the lymphocytes and the endothelial phenogroups, corresponding to the areas of case admixture shown in A. Macrophage populations for each case is represented by one to three phenogroups. (C) tSNE plots are highlighted with

lymphoid (CD3; enlarged in the inset), endothelial (CD34; enlarged in the inset) and myelomonocytic markers.

Figure 2. Composite phenotype of the myelomonocytic and lymphoid infiltrate. (A) Absolute numbers of the inflammatory cells in each case per 6.28 mm2. Note the selective absence of CD16+TAMs in case 8, non neoplastic myometrium. Legend is shown in the bottom right of the graph. (B) Distribution of checkpoint protein and activation markers on myelomonocytic cells. Case 8, non neoplastic myometrium, has a small percentage of inflammatory cells with a coordinated activated phenotype; in all other cases, the expression of markers is uncoordinated. Legend is shown in the bottom right of the graph. (C) Distribution of relevant markers on lymphoid subsets. Note that only cases with enough lymphocytes are represented. CD39, CD69, PD1 and TIM3 are expressed as percentage of all CD3+ lymphocytes. FOXP3 percentages refer to the CD4+ subset. TCF7 refers to the CD8+ subset. Legend is shown at the bottom of the graph.

Figure 3. Relationship between PD1+ and TCF7+ CD8+ T cells subsets. The coexistence of PD1+ TCF7− and of TCF7+ PD1− CD8+ T cells in each case is plotted as percentage of all CD8+ cells. Note that some samples show skewed expression by either population, others have a mixture of both. For complete data see Supplemental Data.

Figure 4. T lymphocyte, macrophage and dendritic cell interactions in sarcomas. Inverted grayscale images of immunostained section detail from Case 14. PDL1+ macrophages expressing HLA-DR, CD16, CD64, CD68, CD163 and TIM3 are highlighted by a dashed line and are surrounded by PD1+ T cells, both CD8 and CD4. In the center, an HLA-DR+, S100AB+ CD14- dendritic cell is highlighted by a solid line. The empty arrowhead shows a CD8+, VISTA +, TCF7+ lymphocyte. The black arrowheads point to CD4+, FOXP3+, TCF7−, CD32±, OX40 ± regulatory T cells. The empty triangle points to a CD3+, CD8+, TCF7+, CD69+, VISTA+ lymphocyte. Note that TCF7 and FOXP3 are mutually exclusive. Negative markers are not represented. Image size 60 × 56 μm.

Figure 5. Heterogeneity of macrophages in sarcomas. Inverted grayscale images of immunostained section detail from Case 13. (A) CD68+, CD14−, CD16−, CD64− CD163 phagocytes are highlighted by a dashed line. The arrowhead points to an endovascular VISTA+, LYZ+, CD14+ inflammatory monocyte. Negative markers are not represented. Image size 105 × 145 μm. (B) A solid line highlights an activated CD83+ PDL1+ TAM (CD14+, CD16+, CD32+, CD64+, CD68+, CD163+, TIM3+). The arrows point to PD1+ Monocytes/ Macrophages undefined (CD14+, CD64±, CD68±, CD163±), one of which co-express PD1 and PDL1. The dashed outlines indicate PD1+, CD68+, TIM3± Phagocytes. Note a PD1+ isolated tumor cell (asterisk). Negative markers are not represented. Image size 76 × 67 μm.

Table 1

Case N.	Age	Menopause	FIGO	TMN	Size cm.	Mitosis/ 10HPF	Necrosis	MHC Class I	MHC Dynamics	Desmin	B7H3	Axl	IDO	Total cells/ sq mm	Segmented/ total DAPI	Predominant T cell phenotype
1	39	No	IVB	pT3aNxM1	6	20	Present	+++	induced	+	+ het	++	pos	900	31%	T_{ex}-chronic
2	63	Yes	IVB	pT1bNxM1	16	11	Present	+++	constitutive	+	+	±	pos weak	1248	23%	T_{ex}-acute
3	34	No	N/A	N/A	2	9	Absent	+	induced	+++	+ het	++	pos	517	15%	Deserted
4	36	No	IB	pT1bN0Mx	11	50	Present	±	constitutive	±	+++	++	pos	210	5%	Deserted
5	55	Yes	IIIB	pT3bNxMx	20	18	Absent	+	induced	neg	+++	++ het	pos	2567	77%	T_{ex}-acute
6	61	Yes	IB	pT1bNxMx	12	25	Focal	+	constitutive	+++	+ het	+ het	pos	474	8%	Deserted
7	44	No	IVB	pT1bNxM1	14	40	Present	+	constitutive	neg		±	pos	363	6%	T_{ex}-acute
8	40	No	IIIB	pT3bN0Mx	4	30	Present	±	constitutive	+++	+het	+	pos	481	8%	Deserted
9	47	No	IB	pT1bN0Mx	9	35	Focal	+++	constitutive	+	+ het	neg	pos weak	593	14%	Deserted
10	62	Yes	IB	pT1bN0Mx	8	25	Focal	+++	induced	+++	+/− het	+	pos weak	1780	38%	T_{ex}-acute
11	61	Yes	IIB	pT2bNxMx	13	25	Focal	+++	constitutive	neg	+	+	pos	1644	44%	Deserted
12	49	No	IB	pT1bNxMx	11	25	Focal	+ focal	induced	neg	+++	±	neg	2890	54%	Deserted
13	59	Yes	IB	pT1bNxMx	7	13	Focal	++ het	induced	neg	+++	++	pos weak	3533	93%	T_{ex}-acute
14	35	No	IB	pT1bNxMx	11	50	Absent	+ het	induced	++	+	++	pos	2756	33%	T_{ex}-mixed
15	57	Yes	IIIB	pT3bN0Mx	7	20	Present	+	constitutive	++	++ het	neg	neg	72	2%	Deserted
16	49	No	IIIC	pT3bN1Mx	14	25	Present	+++	induced	neg	−	+	neg	1631	33%	T_{ex}-chronic
17	60	Yes	IB	pT1bN0Mx	19	11	Focal	+	induced	neg	−	++	pos	984	50%	T_{ex}-mixed
18	71	Yes	IB	pT1bN0Mx	9	35	Focal	+	constitutive	+	−	++	pos	645	13%	T_{ex}-mixed
19	51	Yes	IIIC	pT3bN1Mx	19	60	Present	+++	constitutive	+	+/− het	++het	neg	1296	27%	T_{ex}-chronic
20	69	Yes	IB	pT1bNxMx	10	40	Present	±	induced	++	−	++	pos het	666	21%	Deserted
21	34	No	IB	pT1bN0Mx	3	10	Absent	++ het	constitutive	+++	−	++	pos	502	10%	Deserted

Fig.1

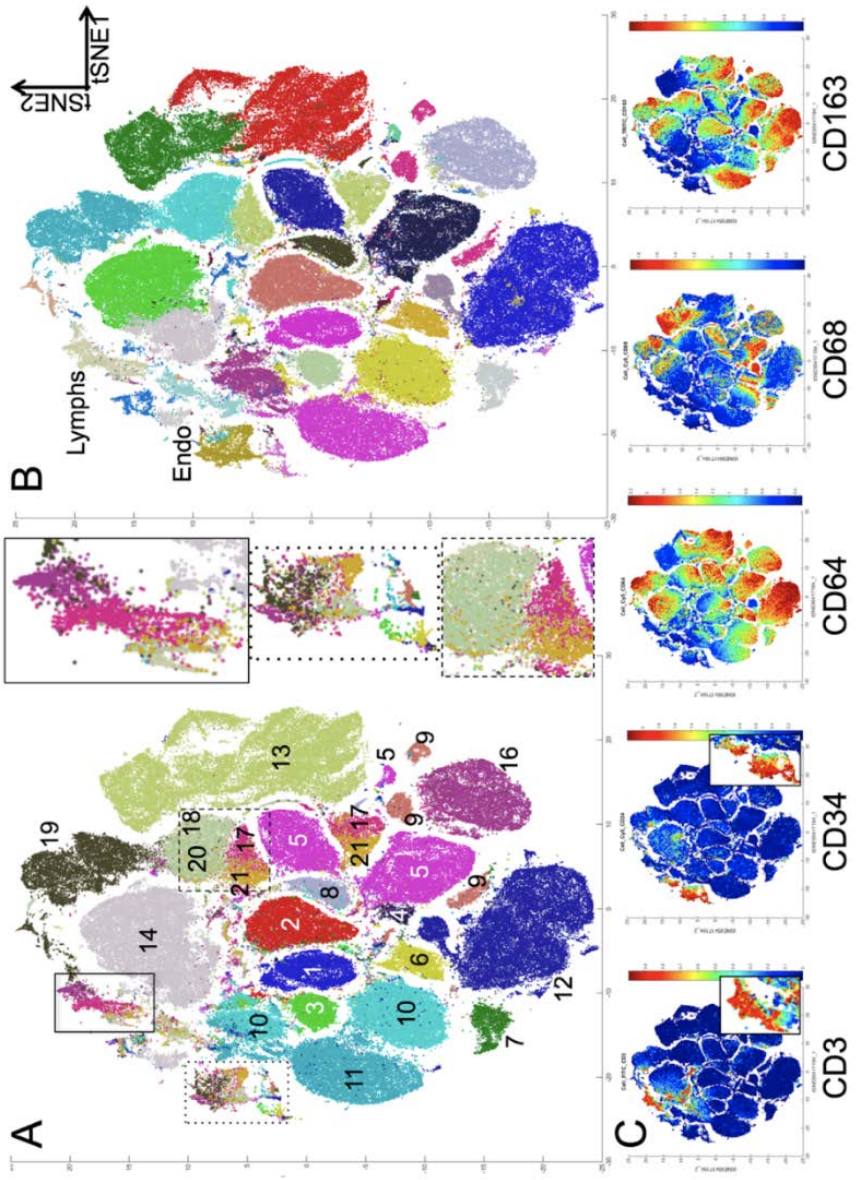


Fig.2

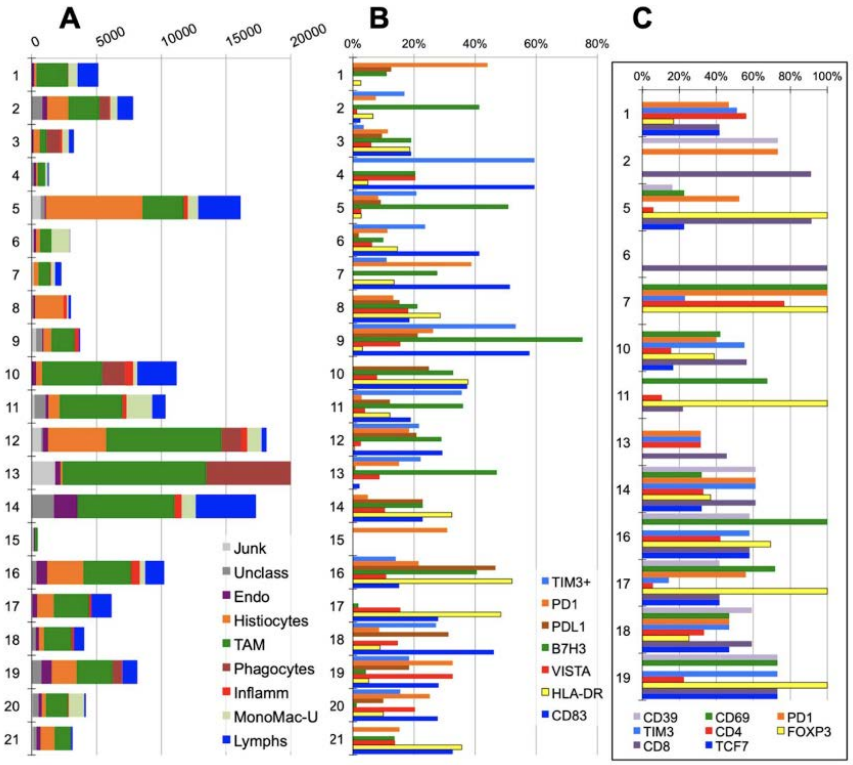

A
0
5000
10000
15000
20000
Junk
Unclass
Endo
Histiocytes
TAM
Phagocytes
Inflamm
MonoMac-U
Lymphs
B
0%
20%
40%
60%
80%
TIM3+
PD1
PDL1
B7H3
VISTA
HLA-DR
CD83
C
0%
20%
40%
60%
80%
100%
CD39
CD69
PD1
TIM3
CD4
FOXP3
CD8
TCF7

Fig.3

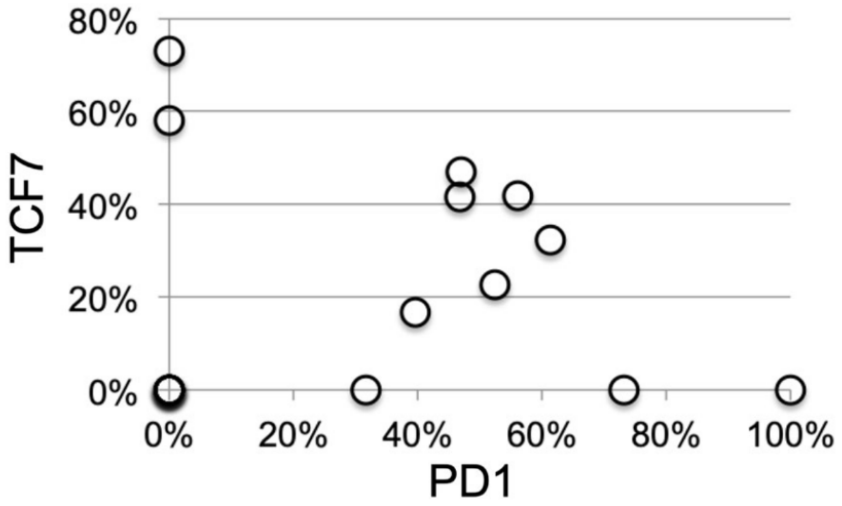

80%
60%
40%
20%
0%
TCF7
0%
20%
40%
60%
80%
100%
PD1

Fig.4

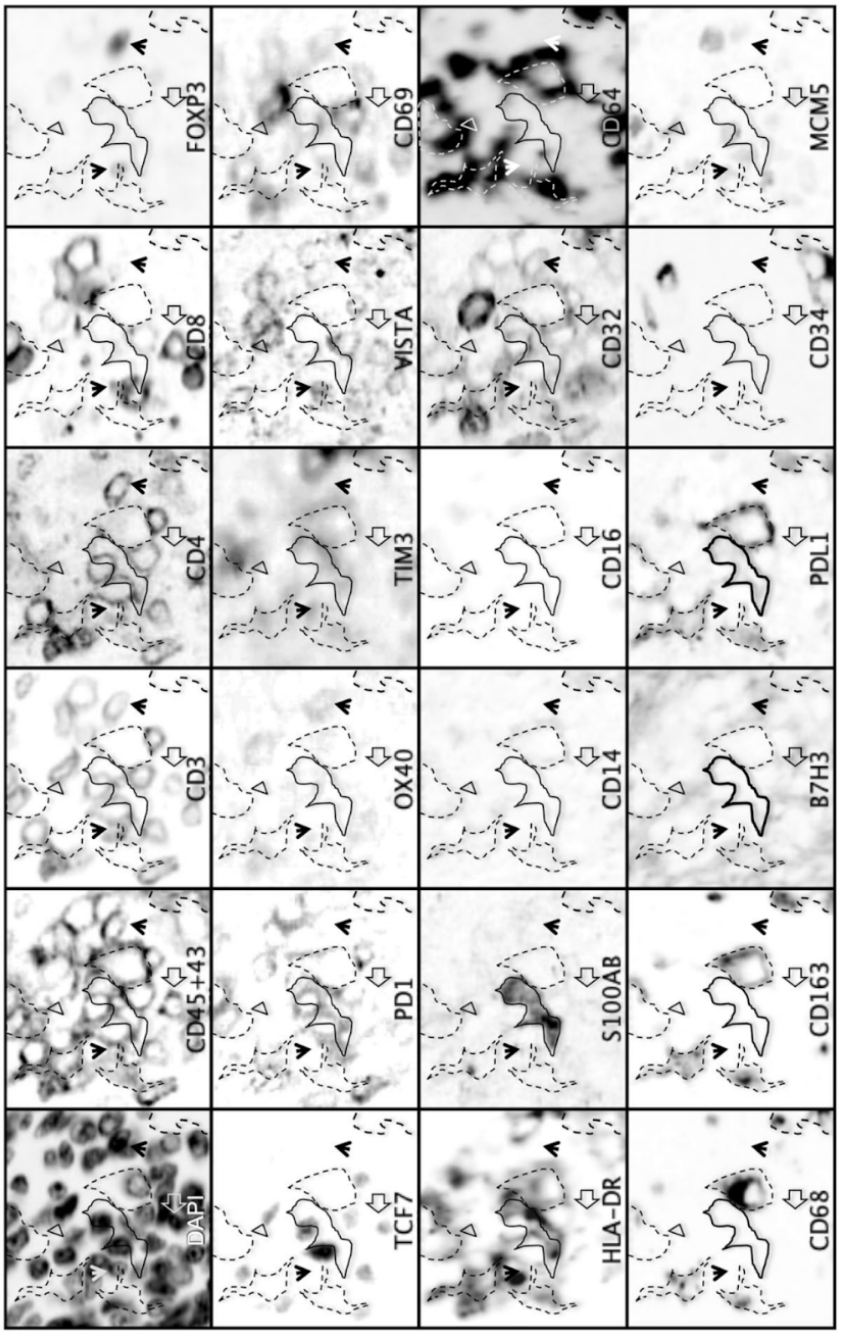
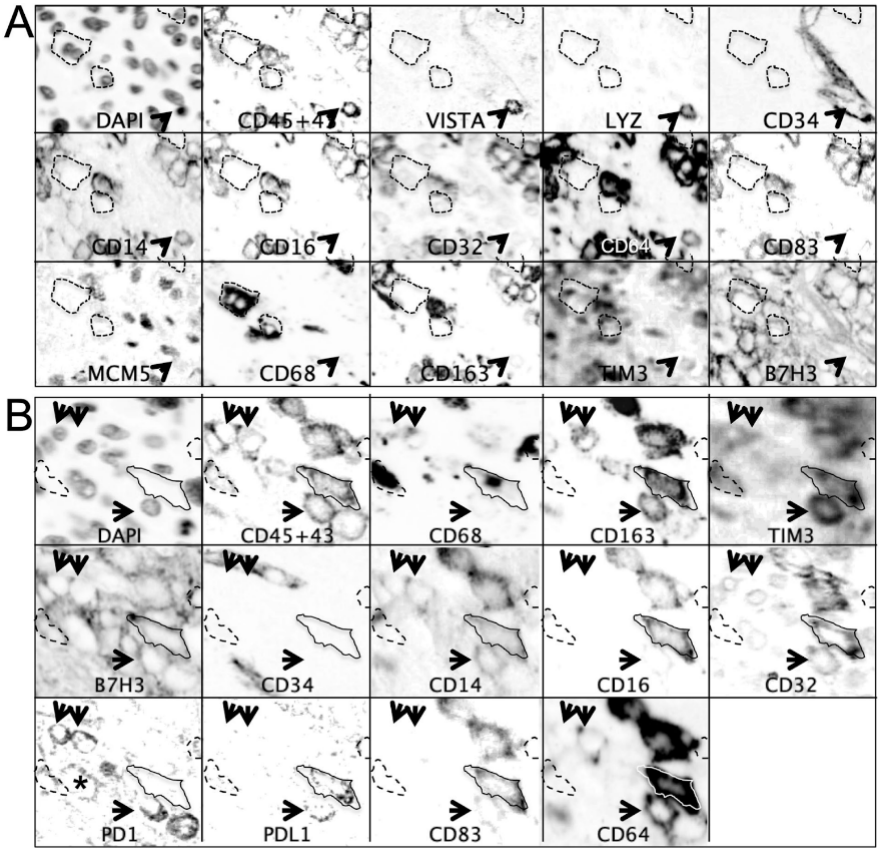

DAPI
CD45+43
CD3
CD4
CD8
FOXP3
TCF7
PD1
OX40
TIM3
VISTA
CD69
HLA-DR
S100AB
CD14
CD16
CD32
CD64
CD68
CD163
B7H3
PDL1
CD34
MCM5

Fig.5



References

1. Gregersen, P. & Behrens, T. Genetics of autoimmune diseases–disorders of immune homeostasis. Nat. Rev. Genet. 7, 917–928, https://doi.org/10.1038/nrg1944 (2006).

2. Hanahan, D. Rethinking the war on cancer. Lancet 383, 558–563, https://doi.org/10.1016/S0140-6736(13)62226-6 (2014).

3. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. Nature 541, 321–330, https://doi.org/10.1038/nature21349 (2017).
4. Palucka, A. K. & Coussens, L. M. The Basis of Oncoimmunology. Cell. 164, 1233–1247, https://doi.org/10.1016/j.cell.2016.01.049 (2016).
5. Wolchok, J. D. & Chan, T. A. Cancer: Antitumour immunity gets a boost. Nature 515, 496–498, https://doi.org/10.1038/515496a (2014).
6. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune Checkpoint Blockade: A Common Denominator Approach to Cancer Therapy. Cancer Cell. 27, 450–461, https://doi.org/10.1016/j.ccell.2015.03.001 (2015).
7. Blank, C. U., Haanen, J. B., Ribas, A. & Schumacher, T. N. CANCER IMMUNOLOGY. The “cancer immunogram”. Science 352, 658–660, https://doi.org/10.1126/science.aaf2834 (2016).
8. Joyce, J. A. & Fearon, D. T. T cell exclusion, immune privilege, and the tumor microenvironment. Science 348, 74–80, https://doi. org/10.1126/science.aaa6204 (2015).
9. Kowanetz, M. et al . Differential regulation of PD-L1 expression by immune and tumor cells in NSCLC and the response to treatment with atezolizumab (anti-PD-L1). Proc. Natl. Acad. Sci. USA 115, E10119–E10126, https://doi.org/10.1073/pnas.1802166115 (2018).
10.Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. Science 348, 69–74, https://doi.org/10.1126/science. aaa4971 (2015).

11.Ribas, A. Tumor immunotherapy directed at PD-1. N. Engl. J. Med. 366, 2517–2519, https://doi.org/10.1056/NEJMe1205943 (2012).
12.Rizvi, N. A. et al . Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348, 124–128, https://doi.org/10.1126/science.aaa1348 (2015).
13. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 357, 409–413, https://doi. org/10.1126/science.aan6733 (2017).
14. McGrail, D. J. et al. Multi-omics analysis reveals neoantigen-independent immune cell infiltration in copy-number driven cancers. Nat. Commun. 9, 1317, https://doi.org/10.1038/s41467-018-03730-x (2018).
15.Ock, C.-Y. et al. Genomic landscape associated with potential response to anti-CTLA-4 treatment in cancers. Nature. Communications 8, 1050, https://doi.org/10.1038/s41467-017-01018-0 (2017).
16. Thorsson, V. et al. The Immune Landscape of Cancer. Immunity 48, 812–830 e814, https://doi.org/10.1016/j.immuni.2018.03.023 (2018).
17.Varn, F. S., Wang, Y., Mullins, D. W., Fiering, S. & Cheng, C. Systematic Pan-Cancer Analysis Reveals Immune Cell Interactions in the Tumor Microenvironment. Cancer Res. 77, 1271–1282, https://doi.org/10.1158/0008-5472.CAN-16-2490 (2017).

18.Doig, T. N. et al . Coexpression analysis of large cancer datasets provides insight into the cellular phenotypes of the tumour microenvironment. BMC Genomics 14, 469, https://doi.org/10.1186/1471-2164-14-469 (2013).

19.Lavin, Y., Mortha, A., Rahman, A. & Merad, M. Regulation of macrophage development and function in peripheral tissues. Nat. Rev. Immunology 15, 731–744, https://doi.org/10.1038/nri3920 (2015).

20.Cancer Genome Atlas Research Network. Electronic address, e. d. s. c. & Cancer Genome Atlas Research, N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell., 171, 950–965 e928, https://doi.org/10.1016/j. cell.2017.10.014 (2017).

21.Ben-Ami, E. et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. Cancer 123, 3285–3290, https://doi.org/10.1002/cncr.30738 (2017).

22. George, S. et al. Loss of PTEN Is Associated with Resistance to Anti-PD-1 Checkpoint Blockade Therapy in Metastatic Uterine Leiomyosarcoma. Immunity 46, 197–204, https://doi.org/10.1016/j.immuni.2017.02.001 (2017).

23. Toulmonde, M. et al. Use of PD-1 Targeting, Macrophage Infiltration, and IDO Pathway Activation in Sarcomas: A Phase 2 Clinical Trial. JAMA Oncol. 4, 93–97, https://doi.org/10.1001/jamaoncol.2017.1617 (2018).

24. Cattoretti, G., Bosisio, F. M., Marcelis, L. & Bolognesi, M. M. Multiple Iteractive Labeling by Antibody Neodeposition (MILAN)

Protocol Exchange, https://doi.org/10.1038/protex.2018.106 (2018).
25.Bolognesi, M. M. et al . Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. J. Histochemistry & Cytochemistry 65, 431–444, https://doi.org/10.1369/0022155417719419 (2017).
26.Carpenter, A. E. et al . CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 7, R100, https://doi.org/10.1186/gb-2006-7-10-r100 (2006).
27. Schapiro, D. et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat. methods 14, 873–876, https://doi.org/10.1038/nmeth.4391 (2017). 28. Seeber, A. et al. High IDO-1 expression in tumor endothelial cells is associated with response to immunotherapy in metastatic renal cell carcinoma. Cancer Sci. 109, 1583–1591, https://doi.org/10.1111/cas.13560 (2018).
29. Egelston, C. A. et al. Human breast tumor-infiltrating CD8+ T cells retain polyfunctionality despite PD-1 expression. Nat. Commun. 9, 4297, https://doi.org/10.1038/s41467-018-06653-9 (2018).
30.Azizi, E. et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 174, 1293–1308.e1236, https://doi.org/10.1016/j.cell.2018.05.060 (2018).
31. Duhen, T. et al. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors.

Nat. Commun. 9, 2724, https://doi.org/10.1038/s41467-018-05072-0 (2018).
32. Simoni, Y. et al. Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. Nat. Publ. Group. 557, 575–579, https://doi.org/10.1038/s41586-018-0130-2 (2018). 33. Sade-Feldman, M. et al. Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. Cell. 175, 998–1013.e1020, https://doi.org/10.1016/j.cell.2018.10.038 (2018).
34.Thommen, D. S. & Schumacher, T. N. T Cell Dysfunction in Cancer. Cancer Cell . 33, 547–562, https://doi.org/10.1016/j.ccell.2018.03.012 (2018).
35. Saito, T. et al. Two FOXP3(+)CD4(+) T cell subpopulations distinctly control the prognosis of colorectal cancers. Nat. Med. 22, 679–684, https://doi.org/10.1038/nm.4086 (2016).
36. Jiménez-Sánchez, A. et al. Heterogeneous Tumor-Immune Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. Cell. 170, 927–938.e920, https://doi.org/10.1016/j.cell.2017.07.025 (2017).
37.Scheper, W. et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. Nat. Med. 25, 89–94, https://doi.org/10.1038/s41591-018-0266-5 (2019).
38.Franklin, R. A. et al . The cellular and molecular origin of tumor-associated macrophages. Science 344, 921–925, https://doi. org/10.1126/science.1252510 (2014).

39. Biswas, S. K., Allavena, P. & Mantovani, A. Tumor-associated macrophages: functional diversity, clinical significance, and open questions. Semin. immunopathology 35, 585–600, https://doi.org/10.1007/s00281-013-0367-7 (2013).
40.Lawrence, T. & Natoli, G. Transcriptional regulation of macrophage polarization: enabling diversity with identity. Nat. Rev. Immunology 11, 750–761, https://doi.org/10.1038/nri3088 (2011).
41.Singhal, S. et al. Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer. Sci. Transl. Med. 11, eaat1500, https://doi.org/10.1126/scitranslmed.aat1500 (2019).
42. Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. Nat. Rev. Immunology 8, 958–969, https://doi. org/10.1038/nri2448 (2008).
43. Xue, J. et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity 40, 274–288, https://doi.org/10.1016/j.immuni.2014.01.006 (2014).
44.Sconocchia, G. et al. Tumor infiltration by FcγRIII (CD16)+ myeloid cells is associated with improved survival in patients with colorectal carcinoma. Int. J. cancer J. Int. du. cancer 128, 2663–2672, https://doi.org/10.1002/ijc.25609 (2011).
45.Gordon, S. R. et al . PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. Nature 545, 495–499, https://doi.org/10.1038/nature22396 (2017).

46.Thibult, M.-L. et al. PD-1 is a novel regulator of human B-cell activation. Int. immunology 25, 129–137, https://doi.org/10.1093/ intimm/dxs098 (2013).
47. Hartley, G. P., Chow, L., Ammons, D. T., Wheat, W. H. & Dow, S. W. Programmed Cell Death Ligand 1 (PD-L1) Signaling Regulates Macrophage Proliferation and Activation. Cancer immunology Res. 6, 1260–1273, https://doi.org/10.1158/2326-6066.CIR-17-0537 (2018).
48. Gao, J. et al. VISTA is an inhibitory immune checkpoint that is increased after ipilimumab therapy in patients with prostate cancer. Nat. Med. 23, 551–555, https://doi.org/10.1038/nm.4308 (2017).
49.Miller, M. J., Hejazi, A. S., Wei, S. H., Cahalan, M. D. & Parker, I. T cell repertoire scanning is promoted by dynamic dendritic cell behavior and random T cell motility in the lymph node. Proc. Natl. Acad. Sci. USA 101, 998–1003, https://doi.org/10.1073/ pnas.0306407101 (2004).

Chapter 3

The normal and fibrotic mouse lung classified by spatial proteomic analysis

Roberta Ciccimarra 1,4, Maddalena M. Bolognesi 2,4, Matteo Zoboli 1, Giorgio Cattoretti 2, Franco F. Stellari 3* & Francesca Ravanetti 1

1 Department of Veterinary Science, Università di Parma, Parma, Italy. 2Department of Medicine and Surgery, Università di Milano-Bicocca, Monza, Italy. 3 Corporate Preclinical R&D, Chiesi Farmaceutici S.P.A., Largo Belloli 11/A, 43122 Parma, Italy. 4 These authors contributed equally: Roberta Ciccimarra and Maddalena M. Bolognesi. *email: FB.stellari@chiesi.com

Scientific Reports(2022) 12:8742
https://doi.org/10.1038/s41598-022-12738-9

Single cell classification is elucidating homeostasis and pathology in tissues and whole organs. We applied in situ spatial proteomics by multiplex antibody staining to routinely processed mouse lung, healthy and during a fibrosis model. With a limited validated antibody panel (24) we classify the normal constituents (alveolar type I and II, bronchial epithelia, endothelial, muscular, stromal and hematopoietic cells) and by quantitative measurements, we show the progress of lung fibrosis over a 4 weeks course, the changing landscape and the cell-specific quantitative variation of a multidrug transporter. An

early decline in AT2 alveolar cells and a progressive increase in stromal cells seems at the core of the fibrotic process.

Single cell biology, brought to fruition by advances in gene sequencing and computational progress, has revolutionized how we understand biological processes in health and in pathology1. Applying these techniques to the analysis of individual cells in-situ, i.e. within the tissue microenvironment, has added the information of the tissue sociology of the specimen, answering to the growing need to investigate it, due to various cellular functions, the spatial organization of molecular targets, the relationship among multiple cell types and morphology. While in situ single cell sequencing can provide information on the RNA species in the thousands, in situ multiplex staining has a limit of around 100 antibodies, in general 30–60 for the time being2. To distinguish the techniques which can add antibodies in the dozens from the other which provide a much limited number (5–6), we will refer to the former as “high-plex” multiplexing techniques.

Single cell biology via high-plex in situ staining can be accomplished both in fresh frozen tissue sections and in sections from routinely processed tissue (formalin fixed, paraffin embedded; FFPE), depending on the technology used. FFPE material has several advantages, including the capability to create high density arrays containing several different samples, the tissue microarray technology (TMA)3. Being the high-plex technology complex and time-consuming by itself, the

combination with the TMA technology allows an enormous, detailed analytical power in the basic and applied science field. Murine models are an extremely helpful, standard companion to the research on human subjects in all fields of medicine. All kinds of diagnostic and therapeutic approaches used on humans have been miniaturized and applied to mice4, including histopathologic and phenotypic examination. The problem of using antibody-based high-plex methods on mouse tissue when using primary antibodies raised in mice is the cross reactivity with endogenous immunoglobulins: we have overcome this problem by using anti-isotype secondary Abs5 on FFPE mouse tissue, therefore we have broadened the antibody portfolio which can be used on mouse histopathology. Mouse specimens have been extensively used for single cell biology studies in all fields, largely via single cell RNA sequencing, rarely for in situ spatial proteomics and high-plex staining. We have applied the MILAN high-plex technique6 to a murine model of Bleomycin-induced (BLM) lung fibrosis. Although, this animal model is the most used in preclinical study and the best characterized either to investigate lung fibrotic mechanisms or to screen drugs, and the American Thoracic Society (ATS) has suggested recommendations for preclinical assessment of antifibrotic compounds, a deep molecular profiling characterization is still lacking 7–9. Bleomycin elicits a time-dependent lung fibrotic process10, which has several similarities with the human counterpart11. The BLM delivered through the airways determines bronchiolocentric accentuated

fibrotic changes through a multistep injury starting with epithelial cell damage (in mice, days 1–3) followed by acute interstitial and intra-alveolar inflammation (days 3–10) and ultimately to fibroblasts activation and remodeling of extracellular matrix leading to fibrosis, extracellular matrix deposition (days 10–21 with a peak around day 14)12. In mice, the alterations induced by Bleomycin are heterogeneous, time-limited and self-resolving, with the drawback of a narrow time window for preclinical testing. In fact, histological analysis revealed fibrosis pattern started from day 7, mainly as single fibrotic masses, and evolved at days 14 as confluent conglomerates of substitutive collagen, which last till day 21 with a tendency to resolve already at 28 day13. To better understand lung fibrosis development, highplex technology has been used to detect target of interest, applying it to routinely processed tissue samples. In this study we consider the recent observation that BLM administration induces an ABC (ATP Binding Cassette) carrier upregulation in lung, with an increasing expression of P-glycoprotein (P-gp) and transporters in C57BL/6 male mice (Park et al. 2020)14. The ABC carriers actively transport multiple xenobiotics across the membrane reducing the intracellular concentration of drugs and leading to a potential decrease in anti-fibrotic activity. We present here the validation of the method and results of the single cell composition of the healthy and pathologic mouse lung, including the changes in expression of a P-gP multidrug transporter in each of the cell types identified by high-dimensional analysis.

Results

TMA validation.

The use of tissue cores, instead of whole sections, entails higher throughput at the expense of representativeness. To address this latter aspect we used single cell analysis to validate the use of the TMA. The image segmentation strategy, centered on the identification of nuclear DAPI + containing regions of interest (ROI), approximate single-cell identification in tissue, with some limits (see supplementary methods and discussions). 23 markers + DAPI (Table 1) were used to classify all the ROI (see M&M and supplementary material). Single cell analysis included an average of 2396 cells for 1 mm core, 10,337 for 2 mm core and 40,904 for the whole section (Fig. 1). Clusters comprising sparse cells, such as Macrophages (Fig. 1) and T cells (not shown) were equally represented in the 2 mm cores and in the whole section (Fig. 1), but ill-identified on 1 mm cores. The 2 mm TMA cores were used for the study.

Single cell analysis of the normal and fibrotic mouse lung

tSNE plots from four control animals showed a superimposition of all cells from all animals as expected (Fig. 2A). Cells from Bleomycin-treated mice were allocated in the same phenoclusters (Fig. 2C) of the controls, but with a slightly different spatial distribution for mice 7, 14 and 21 post BLM treatment (Fig. 2C). Mice at day 28th were very similar to controls (Fig. 2D). The phenoclusters contained cells from every

case (Fig. 2B) demonstrating the absence of batch effect in the data. Phenograph clusters containing all mice from all experimental points were plotted on tSNE and a provisional cell classification was assigned to each phenogroup, according to the defining proteins, as shown in a heatmap (Supplementary Fig. S5). Subsequently, each mouse was analyzed individually (Fig. 3 and Supplementary Fig. S8). After the removal of clusters resulting from artifacts or uninterpretable clusters (see supplementary material), a total of 80,506 cells were analyzed in the 13 cores considered, with an average distribution of 6029 cells/core. The normal mouse lung contained a majority of Alveolar Epithelial cells (67.9% ± 11.9), divided into AT1 (22% ± 5.8), AT2 (14%± 9.7) and a population with co-expression of AT1 and AT2 markers, named transitional AT (32% ± 11.8). Each of the remaining constituents of the normal lung remained below the 10% value. Bronchial epithelial cells and goblet cells averaged 7% ± 7.7, endothelial cells were 3%± 1.4, myofibroblasts/smooth muscle cells 2% ± 1.9, other stromal cells 0.7%± 0.4. Resident hematopoietic cells were T lymphocytes (7%± 1.5), B lymphocytes (3% ± 1.9), macrophages and dendritic cells (4%± 1.9) and neutrophils (1%± 1.2) (Fig. 3 and Supplementary Fig. S8). By plotting the cell types identified on the tissue coordinates, cell types composing the scaffold of the normal lung were organized in morphologically recognizable structures, while lung epithelial cells and other diffusely present cell types were evenly distributed across the tissue (Fig. 4 and supplementary Fig. S7),

mimicking the image obtained by multicolor immunofluorescence (Fig. 4 and supplementary Fig. S4). Single cells from mice treated with Bleomycin and examined at day 7, 14, 21 and 28 partially overlap with the control mice (Fig. 2), highlighting qualitative and quantitative changes in the lung population. By applying the same classification criteria used to classify the normal lung, one could observe a transient reduction in AT1 and a progressive decrease of AT2 pneumocytes (Fig. 3), largely due to changes in an AT2 subpopulation (transitional AT2, see supplementary Fig. S7 and S8). Stromal cells increased progressively, accompanied by a late increase of myofibroblasts/smooth muscle cells, mimicking the histopathologic accumulation of stroma (see supplementary Fig. S1). The remaining lung cells remained stable except for a late decrease of macrophages, neutrophils and B lymphocytes. Plasma cell markers were not included in the panel, thus we could not assess whether the decrease in B cells was due to absence or maturation. The spatial distribution of the cell types revealed a crowded parenchyma and focally increased stromal foci, B cell and macrophage aggregations (Fig. 4). A drug transmembrane transporter, P-gP, constitutively expressed in many cell types, was measured in lung cells identified by high dimensional analysis. Increasing intensity signal was registered in all populations after Bleomycin treatment, except in bronchial cells (Fig. 5 and supplementary Fig. S9). The onset of the time-dependent increase was delayed in the stromal cell type (Fig. 5)

Discussion

We have obtained an in situ spatial representation of the normal and of the fibrotic mouse lung by single cell classification via multiple antibody staining. Analogous to a single cell classification of murine (and human) lung cell component by single cell RNA sequencing 15–18, here we show that the content of a normal or diseased organ such as the lung can be finely dissected at the single cell level, with two additional properties: spatial cell disposition is represented and proteins, instead of RNA, are assessed by a robust, cheap and versatile method. With a rather limited number of validated antibodies, we can identify in situ the main cell types which are relevant for lung homeostasis and for the initiation and establishment of the Bleomycin-induced fibrosis: AT1, AT2, transitional AT1-AT2 pneumocytes, bronchial lining cells, vasculature, stromal cells, T and B lymphocytes, macrophages. In addition, we measured at the single cell level continuously expressed proteins such as a transcellular drug transporter, P-gP. We measured with the same detail the changes occurring upon Bleomycin treatment over time, documenting a change in the AT1/AT2 ratio, the progressive accumulation of stromal cells and the asynchronous increase of P-gP expression.

Lastly, we included in the assay a high throughput method, the tissue microarray technology, after careful validation of the representativeness of tissue core size. We suggest the use of a 2 mm TMA core as a minimum tissue area for the evaluation of

tissue cell composition on mouse lung. We came to this conclusion by performing high-dimensional analysis as a validating tool, applicable to TMAs from tissues of various origin and composition and by taking into account the highest total cell number compatible with TMA sampling of the lung. The most abundant population we found in normal tissue and also in BLM-treated mice is represented by epithelial cells. Data which has been obtained by tissue dissociation and single cell sequencing of normal mouse lungs19,20 do not agree among themselves about the representation of the cell types, probably because of different pre-analytical dissociation methods and inherent selective loss of epithelial cells or enrichment of other cells. The drawbacks of the tissue dissociation methods are known21. By in-situ cell classification we can provide an unbiased estimate of the various lung cell components. In addition, by measuring the end product of the transcription and translation machinery, we complement the data provided by RNA sequencing.

Analogously to scRNA sequencing data, the alveolar epithelial cell populations do aggregate in closed contact in the bi-dimensional tSNE space, reflecting a continuum of phenotypes, interpretable as a transition from AT2 to AT1 cells. We tentatively identified a reproducible subset of AT2 cells bearing AT1 markers, which we dubbed “transitional AT” cells, analogously to a similar subset identified in Bleomycin-treated animals 15,17. Whether the subset we identified by in-situ proteomics and the ones described in scRNAseq experiments

are identical requires additional experiments with parallel analytical tools. We cannot completely exclude that two distinct AT1 and AT2 cells, closely spaced, are segmented as a cell with hybrid phenotype. Collagen deposing stromal cells are the main actor in the fibrotic process produced in this experimental model. We could demonstrate an increase in stromal cells which parallels the histomorphologic changes. According to our data, these cells represent a minority of the lung population, to the point that are inconsistently demonstrated in the normal lung, at variance with data obtained by dissociation and RNA sequencing. Besides the differences in methods, as outlined before, stromal cells are underrepresented in our markers panel; in addition,a nuclear-based cell segmentation is not ideal to identify elongated cells, thus we may underestimate this cell subset, despite showing a treatment-dependent increase. Histochemical stains do not discriminate the cell of origin for collagen deposition, nor if the collagen is deposited by pre-existing cells. On the other hand, we assess stromal cells individually and by intracellular markers, documenting a net increase in nucleated stromal cells upon treatment: it is thus not unexpected that the two represent non-identical assays, as demonstrated by the data. The progressive decrease of the AT2 and particular of the transitional AT over time and the increase of stromal cells we have shown may highlight the key drivers of lung fibrosis in this mouse model: a progressive reduction of the alveolar lining repair by depleting local alveolar progenitors and surfactant producing cells, AT2, coupled with newly produced,

collagen depositing fibroblasts. These results are novel and should be consolidated by independent experiments and increased sample numerosity. The time-course expression of P-gP has been previously published for the same Bleomycin-induced mouse lung model14, with notable differences: the experiments were performed on male mice and the RNAs for the multidrug transporters were measured, both as whole tissue extracts and by in-situ hybridization. We have reproduced the data on male mice (see Supplementary data), and shown that female mice differ in the kinetic of the pathology but not in the type of histopathologic lung changes. In addition, by measuring the multidrug transporter protein at the single cell level on each and every cell type in the specimen, we have detailed a cell-type specific kinetics of P-gP expression, with some notable difference with the RNA data of Park et al.14. Limitations of the study: We employed a limited number of antibodies, compared to the potentiality of the MILAN technique (over 100); this because there is a variety of reagents to choose from which is more limited than what is available for human FFPE tissue, particularly in terms of species origins of the antibodies. Being the MILAN technique based on multiple (3 or 4) unconjugated antibodies per round, we partially overcome these limitations by using mouse antibodies on mouse tissue, background free5. It has been shown previously however, that the discriminating power of the dimensionality reduction algorithms is so high, that even with a reduced number of diagnostic parameters, the cell types of interests can still be

identified22 and this is the case for the present work. An additional limitation of this study is the use of a rather unsophisticated image segmentation technique and manual cell type assignment for the phenogroups. Cell segmentation is one of the most challenging tasks for in situ transcriptomics and proteomics23 and efforts are ongoing to improve it for the mouse lung fibrosis model. In this model, elongated cells such as fibroblasts and stromal cells in general poorly fit cell identification via nuclear DAPI identification. This may be the reason why in normal lungs we occasionally found minimal or no fibroblast cell clusters, together with the small total amount, in common with cell suspension based studies 15–18. Along the same lines, we found clusters containing markers of two distinct cells (e.g. epithelial and macrophages, epithelial and stroma) which we were forced to discard, because of the juxtaposition of elongated cells was not solvable with the cell segmentation strategy used, despite being clearly identifiable by the tissue spatial distribution (see supplementary data/discussion). Lastly, despite the nucleus-centered segmentation strategy, which limits the sampling of nearby cells, signals of diffusible and/or extracellular proteins may leak into another “pseudo-cell”, providing a hybrid signature, but also important biologic information about closely adjacent cells. In summary, we have shown proof of principle that mouse tissue, being a normal organ, a pathology model or a developing tissue, can be dissected on routinely processed material by in-situ high-dimensional proteomics and single cell bioinformatic

analysis. This represents a powerful tool for pre-clinical studies such as drug discovery and novel treatments and can be integrated with other “omics” tools such as scRNA sequencing, in situ transcriptomics etc.24.

Materials and methods

Experimental animals and histology.

A mouse model of Bleomycin-induced lung fibrosis, currently in use in our institution, has been previously published25 and has been approved by the internal AWB (Animal Welfare Body) of Chiesi Farmaceutici under protocol number: 841/2019-PR and comply with the European Directive 2010/63 UE, Italian D.Lgs 26/2014 and the revised “Guide for the Care and Use of Laboratory Animals” and the Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines (www.arriveguidelines.org). In brief it consists of the oropharyngeal BLM administration to 7–8 weeks old C57Bl/6 female mice of 15 μg/mouse at each day of treatment respectively. Animals were sacrificed at 7, 14, 21 and 28 days after the administration. After the sacrifice, lungs were removed, formalin fixed and paraffin embedded (FFPE). Three serial sections, 5 μm thick were obtained for the stainings: Hematoxylin and Eosin, Masson’s trichrome (TM) and a multiplex immunostaining. Whole-slide images were acquired by a NanoZoomer S-60 Digital slide scanner (NanoZoomer S60, Hamamatsu, Japan) at 20× magnification. Fibrotic lung

injury was assessed histologically through the Ashcroft scoring system 26,27 on whole parenchyma by two independent researchers (blinded to the experimental design). Detailed description and histomorphometric characterization are included in the Supplementary methods and in Supplementary Figure S1. The most relevant areas on the slide were marked for subsequent tissue microarray construction (see below).

Tissue microarray design and construction.

For the TMA preparation, we used FFPE tissue blocks of lungs from BLM-treated female mice. 13 cases of treated and control lungs were selected; cores were extracted, selecting specific regions representing normal lung tissue with alveolar parenchyma, bronchioles and vessel for saline samples and fibroproliferative foci with its alterations for Bleomycin ones from the C57BL/6 female miceFFPE lungs of BLM and saline treated mice across four time points (7, 14, 21 and 28 days, see Supplementary Table S1). The TMA was constructed with a Tissue Microarrayer Galileo CK4500 (Tissue Microarrayer Model TMA Galileo CK4500; Integrated Systems Engineering srl, Milano, Italy) using Galileo Software to match the annotated tissue on histological slide with their corresponding areas on the surface of the paraffin donor blocks. After the core transfer in the recipient block, to allow the samples to be properly embedded into the block, the TMA was incubated for 24 h at + 38 °C. Finally, 5 µm thick serial sections were cut from the TMA with a rotary microtome (Slee Cut 6062, Slee Medical, Mainz,

Germany) and placed on Polysine adhesion glass slides (Thermo Fisher Scientific). Routine histology was performed as described previously25. In order to select the appropriate Tissue Microarray (TMA) core dimension, after applying a high-plex MILAN staining method6 on the whole slide of a mouse lung, we selected 4 virtual cores (with ImageJ) of 1 mm and 4 of 2 mm in diameter. Thus we run single cell analysis as previously described on 8 cores and on a whole slide and we compare the results. The level of adequacy of tissue portion was set when all main cell populations found in whole tissue by clustering analysis (Rphenograph package) were rediscovered in cores. Moreover, the 2 mm core better includes the patchy distribution of fibroproliferative foci within the parenchyma.

Highplex immunofluorescence

Lung samples were processed as per the Milan (Multiple Iterative Labeling by Antibody Neodeposition) protocol6 modified for mouse FFPE tissue staining. The protocol consists in the cyclic application of primary antibodies (Table 1) raised in multiple species, including mouse5, applied on the same section, nuclear staining with DAPI, autofluorescence subtraction28. For each round, the stained slides were scanned and saved as digital images using a multichannel fluorescence acquisition instrument (NanoZoomer S60, Hamamatsu, Japan). The antibody stripping preceded a subsequent staining round. A multichannel fluorescence acquisition was performed after each stripping, in order to check the complete antibody

removal. In order to generate an antibody panel which would produce single cell lung classification, we mined existing single cell RNA sequencing (scRNAseq) studies19,20 for highly expressed, lineage-restricted messages which would correspond to proteins against which antibodies would be available. The staining sequence design and antibodies validation and specifications were provided in Table 1. The methods used for the antibody selection and validation are reported in Supplementary methods. Sections were incubated overnight with primary antibodies, which were then revealed by secondary fluorochrome-tagged antibodies; both isotype control and omission of primary have been performed as negative control. Sequential stripping was obtained with a beta-mercaptoethanol and sodium dodecyl sulphate mix28. Mouse tissue preservation after each of multiple staining and stripping rounds showed no or negligible cell loss (Supplementary Figs. S2 and S3), as measured by DAPI nuclear stain and quantification on four other tissues (kidney, liver, lung and heart).

Single cell analysis

After the stainings were acquired, digital slide images (.ndpi) were imported as .tiff and registered with the AMICO software29, based on Fiji. All DAPI images from different staining rounds were registered together and their coordinates of rotation and translation were used to align the individual marker images of each round. Once all images were aligned,

autofluorescence was subtracted from FITC and TRITC channels. DAPI stainings were used for cell segmentation with Fiji (Threshold Default -Watershed_Analize particles and Count mask creation), where each nucleus and surrounding cytoplasm is identified by a single Region of Interest (ROI). Cell segmentation based on DAPI or other markers is a bioinformatic approximation to physical single cell isolation, while maintaining in situ cell position and microenvironment location. The term “single cell” will be used throughout for this approximation. Mean intensity value of all markers within each segmented cell and spatial coordinates of centroids of nuclei were recorded together in a .csv file. Then .csv files were uploaded to the R Studio software (version 1.4) for a more detailed analysis. Rtsne 30, umap 31 and Rphenograph 32 (n = 30) algorithms (R packages) were used respectively for dimensionality reduction and clustering of data. UMAP plots were decorated with single individual relevant marker intensity to evaluate the distribution among phenogroups. A single tSNE plot from homogeneous groups of animals was used to compare treatments versus control (Fig. 2), exclude an experimental batch effect (Fig. 2) and explore the cell type of the clusters identified by Phenograph with a comprehensive heatmap (Supplementary Fig. S5). Subsequently, the samples were analyzed individually. Clusters obtained were further explored via: (i) individual cell immunoprofiling through dedicated hierarchical clustering heatmaps as published 33 (Supplementary Fig. S6), (ii) visualization of the spatial

distribution on tissue (Figs. 4 and S7). Clusters which satisfy both criteria were manually classified according to the expressions of key markers (Table 1). Major cell types were created by merging clusters with a similar phenotype, after removing artifacts. Main types were organized as follows: (1) Alveolar Epithelial cells subdivided in AT1, AT2 or transitional AT1-AT2 subpopulations, (2) Bronchial and Goblet Epithelial cells, (3) Macrophage cells, (4) B cell, (5) T cells, (6) Vasculature and (7) Stromal cells, Neutrophils (8). For graphic spatial representation, all epithelial cells were grouped in a single entity (Alveolar Epithelial Cells). Phenoclusters for which no coherent phenotype was identifiable were excluded from the analysis (see supplementary material for further information). The transporter expression was analyzed by measuring the mean of the intensity of the signal (on 8-bit grayscale images, values 0–255) within each main type.

Ethics approval and consent to participate. All the animal experiments were performed according to the ARRIVE guidelines (www.arriveguidelines.org) and European Directive 2010/63 UE, Italian D.Lgs 26/2014, and associated guidelines.

Acknowledgements We wish to thank Mario Faretta, PhD, Istituto Europeo di Oncologia, Milan, Italy, for providing and updating the A.M.I.C.O. software, all the lab technicians for the excellent skills and support.

Author contributions

R.C., F.R., G.C. and F.F.S. equally designed the experiments; M.M.B., R.C. and M.Z. devised the image analysis algorithms and performed visual and digital image analysis; R.C. and F.R. performed immunostaining experiments; M.M.B., G.C., F.R. and F.F.S. wrote the manuscript.

Funding This work was supported by internal funds from Chiesi Farmaceutici S.p.A., by the Departmental University of Milano-Bicocca funds, by Regione Lombardia POR FESR 2014–2020, Call HUB Ricerca ed Innovazione: ImmunHUB to Giorgio Cattoretti, by the Departmental University of Parma funds.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information

The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-022-12738-9. Correspondence and requests for materials should be addressed to F.F.S. Reprints and permissions information is available at www.nature.com/reprints. Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Figure 1 Core dimension validation for TMA construction. Representative DAPI-stained normal lung whole slide image (**A**) in which 1 (blue) and 2 (red) mm diameter representative cores are drawn. (**B**–**D**) tSNE plots derived from the 1 mm (**B**), 2 mm (**C**) cores and the whole section (**D**), from which the phenogroups have been extracted and plotted onto the tSNE plot. A detail of the plot is magnified at the bottom left of each. Note the lack of discrimination of the macrophage population from the adjacent phenogroups in the 1 mm core example, highlighted by a single color rendering. The distribution of CD206, a macrophage marker, is shown in red on the tSNE plot of the whole section (**E**). The table compares core size, cells contained, number of clusters and whether a macrophage cluster was identifiable or not.

Figure 2 Distribution of control and treated lung cells in dimensionality-reduced 2-dimensional space (tSNE). (**A**) Four saline cases from different time points (7, 14, 21 and 28 days) are plotted according to the tSNE bidimensional coordinates, each in an unique color. Each dot represents a cell. (**B**) tSNE plot of all cases, each with an unique color. (**C**) tSNE plot of all saline samples (blue) and Bleomycin-treated samples at days 7, 14 and 21 (red). (**D**) tSNE plot of all saline samples (blue) and Bleomycin-treated samples at day 28 (red).

Figure 3 Classification and frequency of cell types in controls and treated mice. The cell types identified in the lungs are shown as mean ± SD for individual mice for each treatment. AT2 represents AT2 and transitional AT1/AT2 alveolar cells (see supplementary data); bronchial/goblet = bronchial cells and MUC5B cells merged; all stromal = phenoclusters with mesenchymal markers merged. For details see supplementary Table 1.

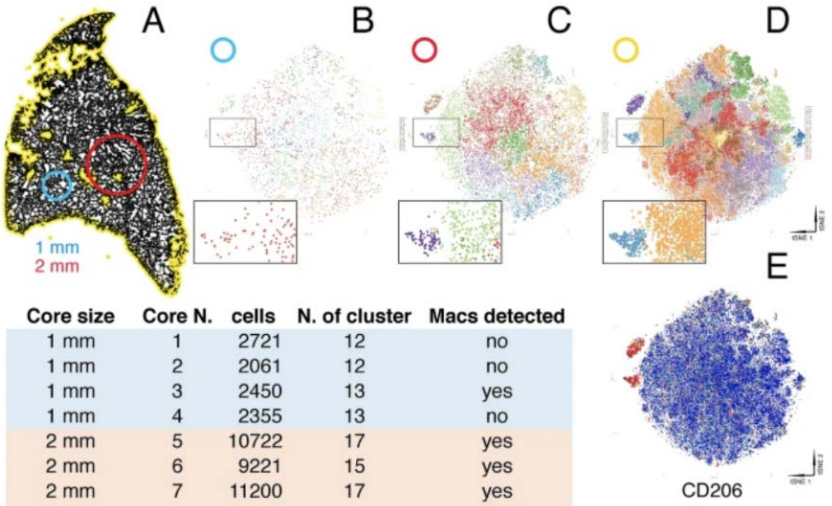
Figure 4 Spatial distribution of cell types. UMAP plots, spatial distribution of phenogroups and immunofluorescence images for representative untreated (**A**–**C**) and day 21 lungs (**D**–**F**). UMAP plots (**A**,**D**) are colored with the cell classification color (bottom legend). The same color-coded groups are plotted on the respective TMA core spatial coordinates (**B**,**E**). A detail is shown in the insets. Note in E a fibroblastic focus next to a bronchial lining population (red) and vessels. On the right five representative markers (color legend at the bottom) are shown

in the whole core or in a high-magnification field (inset) (**C**,**F**). Scale bar 100 µm.

Figure 5 Multidrug transporter (P-gP) levels during Bleomycin treatment in selected cell types. The linear fluorescence levels (0–255), rescaled 0–1 of the multidrug transporter from all experimental animals are plotted as boxplots and outliers for each treatment group and for the selected cell types. AT = all alveolar epithelial cells; Bronch Gobl = bronchial cells and MUC5B + cells merged.

Table 1 Design and antibodies specifications

Figure 1



Core size	Core N.	cells	N. of cluster	Macs detected
1 mm	1	2721	12	no
1 mm	2	2061	12	no
1 mm	3	2450	13	yes
1 mm	4	2355	13	no
2 mm	5	10722	17	yes
2 mm	6	9221	15	yes
2 mm	7	11200	17	yes
2 mm	8	10206	18	yes
whole section	9	40904	22	yes

Figure 2

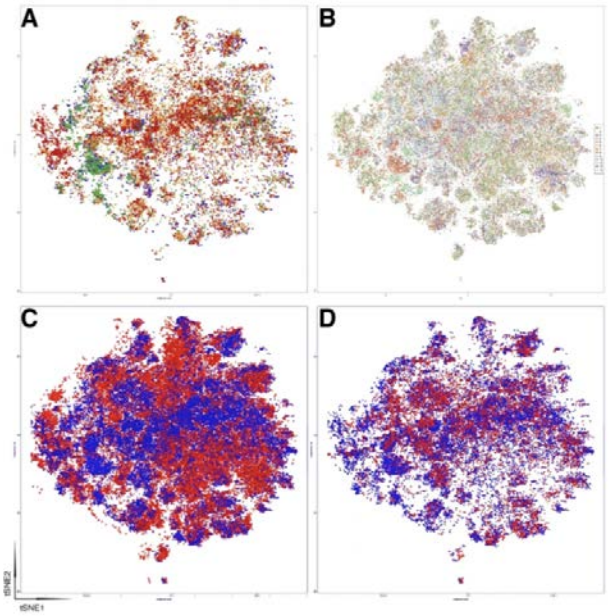


Figure 3

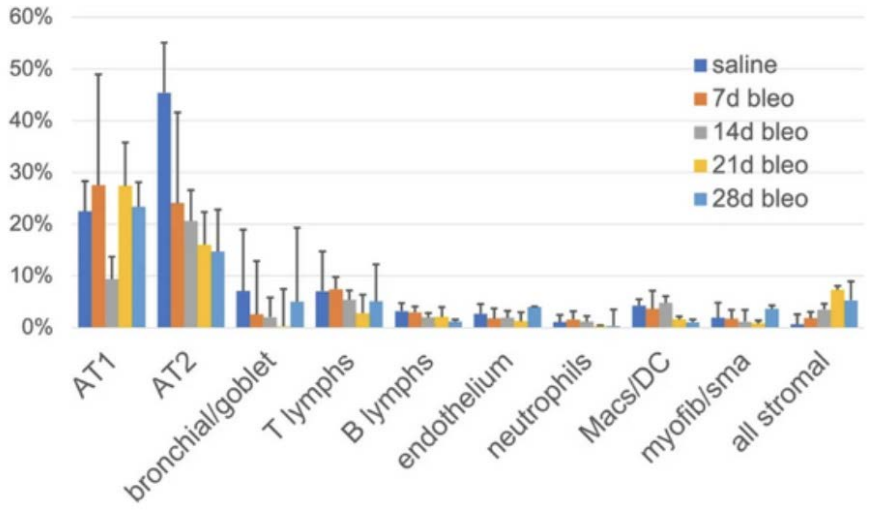


Figure 4

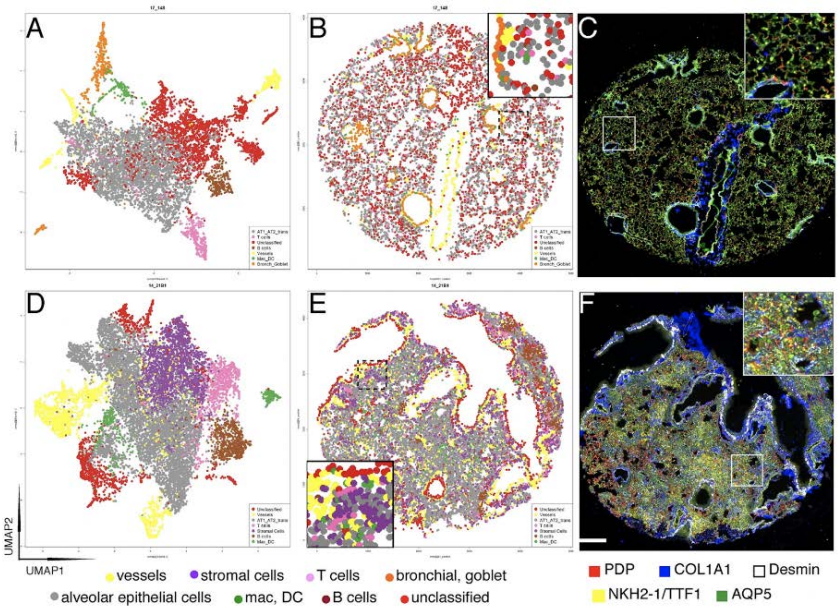

A
B
C
D
E
F
UMAP1
UMAP2
vessels
stromal cells
T cells
bronchial, goblet
alveolar epithelial cells
mac, DC
B cells
unclassified
PDP
COL1A1
Desmin
NKH2-1/TTF1
AQP5

Figure 5

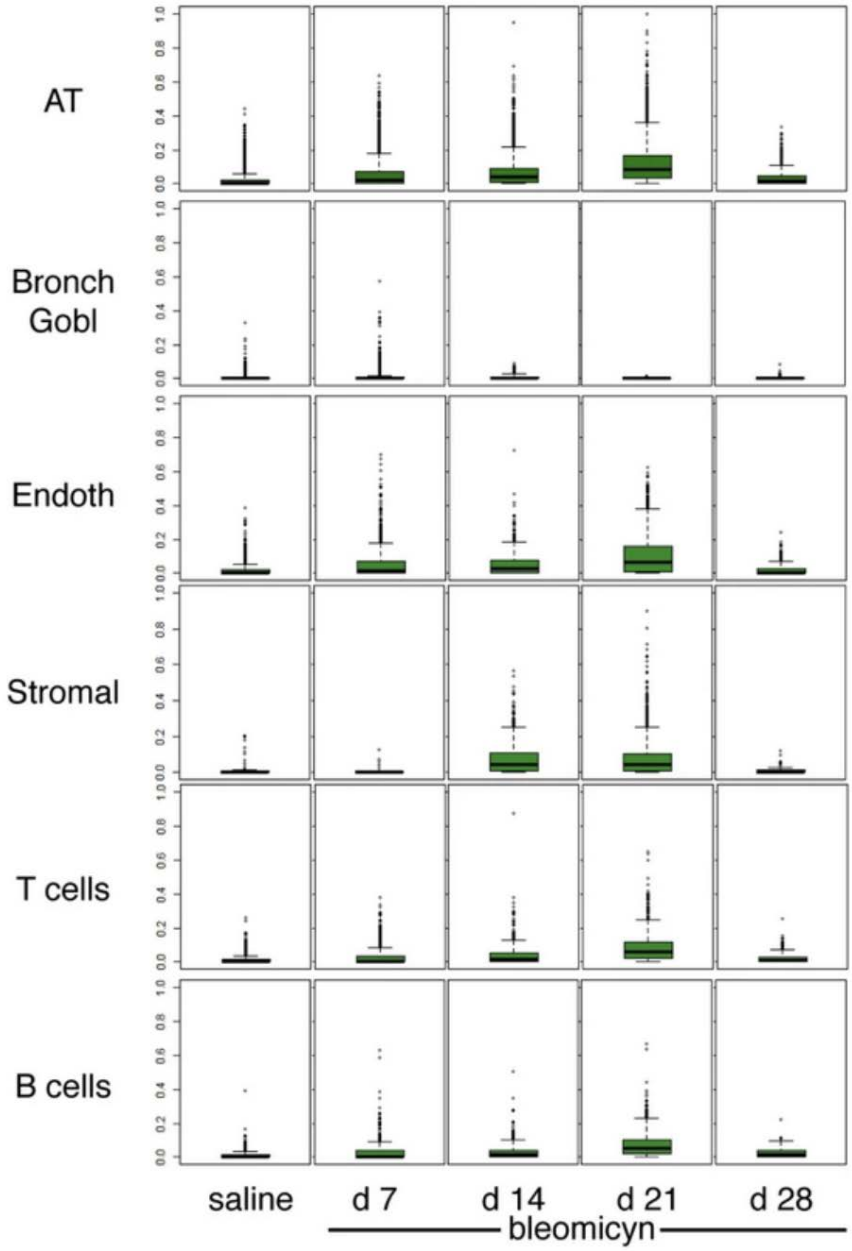

AT
Bronch Gobl
Endoth
Stromal
T cells
B cells
saline
d 7
d 14
d 21
d 28
bleomicyn

Table 1 Design and antibodies specifications

Staining round	Antigen name	Short name	Cell type	scRNAseq ref	Catalog N	Company	Species	Dilution or final concentration (µg/ml)
Primary antibodies								
1	Collagen1	COLL1	Stromal Cells	24 (Col1a1)	ab88147	Abcam	mo IgG3	2.5
1	Alpha smooth muscle actin	aSMA	Vasculature		1A4—M0851	Dako	mo IgG2a	0.7
2	Transgelin	SM22	Vasculature		ab10135	Abcam	goat	4
2	Desmin	DESM	Vasculature		D33—M076	Dako	mo IgG1	1.3
3	Von Willebrand Factor	vWF	Vasculature		AP22470PU-N	Origene	goat	2.5
3	Multiple Cytokeratins	PanCK	Epithelial Cells		MA1-82041	Thermo Fisher	mo IgG1	1:40
4	Mucin	MUC5b	Goblet Cells		ab77995	Abcam	mo IgG1	5
4	Macrophages Mannose Receptor	CD206	Macrophages	23 (Mrc1)	AF2535	R&D	goat	1
5	Surfactant D	SPD	AT2, alveolar progenitors	24 (Sftpd)	Sc-25324	Santa Cruz B	mo IgG1	1
5	Paired box protein 5	PAX5	B Lymphocytes		ab109443	Abcam	rb Mab	1
6	Aquaporin 5	AQP5	AT1	24	Sc-514022	Santa Cruz B	mo IgG1	1
6	Osteopontin	OPN	Macrophages, neutrophils	24	AF808	R&D	goat	1
7	Podoplanin	PDPN	AT1	24	ab11936	Abcam	s.hamster	5
7	Matrix Metalloprotease 7	MMP7	–		D4H5—3801	Cell Signaling	rb Mab	5
8	Thyroid Transcription Factor1	NKX2-1/TTF1	AT2	24 (Nkx2-1)	8G7G3—M3575	Dako	mo IgG1	1
8	T cells	CD3	T Lymphocytes		ab5690	Abcam	rb poly	0.3
9	Advanced Glycosylation End-Product Specific Receptor	RAGE	AT1	23 (Ager)	Sc-80653	Santa Cruz B	mo IgG1	1
9	Myeloperoxidase	MPO	Monocyte progenitors	24	ab9535	Abcam	rb poly	1:200
10	Homeobox only protein	HOP	AT1	24	Sc-398703	Santa Cruz B	mo IgG1	1
10	ATP-binding cassette sub-family A member 3	ABCA3	AT2	24	ab24751	Abcam	mo IgG2a	1
10	Fibroblast growth factor 10	FGF10	Stromal Cells	23	GTX30007	GeneTex	goat	1
11	Vimentin	VIM	Stromal Cells		Sc-373717	Santa Cruz B	mo IgG1	1
11	Multidrug resistance 1	P-gP	–		orb11267	Biorbyt	rb poly	1
11	Breast cancer resistance protein	BCRP	–		ab24115	Abcam	rat IgG2a	1

Name	Dilution	Catalog N
Secondary antibodies		
Alexa Fluor® 488 AffiniPure Donkey Anti Goat IgG (H + L)	1:500	705-545-147
Alexa Fluor® 488 AffiniPure Goat Anti Rabbit IgG (H + L)	1:500	111-545-144
Alexa Fluor® 488 AffiniPure Goat Anti-Mouse IgG, Fcγ subclass 1 specific	1:200	115-545-205
Alexa Fluor® 647 AffiniPure Donkey Anti-Goat igG (H + L)	1:150	705-005-003
Alexa Fluor® 647 AffiniPure Goat Anti Mouse IgG, Fc subclass 1 specific	1:500	115-605-205
Alexa Fluor® 647 AffiniPure Goat Anti Mouse IgG, Fc subclass 3 specific	1:500	115-605-209
Alexa Fluor® 647 AffiniPure Donkey Anti Rat IgG (H + L)	1:500	712-605-153
Rhodamine (TRITC) AffiniPure Goat Anti Mouse IgG, Fc subclass 2a specific	1:200	115-295-206
Rhodamine (TRITC) AffiniPure Donkey Anti-Rabbit IgG (H + L)	1:200	711-295-152

https://www.nature.com/articles/s41598-022-12738-9/tables/1

References

1. Method of the Year. Single-cell multimodal omics. Nat. Methods 17(1), 2020. https://doi.org/10.1038/s41592-019-0703-5 (2019).

2. De Smet, F., Antoranz Martinez, A. & Bosisio, F. M. Next-generation pathology by multiplexed immunohistochemistry. Trends Biochem. Sci. 46, 80–82. https://doi.org/10.1016/j.tibs.2020.09.009 (2021).

3. Kononen, J. et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat. Med. 4, 844–847 (1998).

4. Clohessy, J. G. & Pandolfi, P. P. Mouse hospital and co-clinical trial project—From bench to bedside. Nat. Rev. Clin. Oncol. 12, 491–498. https://doi.org/10.1038/nrclinonc.2015.62 (2015).

5. Mascadri, F. et al. Background-free detection of mouse antibodies on mouse tissue by anti-isotype secondary antibodies. J. Histo chem. Cytochem. 69, 535–541. https://doi.org/10.1369/00221554211033239 (2021).

6. Cattoretti, G., Bosisio, F., Marcelis, L. & Bolognesi, M. M. Multiple Iterative Labeling by Antibody Neodeposition (MILAN). https://
protocolexchange.researchsquare.com/article/nprot-7017/v5 (2019).

7. Miles, T. et al. The contribution of animal models to understanding the role of the immune system in human idiopathic pulmonary fibrosis. Clin. Transl. Immunol. 9, e1153. https://doi.org/10.1002/cti2.1153 (2020).

8. Jenkins, R. G. et al. An Official American Thoracic Society Workshop Report: Use of animal models for the preclinical assessment of potential therapies for pulmonary fibrosis. Am. J. Respir. Cell Mol. Biol. 56, 667–679. https://doi.org/10.1165/rcmb.2017-0096ST (2017).

9. Denayer, T., Stöhr, T. & Roy, M. V. Animal models in translational medicine: Validation and prediction. Eur. J. Mol. Clin. Med. https://doi.org/10.1016/j.nhtm.2014.08.001 (2014).

10. Tashiro, J. et al. Exploring animal models that resemble idiopathic pulmonary fibrosis. Front. Med. (Lausanne) 4, 118. https://doi. org/10.3389/fmed.2017.00118 (2017).

11. Hogan, B. L. M. et al. Repair and regeneration of the respiratory system: Complexity, plasticity, and mechanisms of lung stem cell function. Cell Stem Cell 15, 123–138. https://doi.org/10.1016/j.stem.2014.07.012 (2014).

12.Della Latta, V., Cecchettini, A., Del Ry, S. & Morales, M. A. Bleomycin in the setting of lung fibrosis induction: From biological mechanisms to counteractions. Pharmacol. Res. 97, 122–130. https://doi.org/10.1016/j.phrs.2015.04.012 (2015).

13.Stellari, F. F. et al. Heterologous matrix metalloproteinase gene promoter activity allows in vivo real-time imaging of bleomycininduced lung fibrosis in transiently transgenized mice. Front. Immunol. 8, 199. https://doi.org/10.3389/fimmu.2017.00199 (2017).

14.Park, J. K. et al. Bleomycin induces drug efflux in lungs a pitfall for pharmacological studies of pulmonary fibrosis. Am. J.

Respir. Cell Mol. Biol. 62, 178–190. https://doi.org/10.1165/rcmb.2018-0147OC (2020).
15.Strunz, M. et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. Nat. Commun. 11, 3559. https://doi.org/10.1038/s41467-020-17358-3 (2020).
16.Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. Cell Stem Cell 27, 366-382.e367. https://doi.org/10.1016/j.stem.2020.06.020 (2020).
17.Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. Nat. Cell Biol. 22, 934–946. https://doi.org/10.1038/s41556-020-0542-8 (2020).
18. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature 587, 619–625. https://doi. org/10.1038/s41586-020-2922-4 (2020). 19.Han, X. et al. Mapping the mouse cell atlas by microwell-seq. Cell 172, 1091-1097.e1017. https://doi.org/10.1016/j.cell.2018.02. 001 (2018).
20. Karkanias, J. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature https://doi.org/10.1038/s41586018-0590-4 (2018).
21.Denisenko, E. et al. Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol.

21, 130–125. https://doi.org/10.1186/s13059-020-02048-6 (2020).
22.Amir, E.-A.D. et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol. 31, 545–552. https://doi.org/10.1038/nbt.2594 (2013).
23.Caicedo, J. C. et al. Data-analysis strategies for image-based cell profiling. Nat. Methods 14, 849–863. https://doi.org/10.1038/ nmeth.4397 (2017).
24. Longo, S. K., Guo, M. G., Ji, A. L. & Khavari, P. A. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat. Rev. Genet. 22, 627–644. https://doi.org/10.1038/s41576-021-00370-8 (2021). 25. Mecozzi, L. et al. In-vivo lung fibrosis staging in a bleomycin-mouse model: A new micro-CT guided densitometric approach. Sci. Rep. 10, 18735. https://doi.org/10.1038/s41598-020-71293-3 (2020).
26.Ashcroft, T., Simpson, J. M. & Timbrell, V. Simple method of estimating severity of pulmonary fibrosis on a numerical scale. J. Clin. Pathol. 41, 467–470. https://doi.org/10.1136/jcp.41.4.467 (1988). 27. Hubner, R. H. et al. Standardized quantification of pulmonary fibrosis in histological samples. Biotechniques 44(507–511), 514–507. https://doi.org/10.2144/000112729 (2008). 28.Bolognesi, M. M. et al. Multiplex staining by sequential immunostaining and antibody removal on routine tissue sections. J. Histochem. Cytochem. 65, 431–444.

https://doi.org/10.1369/0022155417719419 (2017). 29.Furia, L., Pelicci, P. G. & Faretta, M. A computational platform for robotized fluorescence microscopy (I): High-content imagebased cell-cycle analysis. Cytometry A 83, 333–343. https://doi.org/10.1002/cyto.a.22266 (2013).

30.van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).

31. McInness, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv.org (2018).

32.Levine, J. H. et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell 162, 184–197. https://doi.org/10.1016/j.cell.2015.05.047 (2015).

33. Manzoni, M. et al. The adaptive and innate immune cell landscape of uterine leiomyosarcomas. Sci. Rep. 10, 702–710. https://doi. org/10.1038/s41598-020-57627-1 (2020).

Chapter 4

Next-Generation Pathology Using Multiplexed Immunohistochemistry: Mapping Tissue Architecture at Single-Cell Level

Francesca Maria Bosisio [1], Yannick Van Herck [2], Julie Messiaen [1 3 4], Maddalena Maria Bolognesi [5 6], Lukas Marcelis [1], Matthias Van Haele [1], Giorgio Cattoretti [5 6], Asier Antoranz [1 3], Frederik De Smet [1 3]

Affiliations [1] Translational Cell and Tissue Research Unit, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium. [2] Department of Oncology, KU Leuven, Leuven, Belgium. [3] The Laboratory for Precision Cancer Medicine, Translational Cell and Tissue Research Unit, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium. [4] Department of Pediatrics, University Hospitals Leuven, Leuven, Belgium. [5] Pathology, Department of Medicine and Surgery, Università di Milano-Bicocca, Monza, Italy. [6] Department of Pathology, Azienda Socio Sanitaria Territoriale (ASST) Monza, Ospedale San Gerardo, Monza, Italy.

Front. Oncol., 05 August 2022 Sec. Cancer Molecular Targets and Therapeutics

https://doi.org/10.3389/fonc.2022.918900

Abstract

Single-cell omics aim at charting the different types and properties of all cells in the human body in health and disease.

Over the past years, myriads of cellular phenotypes have been defined by methods that mostly required cells to be dissociated and removed from their original microenvironment, thus destroying valuable information about their location and interactions. Growing insights, however, are showing that such information is crucial to understand complex disease states. For decades, pathologists have interpreted cells in the context of their tissue using low-plex antibody- and morphology-based methods. Novel technologies for multiplexed immunohistochemistry are now rendering it possible to perform extended single-cell expression profiling using dozens of protein markers in the spatial context of a single tissue section. The combination of these novel technologies with extended data analysis tools allows us now to study cell-cell interactions, define cellular sociology, and describe detailed aberrations in tissue architecture, as such gaining much deeper insights in disease states. In this review, we provide a comprehensive overview of the available technologies for multiplexed immunohistochemistry, their advantages and challenges. We also provide the principles on how to interpret high-dimensional data in a spatial context. Similar to the fact that no one can just "read" a genome, pathological assessments are in dire need of extended digital data repositories to bring diagnostics and tissue interpretation to the next level.

Keywords: methods for spatial profiling; multiplexed immunofluorescencence and immunohistochemistry; single-cell ‘omics; spatial profiling; tissue architecture analysis.

Introduction

For centuries, medical sciences have tried to achieve a deep understanding of the human body, both in health and disease. Twenty years ago, a major hurdle was crossed with the mapping of the human genome (1). However, it is now becoming clear that one cannot just “read” a genome and subsequently understand or predict the principles that underlie human biology or disease. The sole true interpreters of the genome are cells, and understanding how the genome functions within cells, how cells form tissues and dynamically remodel their activities when they progress towards disease, is among the greatest scientific and technological challenges of our era. The goal is no longer to find differences in “bulk” genomic readouts but rather to see explicit changes in a specific set of cells and to predict their behavior. In this light, multiple human cell atlas initiatives are working towards describing every cell of the human body, as a reference map to accelerate progress in biomedical science (2–5). These ambitious projects – similar in scale to the human genome project – aim to chart the different types and molecular properties of all human cells in health and disease, for which a multitude of organ-oriented working groups are mapping the single-cell composition and their spatial architectures. Technological advances in the field of single-cell ‘omics’ such as single cell genomics, epigenomics, transcriptomics, and proteomics, and even their combinations in a multi-omic setting (6, 7) are now rendering it possible to map physiological

features of each individual cell in an organ as a functional unit. However, current methods still mostly require that cells are dissociated and removed from their original microenvironment, thus destroying valuable information about their location and interactions – information that is crucial to understand many disease patho-physiologies. The next, ongoing step is aimed at describing single-cell features in their natural microenvironment (8).At the moment, understanding cellular functions at single-cell level within the context of a tissue is primarily done by expression profiling, using either transcriptional (9) or protein-based multiplex methods in pathological tissue sections (10), even though the realm of spatial omics keeps on growing fast (11, 12). While methods for enhanced spatial omics are only now starting to become available, pathologists have been evaluating cells in tissue sections using classical antibody- and morphology-based (i.e. H&E staining) methods for decades (13). Daily clinical practice is mostly performed using classical (chromogenic) immunohistochemical (IHC) methods that allow the simultaneous assessment of one or two proteins in a single tissue slide, which are mostly evaluated in a visual, semi-quantitative way by a pathologist (14). Novel technologies and methods, that will be discussed in this review, are now making it possible to perform quantitative spatial, antibody-based expression profiling of dozens of protein markers in a single tissue section. This will, when carefully selected, provide deeper insights in disease states while offering the ability to study cell-cell interactions, precisely define

disease-related niches, all within the original context of the tissue (see below).Such technology also comes along with multiple challenges as well. First, a careful selection of the right technology to answer the biological question is crucial: as we will discuss below, this is defined by the type and size of a tissue, the number of samples that need to be processed, and the number of markers that need to be interrogated. Second, besides the technical hurdles that need to be overcome to collect high quality, pathology-grade images in which each marker is carefully monitored, storing and processing the large volumes of image data also pose a significant logistic and infrastructural challenge. Finally, the plethora of information that is obtained from these analyses is also becoming of such magnitude and complexity that mere eyeballing of a tissue by a pathologist or researcher is no longer sufficient to properly extract information and interpret expression patterns. Similar to the fact that mutations in a genome need to be interpreted for biological and clinical relevance, pathological assessments also need the installation of suitable analysis algorithms and extended digital data repositories to bring diagnostics and tissue interpretation to the next level.

In this review, we provide a comprehensive overview of the available technologies for multiplexed immunohistochemistry, their advantages and challenges, and provide the basic principles on how to interpret high-dimensional data in a spatial context.

Methods for multiplexed immunohistochemistry

Lately, the armamentarium of technologies for antibody-based multiplexed IHC is rapidly growing. While several technologies have been described in literature performing manual IHC protocols, there is a clear trend of seemingly ‘plug-and-play’ instruments entering the market that automated the same principles to some extent. Importantly, however, since all these technologies depend on an antibody-based detection of proteins in a large variety of tissue types, the selection, validation and performance of the used antibodies have to be done with sufficient care. Adequate minimal validation guidelines need to be set, including the use of appropriate positive and negative controls (e.g. based on classical chromogenic stains (15)), to guarantee sensitivity, specificity and warrant the collection of biologically relevant data. Indeed, the performance of an antibody can vary enormously depending on the tissue type, the experimental conditions in which the antibody is applied, and whether formalin-fixed paraffin embedded (FFPE) or frozen materials are used (16–18). As such, the development and optimization of suitable antibody panels for multiplexed IHC still requires a significant amount of time, although externally validated reagents to detect commonly used markers are becoming more routinely available, for instance thanks to large scale initiatives such as the Protein Atlas consortium (19) or NIH initiatives (20).

The first step, even before exploring different multiplex staining methods, is figuring out if and/or how spatial analysis can help in answering your biological question. Spatial analysis can provide additional information on proximity-based cell-cell interactions, such as the infiltration of immune cells in a tissue, their proximity to tumor cells (see below). Subsequently, once it has been decided that spatial analysis will be needed to answer a particular biological question, the selection of the method becomes key (Figure 1, Table 1). As indicated above, the selection of the most appropriate method is largely defined by several parameters. First, the type (i.e. frozen *vs* FFPE) and size (i.e. needle biopsy *vs* tissue microarray (TMA) *vs* whole tissue slide) of a tissue, in addition to the number of samples that need to be analyzed (from 1 sample at the time to cohorts of hundreds of patients), will already define the first selection. As we will describe below, not all methods are compatible with FFPE/Frozen or large tissue samples, or can be easily scaled up to analyse hundreds of slides in a practical timeframe and with minimal variance induced by batch-effects. A second important parameter is linked to the number of markers that needs to be interrogated, a feature that is commonly highly project specific (e.g. 5 *vs* 50 markers). Finally, understanding how the currently available methods for multiplexed IHC work will be crucial to select the most appropriate one. (Figure 1, Table 1)

Overall, technologies for multiplexed IHC can be classified by the way the antibodies are administered and detected. Indeed,

depending on the technology, large mixtures of antibodies (+20) can be administered simultaneously to a slide, following which a dedicated instrument is able to image, resolve and unmix the location of each antibody in the tissue. Alternatively, various methods also use cyclic procedures through which smaller amounts of antibodies (2–4) are used for staining and imaging, following which the detected signal is removed. By repeating this cycle multiple times, large numbers of markers can be detected in the same tissue section. Finally, several hybrid methods are available that combine both approaches or even use other technologies (e.g. NGS analysis) to resolve the complex mixtures of markers. In the first part of the review, we provide an extended overview of several of the currently available technologies for multiplexed IHC analysis on tissue sections, including their compatibility with materials, antibodies, throughput, plex level, timing, and compatibility with standard or dedicated instrumentation. Methods are grouped by the type of modification that is used to detect the various antibodies.

Fluorescence-Based Detection of Antibody Mixtures

The largest group of methods for multiplexed IHC depends on the detection of mixtures of antibodies using fluorescent signals. Indeed, because fluorophores harbor specific excitation and emission spectra, they can be resolved using commonly used or more advanced microscopy tools, depending on how many fluorophores require resolving and their spectral overlap. In addition, most imaging instruments achieve a resolution

below 0.6µm and are therefore perfectly compatible with single cell measurements [the average human nucleus is approximately 10 micrometers in diameter (45)]. A common issue with these methods is however the presence of autofluorescence in tissue: upon excitation with specific wavelengths, naturally occurring substances in tissue [for example extracellular matrix components, lipo-pigments, aromatic amino acids and flavins (46)] will emit light which overlaps with the fluorescence of the measured fluorophore, and needs to be considered and dealt with while processing images. Some chemical treatments (e.g. using bleach, sudan black or borohydrate) have been suggested to remove autofluorescence, but need to be used carefully and typically only solve the problem partially, while computational methods often offer better solutions, although in that case sufficient control images that capture autofluorescence need to be collected.

Cyclic Methods Using Fluorescent Antibody Detection

These methods make use of a cyclic procedure in which several steps (Figure 1A), including (i) the staining of small numbers of antibodies (2 to 4; typically defined by the microscope settings), (ii) the imaging of the sample and (iii) the removal of the stain, are repeated multiple times until all markers are detected. While cyclic methods are in general more time consuming, this approach allows to perform *interim* evaluations (allowing the researcher to validate every individual step and if needed repeat them), refine the composition of the panel during the

procedure and adapt to unforeseen problems/results since cyclic methods conserve the tissue during analysis.

In these procedures, either directly labelled primary antibodies or labelled secondary antibodies are used to detect the markers. The former approach is used in the multi-epitope-ligand cartography (MELC) (37), multiplexed immunofluorescence (MxIF) (35), cyclic immunofluorescence (CyCIF) (24), Cyclic Multiplexed-Immunofluorescence (cmIF) (36), MACSima Imaging Cyclic staining (MICS) (25) and Iterative Bleaching Extends multi-pleXity (IBEX) (38) procedures. Indeed, MELC, MxIF, CmIF, IBEX and CyCIF make use of directly labelled antibodies which are usually combined in a triple/quadruple staining using 3/4 distinct fluorophores with non-overlapping spectra. Following staining and imaging, the fluorescent signal is removed using a photo-induced or chemical bleaching step before probing for the next markers (24, 35–38). This is different for the MICS technology, which makes use of recombinant antibodies from which the fluorescent label can be removed using a proprietary enzymatic cleavage reaction (25). In either case, antibodies [or the Fab fragments which are left after cleavage (20)] remain in the tissue when the next round of markers are added to the tissue slide. The latter is different in the MILAN approach (21–23), which makes use of fluorescently labelled secondary antibodies that bind to the specific primary antibodies. This has the advantage that signals can be amplified making it easier to detect weakly expressed markers, which might be an issue

when using directly labeled antibodies without amplification. Another difference is that in the MILAN procedure, antibodies are entirely removed using an SDS/ßMercaptoethanol washing step to denature, inactivate, and remove the antibodies from the tissue (15), avoiding potential issues with antibody crowding or steric hindrance which could theoretically arise in methods that do not remove the antibodies (even though there is no formal evidence for such issues at the moment). The downside, however, is that the usage of secondary antibodies in MILAN forces users to make combinations of primary antibodies that were raised in different hosts or harbor different isotypes to avoid cross-reactivity during primary antibody detection (e.g. combinations of Mouse IgG1, Mouse IgG2, rat, goat and/or rabbit need to be made). On the other hand, directly labelled procedures depend on the direct labelling of antibodies, which requires careful selection of color combinations and carrier-free formulations of the primary antibody solutions (which may require custom made formulations), but, once available and validated, can be combined independent of the species where the antibody was raised.

In either of these methods, antibodies are commonly administered to the tissue in small batches (2 to 4) after which the tissue is imaged. Importantly, this setting allows the usage of regular microscopy and imaging tools using common fluorescent channels that are typically available in laboratories across the world. Moreover, automation of these procedures is gradually increasing, including the use of autostainers and

automated slide scanners with regular fluorescent settings, and novel technologies using microfluidics (such as incorporated in the LABSAT and COMET system from Lunaphore, the CODEX system from Akoya (see below), or the MACSima instrument for MICS) are rendering it possible to further speed up the acquisition for small numbers of slides (26). The MICS method, on the other hand, has been optimized on a proprietary instrument, which allows unassisted, automatic processing of a small number of tissue sections (25). Similarly, chipcytometry (27) makes use of highly specialized equipment in which microchambers containing attached cells or (frozen) tissue can be mounted, and subsequently subjected to large multiplex analysis, which is primarily done in a cyclic way, one antibody at the time.

Finally, these methods are primarily compatible with FFPE tissue sections, while the MELC and MICS method have also been described to be compatible with frozen materials (15, 25, 47, 48). Overall, using these technologies, large numbers of markers can be analyzed (50+ have been reported (see Table 1), and the number of analytes keeps on rising constantly. Here, we added a new example of a melanoma tissue sample that was stained for 82 markers using the MILAN method (23) over the course of 50 rounds (Supplement Figure 1), while the MACSima technology was used to stain 327 markers in tonsil tissue over 160+ rounds (25). Overall, the primary limitation of the number of markers mainly comes from the compatibility of the tissue with the antibody removal procedure: methods

involving chemical bleaching steps (such as in CyCIF and MxIF) eventually lead to tissue destruction, loss of antigenicity and/or tissue loss, and are therefore limited in the number of cycles that can be performed (47, 49). Also, cover slips that have to be added/removed repeatedly between the staining/imaging cycles, may cause tissue damage. Finally, the tissue type and how it was preserved prior to the start of the mIHC largely determines the plex level that can be achieved and needs to be defined experimentally.

Next to tissue damage, also the scalability of methods needs attention: while some methods (e.g. MELC, CyCIF and MILAN) allow the simultaneous processing of multiple slides, others (e.g. MICS, COMET) are limited to 2-4 slides at the time, although robotic systems that allow automated slide loading are in development. Also, the area that can be imaged/scanned can differ greatly among systems (from a few mm (2) to whole slide). Scaling can be further enhanced by compiling tissue microarrays (TMA), in which carefully selected tissue subsamples (also referred to as ‘cores’) from large tissue blocks are assembled on a single slide (typically 60 cores of 2mm diameter can be put on the same slide). Considering there is more heterogeneity observed across the dimensions of a single section than between different sections in tumors (50), one should deliberately design the TMA to capture this potential heterogeneity, by taking a smaller number of cores from small and homogeneous samples and a larger number in big and heterogeneous specimens. The role of the pathologist in

selecting the different regions of interest and design of the TMA remains key. In this manner, it has been shown that different tumor biomarkers are accurately reported through the assessment of TMAs (51–53). As such, cohorts of 50-150 patients can be analyzed within the timeframe of a couple of weeks. This workflow is highly compatible within the research context, for instance in performing retrospective analyses simultaneously on big patient groups avoiding batch effects. However, within the daily clinical routine, implementation of TMAs in the workflow is less obvious, even though there may be a role for them when scan areas of more automated (faster) mIHC systems are too small to cover the entire tissue section as a whole and the analysis of multiple regions of interest is still needed.

Batch Methods Using Fluorescent Antibody Detection

In addition to cyclic procedures, novel methods are appearing that allow more extended, single-step multiplexing (Figure 1B). This is currently achieved by labelling antibodies with specific fluorophores which can subsequently be resolved using spectral unmixing methods, such as used in the RareCyte Orion system (39). The advantage of this approach is that a maximum of currently 21 antibodies are applied simultaneously in a single staining procedure, as such offering a fast staining of the sample. However, the generation of antibody panels containing 20+ different fluorescent dyes can be challenging, and requires dedicated instrumentation for spectral unmixing. Moreover, a

significant amount of time (i.e. multiple hours) is required to collect high-resolution images from a small number of whole slides for each included channel. Finally, the simultaneous addition of all markers limits the flexibility to modify the panel after it has been validated, and the generation of the labelled antibodies remains cumbersome and requires a careful selection and validation of the multicolor panel. On the other hand, for routine purposes where the same panel has to be applied repeatedly, this method may offer excellent options, although this has not been investigated yet. Another recent approach enabling higher level multiplexing involves the UltraPlex, hapten-based technology, which adds labels to primary antibodies, which are subsequently detected by anti-hapten antibodies coupled to various fluorophores. While so far only lower plex panels (4-8-plex) were tested, it bears the potential to scale to routine assays of 12 or more markers (54).

Hybrid Approaches Using Fluorescent Marker Detection

In addition to cyclic and batch procedures, other methods use a hybrid approach to achieve extended multiplexing (Figure 1C). This is currently achieved by labelling antibodies with unique nucleotide barcodes which are subsequently detected using either direct hybridization of fluorescently labelled complementary nucleotide probes (as in the CODEX system (30–32)), or by using an *in-situ* amplification system (as used in the immunoSABER system (40) or the InSituPlex system (Ultivue) (41)). The hybrid nature of these approaches comes from the batch application of all antibodies simultaneously in a

first step, while their subsequent detection using the detector probes is subsequently done in a cyclic fashion where hybridization/denaturation steps are alternated with imaging, with the latter needing to be repeated for each cycle (as such leading to a time scale comparable to the cyclic methods). An alternative approach was recently described in the SeqStain procedure (55), where antibodies labelled with fluorescently-labelled DNA (either primary or secondary) are used to detect protein markers, following which the DNA is removed by an enzymatic reaction using a nuclease. By repeating these cycles, multiple markers could be visualized.

These approaches require that antibodies are labelled with nucleotide probes, which reduces flexibility and speed to design novel panels. Moreover, the nucleotide sequence composition of the probes still requires significant optimization to avoid nonspecific binding. On the other hand, unlimited numbers of antibodies could theoretically be labelled with unique barcodes, and as such be combined in large quantities. It remains to be seen, however, how feasible such extended approach will be, whether and when steric hindrance/overcrowding will become an issue, and how destructive multiple rounds of hybridization/denaturation will be for the tissue structure, a step that will probably define the plex level. The cyclic application of the probes also requires specific instrumentation containing microfluidic devices and temperature control or repeated manual work, currently complicating the scalability of this approach to large batches of slides. Finally, while the CODEX

was initially optimized for frozen materials, both methods are now compatible with either FFPE or frozen tissue sections (32, 40).

Detection Using Fluorescent Precipitates

Finally, while all the above-mentioned technologies allow the measurements of large numbers of protein markers, in many cases, researchers don't need such complicated systems. Alternatively, a limited number of markers can be detected by using fluorescent precipitates (Figure 1D), for example by using tyramide signal amplifications (TSA) reagents, an enzyme-mediated detection method (34). The detection of antibodies in this system is based on a cyclic procedure, where each primary antibody is stained separately, probed with a Horse Radish peroxidase (HRP) secondary antibody, and then a specific fluorescent precipitate of the TSA reagent is generated. At the end of each cycle, the primary and secondary antibodies are removed while leaving the fluorescent precipitate before probing with the next antibody. Because all precipitates harbor a different fluorescent spectrum, imaging is only done once at the end, after which antibodies are spectrally unmixed. This concept is used in for example the OPAL system (34), allowing to perform 6-plex staining of FFPE tissue in an automated fashion using commonly available autostainers (typically present in pathology labs) and a dedicated spectral scanner. This technology is compatible with FFPE materials, and tens of slides can be stained simultaneously depending on

the autostainer instruments used. In addition, it has to be said that some fluorescence-based detection systems are using directly linked primary antibodies. Therefore, it could be suggested that sensitivity is lower compared to the generally used chromogenic staining, which is often used with an amplification (frequently a polymer with several HRP enzymes) of the signal (56). Currently, there is a lack of good comparative studies which investigate this question. Some studies even suggest a comparable sensitivity (57, 58). A proper evaluation of the advantages and disadvantages of each technology regarding specific research/clinical question should be done.

2. Chromogenic Detection of Antibody Mixtures

Next to fluorescent detection methods, the use of more classical chromogens has been employed to develop methods for multiplexed IHC (Figure 1E). The Sequential immunoperoxidase labeling and erasing (SIMPLE (42)) or the multiplexed immunohistochemical consecutive staining on single slide (MICSSS (43, 44)), are methods that use the alcohol-soluble peroxidase substrate 3-amino-9-ethylcarbazole, a reagent that in the presence of the commonly used HRP and H_2O_2, generates a chromogenic, red precipitate that can be imaged using regular white light, brightfield microscopy. Once the image is collected, the red precipitate is washed out using ethanol following which the antibody is eluted using acidified permanganate. As such, 5 to 10-plex stains have been described (42, 43). While this method is largely compatible with

standard equipment and procedures in pathology laboratories, it remains to be seen whether the tissue can withstand increased rounds of the low pH antibody washout buffer. Another recent approach using chromogenic multiplexing involves the UltraPlex, hapten-based technology, allowing to generate low-plex analyses (54). So far mainly FFPE tissues were processed, while frozen samples were not analyzed yet. Finally, while easy to assess whether markers are highly expressed or completely absent, the quantification of chromogens is far less quantitative than fluorescent dyes.

3. Mass-Cytometry Based Detection of Antibody Mixtures

As an alternative for chromophore/fluorescence-based imaging, imaging mass cytometry (IMC) makes use of metal-labelled antibodies, which are resolved using mass-spectrometry, an approach that is currently used in the Hyperion (28) or the multiplexed ion beam imaging (MIBI/IONPath) (29) systems (Figure 1F). At the moment, 42 purified metal isotopes, mostly from the lanthanide series, are commercially available for labelling purposes, although the theoretical amount could be 135 based on the possible isotopes, a number that is mostly limited by the excavation of these metal isotopes in sufficient amounts and purity. The analysis of tissues using IMC involves the staining of the tissue with the mixture of all pre-titrated antibodies together, following which laser-assisted ionization allows the analysis of the generated cloud containing the metal ions in a connected mass cytometer. The batch application of

antibodies also makes it compatible with both frozen and FFPE materials. The resolution is currently 0.3-1µm, dependent on the system and settings, but imaging is rather slow at a rate of 2h/mm (2), which puts some constraints on the analysis of whole tissue slides or large cohorts of patients. Finally, similar to several other methods, labelling of antibodies is required, although commercial kits are available that are sufficiently easy to use (59).

4. Sequencing-Based Detection of Antibody Mixtures

A final approach to perform high-dimensional multiplexed IHC involves the combination of nucleotide-labelled antibodies and next-generation sequencing (NGS) (Figure 1G). Indeed, the Digital Spatial Profiling platform (DSP) makes use of photo-cleavable probes to label primary antibodies (33). By staining tissues with these antibodies, precise illumination of specific regions of interest in the tissue using a dedicated platform allows the isolation of the cleaved nucleotides which are subsequently quantified using NGS. As such, multiplexing of up to 40 proteins (or 5,000 mRNA probes) has been described (33), but this technology allows up to 800- of 80-plex profiling of either mRNA or protein, respectively, using an optical barcode readout and has the potential for even greater multiplexing using an NGS readout. This method is also compatible with both FFPE and frozen materials. However a minimum of (non-adjacent) cells (10-20 cells for protein, 50-200 cells for RNA (60)) is required to obtain sufficient probes to

achieve high quality NGS data. As such, this technology does not yet achieve true single cell analysis, although specific cell populations could be profiled in depth. The selection of the cells to study, a critical step in this technology, can still be done by performing a lowplex IHC staining prior to probe isolation, but the isolation of specific, more complex, rare phenotypes could be hampered by this low-plex staining procedure.

Data analysis tools for pathological interpretation

As described above, a multitude of technologies are currently available to measure dozens of protein markers using multiplexed IHC at single-cell level in a tissue slide. Most technologies make use of automated slide scanners, which are able to image a wide range of tissue areas, from pre-selected (small) regions of interest (ROIs) to entire slides/tissue samples. Such output is subsequently subject to a detailed analysis process. Indeed, while the 'wet lab' procedures may still be adopted relatively quickly and easily by laboratories, the subsequent image analysis and interpretation of the resulting high-dimensional data still faces enormous challenges. In the second part, we therefore describe a general workflow of methods to perform data-analysis and extract the most relevant information (Figure 2). Also, while methods for imaging-based cellular analysis are well described (61) (e.g. as used in high content screening approaches of *in vitro* cultured cells), analysing images from tissue samples typically comes along with various additional challenges. Overall, analysing images

from multiplexed IHC consists of 2 major steps, including (i) image-analysis and (ii) high-dimensional, spatially resolved data analysis. Image analysis refers to the extraction of quantitative and meaningful information from the image by means of digital processing techniques and can be further divided in two fractions: low-level and high-level processing. In low-level image processing, a digital image is used as input and another digital image is obtained as output (e.g., a corrected image for improved visualization/analysis), while high-level processing involves functions whose outcome is a description of the content of the input image such as cell edges or tissue regions. The reader should note that upstream steps such as image-acquisition fall outside the scope of this review and are not covered here.

Low-Level Image Analysis

During low-level image analysis, all collected images are prepared for data extraction. While the procedures can be different between the various technology platforms, it generally shares a common strategy. Considering that the majority of technologies make use of fluorescent detection, we will primarily focus on these procedures.

When dealing with immunofluorescent images of tissue sections, these are often affected by aberrations which are critical to the quality of the results (62). Examples of these aberrations include out-of-focus regions (blurriness), vignetting effects, saturation debris, and artifacts due to e.g. air bubbles

and tissue folds. Various methods have been described to tackle some of these aberrations. Examples include ConvFocus (63) from Google AI or FQPath (64) for blurriness detection (focus), CIDRE (65) that uses an energy minimization model and a flat-field correction method developed by Kask and colleagues which uses an additive and a multiplicative component to correct for field-of-view artifacts (vignetting) (66), a protocol implemented in CellProfiler (67) which uses supervised machine learning for automatic quality control of image-based measurements, and HistoQC (62) which implements an automated, quantifiable, quality control process for identifying artifacts and measuring slide quality.

On top of undesired impurities, IF images are also subject to other sources of fluorescent signals such as background and tissue autofluorescence that need to be tackled before quantifying true signals. Examples of methods addressing this issue include dark pixel intensity identification (68) which estimates background signals by acquiring images at a different set of exposure times, and an autofluorescence removal method that uses non-negative matrix factorization that separates the signal into true signal and autofluorescence components (69). Moreover, high-throughput experiments can be subject to batch effects, that is, part of the acquired signal is described by undesired technical variation (for example, sample manipulation) rather than biological sources (61). An example of a method to tackle this issue is RESTORE (70) which identifies negative control cells for each marker and uses their

expression levels to normalize and remove sample-to-sample variation. Local/grid-based normalization tools (71) have also been developed, although their general implementation still requires a broader implementation.

In cyclic methods where images are acquired in consecutive rounds, the next step consists of the exact super-positioning of images acquired from consecutive imaging rounds (commonly referred to as ‘image registration (54)). Registration of medical images is a need that has been around for a while and thus a large number of methods are available. These methods can be broadly classified as rigid (image transformation is limited to translation, rotation, and scaling) (72, 73) and plastic (the moving image can be elastically deformed to best match the fixed image) (74). Historically, image registration was meeting a macroscopic need (for example, overlapping of MRI and Xray images). For multiplexing however, microscopic precision is required as a shift of a few pixels can overlap completely different cells. Therefore, recently, new registration methods have been published in the literature which aim to reach the required precision for exact-cell overlapping (75–78). While usually not required in “batch” methods (see above), this seemingly trivial step can be extremely computationally ‘expensive’ for large images (>100 million pixels) when applying plastic methods. Therefore, the latest algorithms come with GPU implementation. Also, depending on how images are acquired in the imager (e.g. some imagers scan all fluorescent channels at each step, while others scan the entire sample for

each channel separately), additional registration of the acquired fluorescent channels may be required.

Image pre-processing requires a very systematic approach and should be robust and identical across all samples/images of the same project to avoid the loss of biological information and a potentially biased interpretation of the results. While computational tools exist tackling the individual issues, their implementation for non-expert users is challenging and requires some level of computational knowledge. Moreover, linking the inputs and outputs of consecutive steps is far from trivial and integrated workflows are still missing. All of this together makes that to date, image pre-processing is mainly done by manual/visual inspection in the routine setting, thus largely limiting the throughput of these techniques.

High-Level Image Analysis

During high-level image analysis, the output of the different functions is a description of the content of the input image(s) such as cell edges, cell features or tissue regions. Once a set of images is properly aligned and ‘cleaned’, cells need to be precisely delineated (commonly described as ‘cell segmentation’). The quality of this step is very important for proper cell identification (see below). In tissue sections, this step is commonly done using nuclear segmentation. Historically nuclear segmentation has been tackled mechanistically (watershed algorithm and variations (79, 80). However, the introduction of deep learning algorithms (specifically

convolutional neural networks (CNNs) like uNets) have outperformed all the mechanistic algorithms and now represent the state-of-the-art (81, 82). The main drawback of deep-learning based algorithms is the need for large training datasets. However, current efforts are ongoing to implement machine learning approaches (83), which are gradually becoming more efficient at recognizing and splitting cells requiring minimal hands-on image training (84, 85). While a few algorithms have explored the possibility of segmenting full cells (including cytoplasm and membrane) (86, 87) cell nuclei are far less heterogeneous making nuclear segmentation more robust. Regarding the clinical implementation of cell segmentation approaches, deep-learning based methods will require to fine-tune existing algorithms to optimize the accuracy of the predictions and adjust it to the specific acquisition instrument and sample material. To the best of our knowledge, such an ambitious comparison in immunofluorescent images and in a multi-center setting with a variety of acquisition instruments is not available in the literature. However, a similar study has been carried out for Ki-67 expression, a prognostic marker in breast cancer, which has shown excellent accuracy and reproducibility in different analysis platforms performed by multiple operators (88). Accuracy and reproducibility are well known issues in state-of-the-art histopathology where the sample readouts are extracted by the subjective eye-rolling of an expert pathologist which causes intra- and inter-observer variability (89, 90). In fact, the main limitation for the automation (human or

computational) of pathological evaluations is the lack of standardization (91). However, we believe that with the right standardized roadmaps to histopathological analysis integrated computational pipelines will become state-of-the-art and that the future of pathology is mainly digital (92).

Finally, the acquired signals, such as fluorescence intensity, amount of metal isotopes, read-counts, etc., are quantified in every cell together with morphological features of the nucleus (nuclear size, shape, etc.) and topological features (X/Y coordinates) in a step called ‘feature extraction’, and collected in a structured data matrix.

Overall image pre-processing is highly dependent on the technology that was used to acquire the images – each of them will still require significant tweaks and adaptations to translate the procedures of one technology to another. While most vendors of the above described instruments provide accompanying software, these commonly don’t extend beyond mere viewing of images, a step that is crucial for quality assessments of the stains, but don’t allow quantitative and extended analysis. More dedicated (commercial) image analysis platforms are also gradually becoming available (e.g. Halo (Indica labs) or Visiopharm software packages for digital pathology, and the open source tools such as QuPath (93), CellProfiler (94), or HistoCat (95), are continuously updated, but considering the enormous amounts of biological questions that still require downstream analysis of the spatially resolved cell types, such packages typically remain constricted to initial

groundwork, or highly dedicated to one particular type of analysis. Some more integrative pipelines are also gradually being released, such as MCMIRCO (96) or SIMPLI (97), although these still require a high level of bioinformatics skills.

High-Dimensional, Spatially Resolved Data Analysis

Next, the obtained data matrix containing the high-dimensional data still requires further analysis and interpretation. A first step in this process involves the identification of the various cellular (pheno)types (i.e. epithelial cells, particular T cell subtypes, tumor cells, blood vessels, etc), a step that is usually done using clustering analysis and manual interpretation according to methods that resemble many other single cell methodologies. More automated algorithms that can be trained to assign labels to each cell (e.g. using convolutional neural networks (43)), and pretrained templates (98) do exist, but still require significant training, as cellular phenotypes can vary extensively between organs and/or disease conditions. In addition, this step is commonly further complicated by imperfect cell segmentation, as markers of adjacent cells can “pollute” neighboring cells, making interpretation difficult. This can require an interactive iteration of the settings for cell segmentation, next to a step-wise approach in clustering, where in a first step the major cell types are defined, after which the different subtypes of these cells can be defined. The latter is mainly used to avoid that large populations outcompete small/rare populations.

Regarding clustering, there are several methods publicly available without any single method proving to perform better than the others. While many studies have used one or another clustering algorithm (49, 99–101), our group implemented a *consensus* clustering procedure (31, 35) following a “Wisdom of the crowds” type of approach, where three independent algorithms (to choose between Phenograph, K-means, FlowSom, ClusterX, Clara, Hierarchical clustering, among various others) were used to cluster the identified cells. Each of the identified clusters is then annotated manually by an expert pathologist/immunologist, and only cells that remain in the same annotated cell type across minimal 2 methods are retained for further analysis. Since multiplexing datasets can identify millions of cells, clustering is done in a subset of these cells. There are different sampling strategies available, including random sampling, stratified random sampling, stratified proportional random sampling, etc. The annotated subset is then projected on the entire dataset using cell-type-specific fingerprints. The above-described methodology, however, still requires significant manual work from expert pathologists/immunologists, a step that could still largely benefit from properly curated databases, which will allow for machine learning tools to automatically recognize and annotate cell types.

Once this information has been gathered, data is subsequently projected against the spatial context of the tissue. Indeed, each cell is composed of a precise subset of pixels with specific

coordinates that span a specific surface in the image. This information therefore allows the analysis the spatial distribution of the cells that can be used to model cell-cell interactions (32) (typically referred to as 'neighborhood analysis') and any other spatial measurement (e.g. distance measurements to structures of interest, nearest-neighbor analysis, etc), but also the definition of larger cell communities and tissue architectures (102). For any of these analysis subtypes, in spite of the presence of several methods implemented in papers published in literature (102–106), there are currently very few software packages available, mostly requiring specific and specialized bioinformatics skills. However, from a biological point of view, this step is the most crucial as it will allow researchers to explore complex biological systems in a quantitative way, so sufficient efforts will still be required in the coming years to develop standardized methods. Moreover, the high complexity of these analyses, together with the relatively small scanned areas or amount of slides/patients (which cannot capture the full heterogeneity of a tissue or disease group), typically results in large patient-to-patient variation, which will require that significantly large cohorts are interrogated to come to statistically robust conclusions. Whether sufficiently large cohorts can be analyzed will in great part be defined by the used platform (see above). Also, diagnostic tools using mIHC/IF have shown their benefits for the treatment of the patient. Standardized methods to perform image analysis will have to

be developed and properly validated before their clinical implementation can even be considered.

Finally, each described method will eventually generate terabytes of images containing expression profiles of each included marker. The size of such data sets poses an additional challenge to store, label, transfer, computationally process and interpret the data in a reasonable amount of time. Current methods for big data processing, however, should be easy to adopt, but are only now becoming available in most institutes. In line with this, over the coming years, this “next-generation” pathology field will also need extended digital data repositories and standardized analysis methods to bring diagnostics and tissue interpretation to the next level. Implementing the required safety measures in line with the current General Data Protection Regulation (GDPR) regulations will also be of primary importance (107).

Making the difference with spatial IHC profiling

In the last part of this review, we provide an overview on how multiplexed IHC has been used and implemented in research over the past years. Considering the growing number of papers (all to which we can unfortunately not refer), we chose to focus on the most important concepts and provide examples of how researchers have used (some) of the above described methods together with dedicated bioinformatics analysis pipelines to gain additional insights in complex biological processes.

The General use of Multiplexed IHC Across Research

First, to get a top-level view on how multiplexed IHC has been used over the past years, we have performed an in-depth literature search, for which keywords and co-occurring terms were extracted from ~1000 papers from PUBMED using a query related to “multiplexed immunohistochemistry” (see *Methods*), and generated scientific networks. Based on the most occurring terms in the titles and abstracts of these papers, we can draw two main conclusions (Figure 3):

First, considering the topics of research, we identified 3 main topics: (i) technology/methodology development for multiplexed IHC using antibodies in tissue sections with the majority of methods focused on the usage of immunofluorescence; (ii) a strong focus on cancer research, with the aim to better predict prognosis, outcome and survival of patients based on the analysis of tissue sections; (iii) the vast majority of research papers used multiplexed IHC for the analysis of immune infiltrates, with a primary focus on T cell biology, checkpoint inhibition, macrophages and B cells. In addition, a more detailed density map also shows several prominent biological markers (e.g. CD3, CD4, CD8, CD68, CD163, CD20, PD1/PDL1, CTLA4, etc) that are generally used for T cell profiling, checkpoint analysis, and the investigation of macrophages and B-cells.

Second, when overlaying the generated network with the dates of when the papers containing the indicated terms were published, we identify a trend where research was initially

focused on technology development (before 2016), while gradually moving to its implementation in the research community to study immune infiltrates related to cancer development and its treatment using immunotherapy (2016-2020).

Using multiplexed IHC to answer complex biological questions

Considering the rise in available technologies for multiplexed IHC over the past few years and their gradual implementation, only the tip of the iceberg has been uncovered. Indeed, next to the revolution in single cell profiling, for which single-cell RNA (scRNA) sequencing is still the primary method for cellular phenotype exploration (108), the implementation of spatial technologies is becoming increasingly important to precisely locate the identified cell types and phenotypes in the context of a tissue and across a patient cohort. In this way, accurate atlases of healthy and disease-specific microenvironments can be compiled, both for human and mouse tissues (109, 110). For instance, the myeloid compartment of glioblastoma brain tumors was recently described using the combination of scRNAseq, CITEseq and multiplexed IHC through which the various ontologies of the infiltrating macrophages and resident microglial cells could be identified, while their distributions were spatially separated over various niches, which, moreover, evolved from the newly diagnosed to the recurrent setting (111). The distribution of the various cells across a tissue section was

determined by computationally breaking up the tissue in smaller “tiles” in which the relative distributions of the identified cell types were determined. The same approach is directly applicable to define tumoral, peritumoral, perivascular and non-tumoral areas (112), and, while these insights are key to define the macrostructure of a tissue, such analysis still remains on the level of one cell type at the time. Once macro-level areas are defined, it becomes key to understand local cellular distributions, an analysis type that is done by performing neighborhood or nearest neighbor analyses. In this analysis, the local neighborhood from each individual cell is defined by assessing which other cell types are present within a certain radius or distance. We and others have used this approach to define local cell-cell interactions between all identified immune cell populations in melanoma samples (23, 96). As such, it was found that exhausted cytotoxic CD8+ T cells were typically residing in close vicinity to TIM3- or PDL1-positive macrophages (23, 113, 114). Similarly, it was found that the CD8+/TCF7+ double positive T cells that were residing in the tumor could be reactivated and showed a positive correlation to responsiveness to checkpoint blockers in melanoma (115).

Such inference could only be achieved by performing multiplex analysis in tissue sections, allowing researchers to simultaneously define the phenotypic and functional status of each identified cell type but also define the spatial distribution of each cell and how they relate to each other. A similar cellular sociology was identified in breast tumors using IMC through

which *communities* and local niches could be determined which typically consisted of a dominant tumor clone in combination with a variety of immune cells (102). Cellular neighborhoods could also be identified in colorectal tumors (116). Moreover, by measuring such neighborhoods, the presence of particular CD4+ T cells in cellular neighborhoods that were also enriched for granulocytes was identified as a positive markers for survival in CRC, while the presence of macrophages in breast tumor communities rather correlated with poor prognosis (102, 116). Also in lung, ovarian, liver, kidney and various other cancer types, the presence of specific immune cell infiltrates were shown to correlate with good or poor outcomes (117–119) or with response to immunotherapy (120). The number of research papers using this type of analysis keeps on rising steeply (and we apologize to all the authors of papers we have not mentioned here), even though the scale at which such projects can be performed should become even larger to achieve the next clinical revolution.

Multiplexed IHC vs. spatial transcriptomics: a complementary duo

In this review, we have mainly focused on methods for multiplexed IHC. The main advantage of these methods resides in their prompt translatability to the clinical setting: indeed, pathology labs have been performing antibody-based assessments for decades and adding mIHC should be more easily adoptable in such setting, even though a transition

towards digital pathology will be required. Recently, methods for spatial transcriptomics are being developed at rapid pace: the main advantage of the latter methods (for which a plethora of technical approaches is currently available (121)) is their more universal character, whereby complementary detection probes can be developed in a more generic and species independent way, while mIHC largely depends on the availability of specific, high quality antibodies, which are not always available and often not cross-reactive over species (e.g. mouse vs human). While such approach enables researchers to uncover unknown pathways and patterns, RNA-based methods also harbor some pitfalls as well. Indeed, for the majority of (archival) samples, the currently available standard quality of materials (as they are mostly available as FFPE materials in biobanks across the world, and less often in the form of fresh frozen tissue blocks) is typically directly amenable for antibody-based approaches leading to robust insights. The quality of samples for RNA based analyses will, on the other hand, require very close monitoring of sample/RNA quality, as the stability and longevity of RNA molecules is much less compared to proteins. The translational character of RNA-based methods will therefore still require proper benchmarking to ensure robust pathology grade readouts. The coming months and years will have to show how this exciting field keeps on evolving.

Finally, both protein and RNA-based methods are gaining more and more traction to unravel complex cellular networks and architectures. However, the main setting currently still relies on

the “discrete” utilization of one method at the time. Indeed, while each method can already provide highly valuable insights at single-cell resolution, each approach also harbors various downsides as well. For instance, while transcriptome analysis is excellent at unravelling transcription factor networks or identify the source of cytokine expression in complex tissues (122), signal-transduction events, protein-protein interactions or particular immune cell states can be measured more reliably at the protein level (123). As such, combining multiple methodologies in an orchestrated fashion can produce highly synergistic insights that cannot be achieved by either method alone. This evolution towards multi-omics approaches is a very active domain, where various challenges will have to be overcome as well. Not only will methods have to be adapted in such way that capturing RNA and protein based features will remain possible, also data integration will require further evolution, even though steps are being put in that direction (121, 122).

Conclusion

A large variety of technologies for multiplexed immunohistochemistry has been developed over the past years, each with their own advantages and downsides, and it is expected that these will keep on improving over the coming years. Moreover, combinations with other spatial omics, such spatial transcriptomics, genomics, chromatin accessibility, lipidomics or metabolomics, are underway which bring yet

another level of technological and computational challenges. Regardless of the approach, each technology has the goal to interrogate increasing numbers of analytes in pathological tissue samples at single-cell and spatial resolution. This revolution allows us now to investigate complex patho-biological processes at unseen resolution- insights that bear the potential of becoming the next generation of higher order biomarkers. This can, however, only be achieved if appropriate computational tools and infrastructure – the hallmark of a true shift towards digital pathology – are implemented that can deal with this approach and complexity. Moreover, considering the number of parameters that will be measured, it will be paramount to investigate sufficiently large populations of patients so that we can evolve from anecdotical case reports to more fundamental, robust and clinically useful insights. The latter will require the combination of highly standardized methodological processes and concurrent validated analysis pipelines. Finally, such next-generation pathology will also require standardized digital data repositories to set appropriate standards and benchmarks to bring the field to the next level.

Methods

Digital Literature Analysis

The VOSViewer tool and algorithms (Visualizing scientific landscapes; https://www.vosviewer.com (124);) were used to analyse extracted publication data from a PubMed search using the following search criteria:

multiplex immunohistochemistry[Title/Abstract]) OR
(multiplexed immunohistochemistry[Title/Abstract]) OR
(multiplexed immunofluorescence[Title/Abstract]) OR
(multiplexed immunophenotyping[Title/Abstract]) OR
(multiplex immunofluorescence[Title/Abstract]) OR
(multiplex immunophenotyping[Title/Abstract])

Multiplexed Analysis using MILAN

A 3µm thick tissue section of a melanoma tissue microarray was subjected to multiplexed immunohistochemistry according to the MILAN protocol as previously described (21). Overall 83 antibodies were succesfully used as have been described here (23, 125).

Author Contributions

FS and FB conceptualized the manuscript. All authors contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by KULeuven funding (C3/19/053, C1/17/084), the Opening the Future foundation, the Kom op tegen Kanker (Stand up to Cancer) foundation, FWO infrastructure grant (I005920N) and an FWO Fundamenteel Klinisch Mandaat (EMH-D8972-FKM/20) to FB.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

The Supplementary Material for this article can be found online at:https://www.frontiersin.org/articles/10.3389/fonc.2022.918900/full#supplementary-material

Supplementary Figure 1 | GIF representation covering an 82-plex of a melanoma tissue sample stained with 52 consecutive rounds of MILAN staining.

Figure 1 Schematic overview of the currently available methods for multiplexed immunohistochemistry (IHC). **(A)** Currently, the most common approach for multiplexed IHC makes use of fluorescently labelled probes, which are either directly coupled to the primary antibody or indirectly provided by a secondary antibody, that are detected in a cyclic fashion consisting of a staining protocol, followed by tissue imaging and signal removal. **(B)** In contrast to cyclic methods, single-step spectral methods detect all dyes in the tissue simultaneously: these can either be provided by directly labelled antibodies that are all simultaneously present in the tissue section or by the cyclic generation of TSA precipitates which are subsequently spectrally unmixed in a single imaging step. **(C)** Antibodies can

also be detected by covalently linked nucleotide labels to which fluorescently labelled probes are hybridized in a cyclic fashion for which each cycle gets imaged. **(D)** Non-fluorescent mIHC methods involve the cyclic generation of chromogenic substrates that are washed away following an imaging step in between each cycle. **(E)** For imaging mass cytometry (IMC), antibodies are labelled with metal isotopes which are detected by the local vaporization of the metal ions by a UV laser, following which the present isotopes are resolved using atomic spectrometry. **(F)** Finally, nucleotide labelled antibodies can be detected by removing the nucleotide labels from the antibodies using a laser beam, following which the nucleotides that were collected from a precise region of interest are sequenced to quantify the amount of available proteins in that region.

Figure 2 Schematic overview of the required steps for downstream image analysis using the most commonly used fluorescent, cyclic methods for multiplexed IHC. Images are collected across multiple cycles but still need to be cleaned (QC), corrected (PP), registered/aligned (REG), autofluorescence removed (AF), segmented (SEG), feature extracted (FE), phenotypically annotated (PI), and spatially resolved (SA).

Figure 3 The bibliometric map of multiplexed IHC over the years. Using VOS Viewer, a software tool for constructing and visualizing the bibliometric network related to “multiplexed

immunohistochemistry" on PUBMED (see *Methods*), we observed a shift from 2010 where technology development started (blue circles), to its use to unravel complex cellular networks in 2020 (yellow areas) with a strong focus on immuno-oncology and T-cell biology.

Table 1 Technical overview of multiplex IHC staining methodologies.

Figure 1

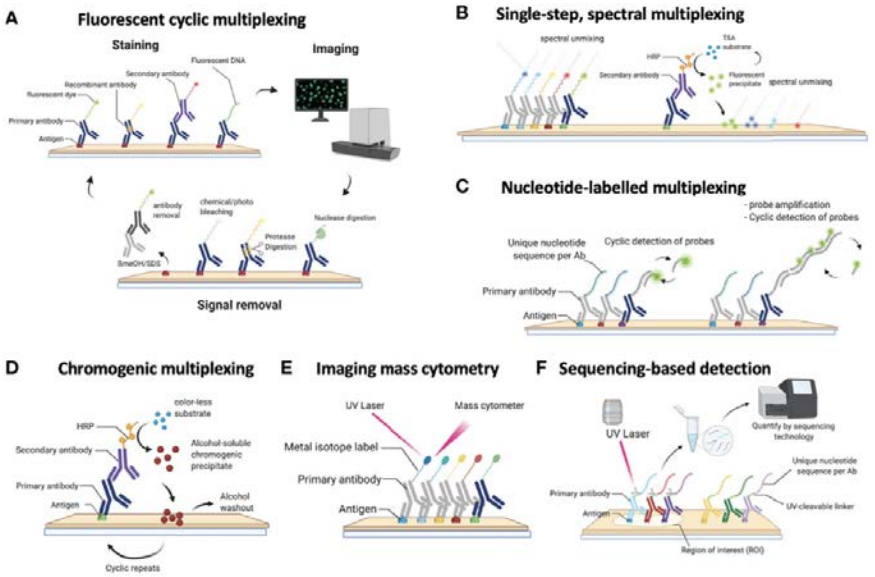


Figure2

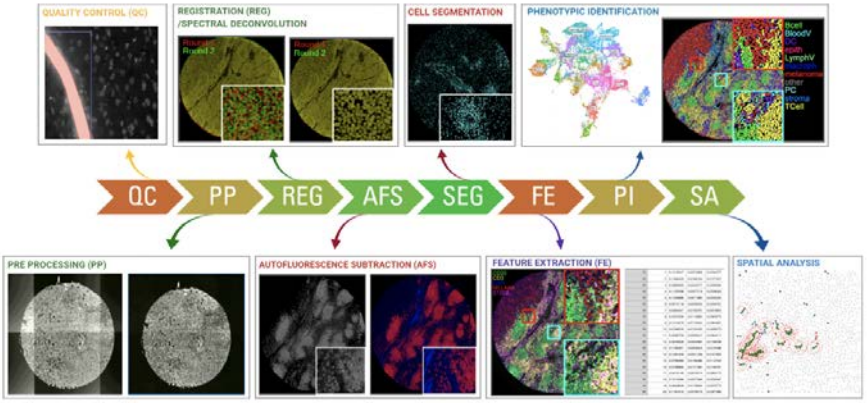


Figure 3

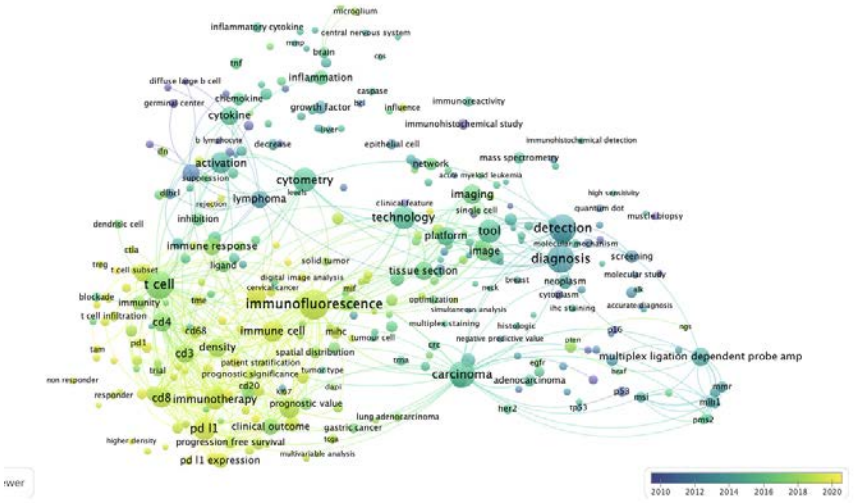


Table 1

Method Name	Commercial Name	Sample	Max. n of Markers	Flexibility to implement new markers[a]	Primary Ab	Ab tag	Signal removal technique	Tissue size	Automation	n of slides/ experiment	Time (for 40-plex staining of 1 large slide)	Resolution	Tissue preservation	Current Scientific use*	Refs
MILAN	NA	FFPE	82	High	Unconj. and Conj.	FL sec ab	Ab stripping	WS/TMA	Partially automated	30+	2-3 weeks	subcellular	Yes	Mature	(21–23)
CyclF	NA	FFPE	60	High	Conj.	FL	bleaching	WS/TMA	Partially automated	30+	2-3 weeks	subcellular	Yes	Mature	(24)
MICS	MaxSima™	FFPE/Frozen	400	Limited	Conj.	FL	bleaching/ fluorochrome release	WS/TMA	Fully automated	2	<1 week	subcellular	No	Limited	(25)
sequIF™	COMET™ / Labsat™	FFPE	40	High	Unconj. and Conj.	FL sec ab	Ab stripping	TMA/up to 2.2x2.2cm section	fully automated	1-4	1-2 days	subcellular	Yes	Limited	(26)
Chipcytometry	CellScape™ / ZeissScanner ONE™	Frozen (FFPE not recommended)	65	High	Conj.	FL	bleaching	TMA/2x1 cm section	Service/Fully automated	NA	<1 week	subcellular	Yes	Mature	(27)
IMC™	Hyperion	FFPE/Frozen/Liquid biopsies	37	Limited	Conj.	Metal	None	WS+ROI/ TMA/slow	Service/automated detection	1	2 weeks	subcellular, 1 µm/px	No	Mature	(28)
MIBI™	IONpath/ MIBIscope™	Frozen	40+	Limited	Conj.	Metal	None	TMA/slow (ROIs 800 x 800 µm²)	automated detection	1	2 weeks	subcellular, 650 nm/px	No	Mature	(29)
NA	CODEX®	Frozen/FFPE	40+	Limited	Conj.	NUC	None	WS/TMA	automated detection	1	<1 week	subcellular	No	Mature	(30–32)
DSP	NanoString's GeoMx® Digital Spatial Profiler (DSP)	FFPE	96	Limited	Conj.	NUC	None	ROI	automated, sequencing still required	1	<1 day	Subcellular/ grid	No	Mature	(33)
OPAL	Akoya's Opal™ Multiplex IHC +/- Vectra® Polaris™ spectral scanner	FFPE	9	High	Conj.	HRP-TSA	Heat based Ab stripping (ER)	WS/TMA	Partially automated	30+	<1 day (for 9-plex)	subcellular	No	Mature	(34)
MxIF	Cell Dive (Leica)	FFPE	61	High	Conj. and unconj.	FL	Chemical inactivation	WS/TMA	Partially automated	30+	2-3 weeks	subcellular	Yes	Method published	(35)
cmIF	NA	FFPE	60	High	Conj.	FL	Chemical inactivation	WS/TMA	Partially automated	30+	2-3 weeks	subcellular	Yes		(36)
MELC	NA	Frozen	50+	High	Conj.	FL	bleaching	WS/TMA	Partially automated	1	2-3 weeks	subcellular	Yes	Mature	(37)
IBEX	NA	Frozen/FFPE	65+	High	Conj.	FL	Chemical inactivation	WS/TMA	Partially automated	30+	1-2 weeks	subcellular	Yes	Method published	(38)
NA	Orion RareCyte® platform	Frozen/FFPE	21	High	Conj.	FL	None	WS/TMA	Fully automated	NR	<1 day (for 6-plex)	subcellular	No	Method published	(39)
S4BER	ImmunoSABER	FFPE/Frozen/Liquid biopsies	10	Limited	Conj.	NUC	None	WS/TMA	Partially automated	1	<1 day (for 10-plex)	subcellular	No	Method published	(40)
Ultivue	NA	FFPE/Frozen	8	Limited	Conj.	NUC	None	WS/TMA	partially automated	30+	<1 day (for 10-plex)	subcellular	Yes	Mature	(41)
SIMPLE	NA	FFPE	5+	High	Unconj.	HRP	Chemical Ab stripping	WS/TMA	Partially automated	30+	<1 day (for 5-plex)	subcellular	Yes	Method published	(42)
MICSSS	NA	FFPE	10	High	Unconj.	HRP	Chemical destaining + Ab blocking	WS/TMA	Partially automated	30+	<1 week (for 10-plex)	subcellular	Yes	Method published	(43–44)

NA, Not applicable; WS, whole slide; TMA, tissue microarray; FL, fluorescent; FFPE, formalin fixed paraffin embedded; Conj., conjugated; Ab, antibody; ROI, region of interest.

https://www.frontiersin.org/files/Articles/918900/fonc-12-918900-HTML-r1/image_m/fonc-12-918900-t001.jpg

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial Sequencing and Analysis of the Human Genome. *Nature* (2001) 409:860–921. doi: 10.1038/35057062

2. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. *Elife* (2017) 6:e27041. doi: 10.7554/eLife.27041

3. Rajewsky N, Almouzni G, Gorski SA, Aerts S, Amit I, Bertero MG, et al. LifeTime and Improving European Healthcare Through Cell-Based Interceptive Medicine. *Nature* (2020) 587:377–86. doi: 10.1038/s41586-020-2715-9

4. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, Rood JE, et al. The Human Tumor Atlas Network: Charting Tumor Transitions Across Space and Time at Single-Cell Resolution. *Cell* (2020) 181:236–49. doi: 10.1016/j.cell.2020.03.053

5. HuBMAP Consortium. The Human Body at Cellular Resolution: The NIH Human Biomolecular Atlas Program. *Nature* (2019) 574:187–92. doi: 10.1038/s41586-019-1629-x

6. Lee J, Hyeon DY, Hwang D. Single-Cell Multiomics: Technologies and Data Analysis Methods. *Exp Mol Med* (2020) 52:1428–42. doi: 10.1038/s12276-020-0420-2

7. Stuart T, Satija R. Integrative Single-Cell Analysis. *Nat Rev Genet* (2019) 20:257–72. doi: 10.1038/s41576-019-0093-7

8. Allam M, Cai S, Coskun AF. Multiplex Bioimaging of Single-Cell Spatial Profiles for Precision Cancer Diagnostics and Therapeutics. *NPJ Precis Oncol* (2020) 4:1–14. doi: 10.1038/s41698-020-0114-1

9. Liao J, Lu X, Shao X, Zhu L, Fan X. Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol* (2021) 39:43–58. doi: 10.1016/j.tibtech.2020.05.006

10. Blom S, Paavolainen L, Bychkov D, Turkki R, Mäki-Teeri P, Hemmes A, et al. Systems Pathology by Multiplexed Immunohistochemistry and Whole- Slide Digital Image Analysis. *Sci Rep* (2017) 7(1):15580. doi: 10.1038/s41598-017-15798-4

11. Thornton CA, Mulqueen RM, Torkenczy KA, Nishida A, Lowenstein EG, Fields AJ, et al. Spatially Mapped Single-Cell Chromatin Accessibility. *Nat Commun 2021 121* (2021) 12:1–16. doi: 10.1038/s41467-021-21515-7

12. Zhao T, Chiang ZD, Morriss JW, LaFave LM, Murray EM, Del Priore I, et al. Spatial Genomics Enables Multi-Modal Study of Clonal Heterogeneity in Tissues. *Nat 2021 6017891* (2021) 601:85–91. doi: 10.1038/s41586-021-04217-4

13. Van Den Tweel JG, Taylor CR. A Brief History of Pathology Preface to a Forthcoming Series That Highlights Milestones in the Evolution of Pathology as a Discipline. *Virchows Arch* (2010) 457:3–10. doi: 10.1007/s00428-010-0934-4

14. Kim SW, Roh J, Park CS. Immunohistochemistry for Pathologists: Protocols, Pitfalls, and Tips. *J Pathol Transl Med* (2016) 50:411–8. doi: 10.4132/jptm.2016.08.08

15. Bolognesi MM, Manzoni M, Scalia CR, Zannella S, Bosisio FM, Faretta M, et al. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *J Histochem Cytochem* (2017) 65:431–44. doi: 10.1369/0022155417719419

16. Nielsen S. External Quality Assessment for Immunohistochemistry: Experiences From NordiQC. *Biotech Histochem* (2015) 90:331–40. doi: 10.3109/10520295.2015.1033462

17. Mascadri F, Bolognesi MM, Pilla D, Cattoretti G. Rejuvenated Vintage Tissue Sections Highlight Individual Antigen Fate During Processing and Long-Term Storage. *J Histochem Cytochem* (2021) 69:659. doi: 10.1369/00221554211047287

18. MM B, FM B. & G, C. Unidentified Variables May Account for Variability in Multiplexing Results. *J Histochem Cytochem* (2020) 68:351–3. doi: 10.1369/0022155420911050

19. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-Based Map of the Human Proteome. *Sci* (2015) 347(6220):1260419. doi: 10.1126/science.1260419

20. Hickey JW, Neumann EK, Radtke AJ, Camarillo JM, Beuschel RT, Albanese A, et al. Spatial Mapping of Protein Composition and Tissue Organization: A Primer for Multiplexed Antibody-Based Imaging. *Nat Methods* (2022) 19:284–95. doi: 10.1038/s41592-021-01316-y

21. Cattoretti C, Bosisio FM, Marcelis L, Bolognesi MM. Multiple Iterative Labeling by Antibody Neodeposition (MILAN) *PROTOCOL (Version 5) available at Research Square*. (2018). doi: 10.21203/rs.2.1646/v5

22. De Smet F, Antoranz Martinez A, Bosisio FM. Next-Generation Pathology by Multiplexed Immunohistochemistry. *Trends Biochem Sci* (2021) 46:80–2. doi: 10.1016/j.tibs.2020.09.009

23. Bosisio FM, Antoranz A, van Herck Y, Bolognesi MM, Marcelis L, Chinello C, et al. Functional Heterogeneity of Lymphocytic Patterns in Primary Melanoma Dissected Through Single-Cell Multiplexing. *Elife* (2020) 9:e53008. doi: 10.7554/eLife.53008

24. Lin J-R, Fallahi-Sichani M, Sorger PK. Highly Multiplexed Imaging of Single Cells Using a High-Throughput Cyclic Immunofluorescence Method. *Nat Commun* (2015) 6:8390. doi: 10.1038/ncomms9390

25. Kinkhabwala A, Herbel C, Pankratz J, Yushchenko DA, Rüberg S, Praveen P, et al. MACSima Imaging Cyclic Staining (MICS) Technology Reveals Combinatorial Target Pairs for CAR T Treatment of Solid Tumors. *Sci Rep* (2022) 12(1):1911. doi: 10.1038/s41598-022-05841-4.

26. Migliozzi D, Pelz B, Dupouy DG, Leblond AL, Soltermann A, Gijs MAM. Microfluidics-Assisted Multiplexed Biomarker Detection for *in Situ* Mapping of Immune Cells in Tumor Sections. *Microsystems Nanoeng* (2019) 5:1–12. doi: 10.1038/s41378-019-0104-z

27. Consentius C, Mirenska A, Jurisch A, Reinke S, Scharm M, Zenclussen AC, et al. *In Situ* Detection of CD73+ CD90+ CD105+ Lineage: Mesenchymal Stromal Cells in Human Placenta and Bone Marrow Specimens by Chipcytometry. *Cytom Part A* (2018) 93:889–93. doi: 10.1002/cyto.a.23509

28. Giesen C, Wang HA, Schapiro D, Zivanovic N, Jacobs A, Hattendorf B, et al. Highly Multiplexed Imaging of Tumor Tissues With Subcellular Resolution by Mass Cytometry. *Nat Methods* (2014) 11:417–22. doi: 10.1038/nmeth.2869

29. Baharlou H, Canete NP, Cunningham AL, Harman AN, Patrick E. Mass Cytometry Imaging for the Study of Human Diseases-Applications and Data Analysis Strategies. *Front Immunol* (2019) 10:2657. doi: 10.3389/fimmu.2019.02657

30. Dakshinamoorthy G, Singh J, Kim J, Nikulina N, Bashier R, Mistry S, et al. Abstract 490: Highly Multiplexed Single-Cell Spatial Analysis of Tissue Specimens Using CODEX. *Cancer Res* (2019) 79(13_Supplement):490. doi: 10.1158/1538-7445.AM2019-490

31. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture With CODEX Multiplexed Imaging. *Cell* (2018) 174:968–981.e15. doi: 10.1016/j.cell.2018.07.010

32. Black S, Phillips D, Hickey JW, Kennedy-Darling J, Venkataraaman VG, Samusik N, et al. CODEX Multiplexed Tissue Imaging With DNA-Conjugated Antibodies. *Nat Protoc* (2021) 16:3802–35. doi: 10.1038/s41596-021-00556-8

33. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex Digital Spatial Profiling of Proteins and RNA in Fixed Tissue. *Nat Biotechnol* (2020) 38:586–99. doi: 10.1038/s41587-020-0472-9

34. Parra ER, Uraoka N, Jiang M, Cook P, Gibbons D, Forget MA, et al. Validation of Multiplex Immunofluorescence Panels Using Multispectral Microscopy for Immune-Profiling of Formalin-Fixed and Paraffin-Embedded Human Tumor Tissues. *Sci Rep* (2017) 7:1–11. doi: 10.1038/s41598-017-13942-8

35. Gerdes MJ, et al. Highly Multiplexed Single-Cell Analysis of Formalinfixed, Paraffin-Embedded Cancer Tissue. *Proc Natl Acad Sci U. S. A.* (2013) 110:11982–7. doi: 10.1073/pnas.1300136110

36. Eng J, et al. Cyclic Multiplexed-Immunofluorescence (cmIF), a Highly Multiplexed Method for Single-Cell Analysis. *Methods Mol Biol* (2020) 2055:521–62. doi: 10.1007/978-1-4939-9773-2_24

37. Schubert W, et al. Analyzing Proteome Topology and Function by Automated Multidimensional Fluorescence Microscopy. *Nat Biotechnol* (2006) 24:1270–8. doi: 10.1038/nbt1250

38. Radtke AJ, Kandov E, Lowekamp B, Speranza E, Chu CJ, Gola A, et al. IBEX-A Versatile Multi-Plex Optical Imaging Approach for Deep Phenotyping and Spatial 2 Analysis of Cells in Complex Tissues 3 High Dimensional Imaging, Immune System, Quantitative Microscopy, Tissue Immunity. *Proc Natl*

Acad Sci U S A (2020) 117(52):33455–33465. doi: 10.1073/pnas.2018488117

39. EP K, et al. The RareCyte® Platform for Next-Generation Analysis of Circulating Tumor Cells. *Cytometry A* (2018) 93:1220–5. doi: 10.1002/cyto.a.23619

40. Saka SK, et al. Immuno-SABER Enables Highly Multiplexed and Amplified Protein Imaging in Tissues. *Nat Biotechnol* (2019) 37:1080–90. doi: 10.1038/s41587-019-0207-y

41. Manesse M, Patel KK, Bobrow M, Downing SR. The InSituPlex ® Staining Method for Multiplexed Immunofluorescence Cell Phenotyping and Spatial Profiling of Tumor FFPE Samples. *Methods Mol Biol* (2020) 2055:585–92. doi: 10.1007/978-1-4939-9773-2_26

42. Glass G, Papin JA, Mandell JW. SIMPLE: A Sequential Immunoperoxidase Labeling and Erasing Method. *J Histochem Cytochem* (2009) 57:899–905. doi: 10.1369/jhc.2009.953612

43. Remark R, Merghoub T, Grabe N, Litjens G, Damotte D, Wolchok JD, et al. In-Depth Tissue Profiling Using Multiplexed Immunohistochemical Consecutive Staining on Single Slide. *Sci Immunol* (2016) 1(1):aaf6925. doi: 10.1126/sciimmunol.aaf6925

44. Akturk G, Sweeney R, Remark R, Merad M, Gnjatic S. Multiplexed Immunohistochemical Consecutive Staining on Single Slide (MICSSS): Multiplexed Chromogenic IHC Assay for High-Dimensional Tissue Analysis. *Methods Mol Biol* (2020) 2055:497–519. doi: 10.1007/978-1-4939-9773-2_23

45. Sun H, Shen J. & Yokota, H. Size-Dependent Positioning of Human Chromosomes in Interphase Nuclei. *Biophys J* (2000) 79:184–90. doi: 10.1016/S0006-3495(00)76282-5

46. Monici M. Cell and Tissue Autofluorescence Research and Diagnostic Applications. *Biotechnol Annu Rev* (2005) 11:227–56. doi: 10.1016/S1387-2656(05)11007-2

47. Lin JR, Izar B, Wang S, Yapp C, Mei S, Shah PM, et al. Highly Multiplexed Immunofluorescence Imaging of Human Tissues and Tumors Using T-CyCIF and Conventional Optical Microscopes. *Elife* (2018) 7:e31657. doi: 10.7554/eLife.31657

48. Pascual-Reguant A, et al. Multiplexed Histology Analyses for the Phenotypic and Spatial Characterization of Human Innate Lymphoid Cells. *Nat Commun* (2021) 12:1–15. doi: 10.1038/s41467-021-21994-8

49. Sood A, et al. Multiplexed Immunofluorescence Delineates Proteomic Cancer Cell States Associated With Metabolism. *JCI Insight* (2016) 1:87030. doi: 10.1172/jci.insight.87030

50. Mani NL, et al. Quantitative Assessment of the Spatial Heterogeneity of Tumor-Infiltrating Lymphocytes in Breast Cancer. *Breast Cancer Res 2016 181* (2016) 18:1–10. doi: 10.1186/s13058-016-0737-x

51. Rubin MA, Dunn R, Strawderman M, Pienta KJ. Tissue Microarray Sampling Strategy for Prostate Cancer Biomarker Analysis. *Am J Surg Pathol* (2002) 26:312–9. doi: 10.1097/00000478-200203000-00004

52. RL C, LA C, DL R. Validation of Tissue Microarray Technology in Breast Carcinoma. *Lab Invest* (2000) 80:1943–9. doi: 10.1038/labinvest.3780204

53. RL C. V, N. & DL, R. A Decade of Tissue Microarrays: Progress in the Discovery and Validation of Cancer Biomarkers. *J Clin Oncol* (2008) 26:5630–7. doi: 10.1200/JCO.2008.17.3567

54. Levin M, Flor AC, Snyder H, Kron SJ, Schwartz D. UltraPlex Hapten-Based Multiplexed Fluorescent Immunohistochemistry. *Methods Mol Biol* (2021) 2350:267–87. doi: 10.1007/978-1-0716-1593-5_17

55. Rajagopalan A, Venkatesh I, Aslam R, Kirchenbuechler D, Khanna S, Cimbaluk D, et al. SeqStain Using Fluorescent-DNA Conjugated Antibodies Allows Efficient, Multiplexed, Spatialomic Profiling of Human and Murine Tissues. *Cell Rep Methods*. (2021) 1(2):100006. doi: 10.1016/j.crmeth.2021.100006

56. Taube JM, et al. The Society for Immunotherapy in Cancer Statement on Best Practices for Multiplex Immunohistochemistry (IHC) and Immunofluorescence (IF) Staining and Validation. *J Immunother Cancer* (2020) 8:155. doi: 10.1136/jitc-2019-000155

57. Molne J, Breimer ME, Svalander CT. Immunoperoxidase Versus Immunofluorescence in the Assessment of Human Renal Biopsies. *Am J Kidney Dis* (2005) 45:674–83. doi: 10.1053/j.ajkd.2004.12.019

58. LL W, et al. Comparison of C3d Immunohistochemical Staining to Enzyme-Linked Immunosorbent Assay and

Immunofluorescence for Diagnosis of Bullous Pemphigoid. *J Am Acad Dermatol* (2020) 83:172–8. doi: 10.1016/j.jaad.2020.02.020

59. Hartmann FJ, et al. Scalable Conjugation and Characterization of Immunoglobulins With Stable Mass Isotope Reporters for Single-Cell Mass Cytometry Analysis. in. *Methods Mol Biol* (2019) 1989:55–81. doi: 10.1007/978-1-4939-9454-0_5

60. Bergholtz H, Carter JM, Cesano A, Cheang MCU, Church SE, Divakar P, et al. Best Practices for Spatial Profiling for Breast Cancer Research With the GeoMx ® Digital Spatial Profiler. *Cancers (Basel)* (2021) 13(17):4456. doi: 10.3390/cancers13174456

61. Caicedo JC, et al. Data-Analysis Strategies for Image-Based Cell Profiling. *Nat Methods* (2017) 14:849–63. doi: 10.1038/nmeth.4397

62. Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. *JCO Clin Cancer Inf* (2019) 3:1–7. doi: 10.1200/CCI.18.00157

63. Kohlberger T, Liu Y, Moran M, Chen PC, Brown T, Hipp JD, et al. Whole-Slide Image Focus Quality: Automatic Assessment and Impact on AI Cancer Detection. *J Pathol Inform* (2019) 10:39. doi: 10.4103/jpi.jpi_11_19

64. Hosseini MS, Brawley-Hayes JAZ, Zhang Y, Chan L, Plataniotis K, Damaskinos S. Focus Quality Assessment of High-Throughput Whole Slide Imaging in Digital Pathology.

IEEE Trans Med Imaging (2020) 39(1):62–74. doi: 10.1109/TMI.2019.2919722.

65. Smith K, et al. CIDRE: An Illumination-Correction Method for Optical Microscopy. *Nat Methods* (2015) 12:404–6. doi: 10.1038/nmeth.3323

66. Kask P, Palo K, Hinnah C, Pommerencke T. Flat Field Correction for High-Throughput Imaging of Fluorescent Samples. *J Microsc* (2016) 263:328–40. doi: 10.1111/jmi.12404

67. Bray M-A, Carpenter AE. Quality Control for High-Throughput Imaging Experiments Using Machine Learning in Cellprofiler. *Methods Mol Biol* (2018) 1683:89–112. doi: 10.1007/978-1-4939-7357-6_7

68. Pang Z, Laplante NE, Filkins RJ. Dark Pixel Intensity Determination and its Applications in Normalizing Different Exposure Time and Autofluorescence Removal. *J Microsc* (2012) 246:1–10. doi: 10.1111/j.1365-2818.2011.03581.x

69. Woolfe F, Gerdes M, Bello M, Tao X, Can A. Autofluorescence Removal by non-Negative Matrix Factorization. *IEEE Trans Image Process* (2011) 20:1085–93. doi: 10.1109/TIP.2010.2079810

70. Chang YH, et al. RESTORE: Robust Intensity Normalization Method for Multiplexed Imaging. *Commun Biol* (2020) 3:1–9. doi: 10.1038/s42003-020-0828-1

71. Graf J, et al. FLINO: A New Method for Immunofluorescence Bioimage Normalization. *Bioinformatics* (2022) 38:520–6. doi: 10.1093/bioinformatics/btab686

72. Srinivasa Reddy B, Chatterji BN. An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration. *IEEE Trans Image Process* (1996) 5:1266–71. doi: 10.1109/83.506761

73. Matungka R, Zheng YF, Ewing RL. Image Registration Using Adaptive Polar Transform. *IEEE Trans Image Process* (2009) 18:2340–54. doi: 10.1109/TIP.2009.2025010

74. Maintz JBA, Viergever MA. An Overview of Medical Image Registration Methods. *Imaging Science Department, Imaging Center Utrecht, Technical Report* (1996). UU-CS-1998–22

75. Ri Y, Fujimoto H. Drift-Free Motion Estimation From Video Images Using Phase Correlation and Linear Optimization. *Proc - 2018 IEEE 15th Int Work. Adv Motion Control (AMC)* (2018), 295–300. doi: 10.1109/AMC.2019.8371106

76. Muhlich J, Chen Y-A, Russell D, Sorger PK. Stitching and Registering Highly Multiplexed Whole Slide Images of Tissues and Tumors Using ASHLAR Software. *bioRxiv* (2021), 440625. doi: 10.1101/2021.04.20.440625

77. Beroiz M, Cabral JB, Sanchez B. Astroalign: A Python Module for Astronomical Image Registration. *Astron Comput* (2020) 32:100384. doi: 10.1016/j.ascom.2020.100384

78. Sandkühler R, Jud C, Andermatt S, Cattin PC. AirLab: Autograd Image Registration Laboratory. *arXiv* (2018) preprint arXiv:1806.09907.

79. Beucher S, Meyer F. *79*. p. (2018). doi: 10.1201/9781482277234-12

80. Gamarra M, Zurek E, Escalante HJ, Hurtado L, San-Juan-Vergara H. Split and Merge Watershed: A Two-Step Method for Cell Segmentation in Fluorescence Microscopy Images. *Biomed Signal Process Control* (2019) 53:101575. doi: 10.1016/j.bspc.2019.101575

81. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28

82. Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection With Star-Convex Polygons. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G (eds) *Medical Image Computing and Computer Assisted Intervention – MICCAI* (2018). MICCAI 2018. Lecture Notes in Computer Science(), vol 11071. Springer, Cham. doi: 10.1007/978-3-030-00934-2_30

83. Korfhage N, Mühling M, Ringshandl S, Becker A, Schmeck B, Freisleben B. Detection and Segmentation of Morphologically Complex Eukaryotic Cells in Fluorescence Microscopy Images *via* Feature Pyramid Fusion. *PloS Comput Biol* (2020) 16(9):e1008179. doi: 10.1371/journal.pcbi.1008179

84. Schwendy M, Unger RE, Parekh SH. EVICAN—a Balanced Dataset for Algorithm Development in Cell and Nucleus Segmentation. *Bioinformatics* (2020) 36:3863–70. doi: 10.1093/bioinformatics/btaa225

85. Kumar N, et al. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Trans Med Imaging* (2017) 36:1550–60. doi: 10.1109/TMI.2017.2677499

86. Al-Kofahi Y, Zaltsman A, Graves R, Marshall W, Rusu M. A Deep Learning-Based Algorithm for 2-D Cell Segmentation in Microscopy Images. *BMC Bioinf* (2018) 19:365. doi: 10.1186/s12859-018-2375-z

87. Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-Cell Segmentation of Tissue Images With Human-Level Performance Using Large-Scale Data Annotation and Deep Learning. *Nat Biotechnol* (2022) 40(4):555–565. doi: 10.1038/s41587-021-01094-0

88. Acs B, et al. Ki67 Reproducibility Using Digital Image Analysis: An Inter-Platform and Inter-Operator Study. *Lab Investig* (2018) 99:107–17. doi: 10.1038/s41374-018-0123-7

89. Franc B, et al. Interobserver and Intraobserver Reproducibility in the Histopathology of Follicular Thyroid Carcinoma. *Hum Pathol* (2003) 34:1092–100. doi: 10.1016/S0046-8177(03)00403-9

90. Pearson RG, Kurien T, Shu KSS, Scammell BE. Histopathology Grading Systems for Characterisation of Human Knee Osteoarthritis – Reproducibility, Variability, Reliability, Correlation, and Validity. *Osteoarthr Cartil* (2011) 19:324–31. doi: 10.1016/j.joca.2010.12.005

91. Beaton C, Twine CP, Williams GL, Radcliffe AG. Systematic Review and Meta-Analysis of Histopathological Factors

Influencing the Risk of Lymph Node Metastasis in Early Colorectal Cancer. *Color Dis* (2013) 15:788–97. doi: 10.1111/codi.12129

92. Pallua JD, Brunner A, Zelger B, Schirmer M, Haybaeck J. The Future of Pathology is Digital. *Pathol - Res Pract* (2020) 216:153040. doi: 10.1016/j.prp.2020.153040

93. Bankhead P, et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci Rep* (2017) 7:16878. doi: 10.1038/s41598-017-17204-5

94. Stirling DR, Carpenter AE, Cimini BA. CellProfiler Analyst 3.0: Accessible Data Exploration and Machine Learning for Image Analysis. *Bioinformatics* (2021) 37:3992–4. doi: 10.1093/bioinformatics/btab634

95. Schapiro D, et al. histoCAT: Analysis of Cell Phenotypes and Interactions in Multiplex Image Cytometry Data. *Nat Methods* (2017) 14:873–6. doi: 10.1038/nmeth.4391

96. Schapiro D, et al. MCMICRO: A Scalable, Modular Image-Processing Pipeline for Multiplexed Tissue Imaging. *Nat Methods* (2021) 19:311–5. doi: 10.1038/s41592-021-01308-y

97. Bortolomeazzi M, et al. A SIMPLI (Single-Cell Identification From MultiPLexed Images) Approach for Spatially-Resolved Tissue Phenotyping at Single-Cell Resolution. *Nat Commun* (2022) 13:1–14. doi: 10.1038/s41467-022-28470-x

98. Hoshida Y. Nearest Template Prediction: A Single-Sample-Based Flexible Class Prediction With Confidence Assessment. *PloS One* (2010) 5:e15543. doi: 10.1371/journal.pone.0015543

99. Hickey JW, Tan Y, Nolan GP, Goltsev Y. Strategies for Accurate Cell Type Identification in CODEX Multiplexed Imaging Data. *Front Immunol* (2021) 12. doi: 10.3389/fimmu.2021.727626

100. Berens ME, Sood A, Barnholtz-Sloan JS, Graf JF, Cho S, Kim S, et al. Multiscale, Multimodal Analysis of Tumor Heterogeneity in IDH1 Mutant vs Wild-Type Diffuse Gliomas. *PloS One* (2019) 14(12):e0219724. doi: 10.1371/journal.pone.0219724

101. Badve SS, et al. Multi-Protein Spatial Signatures in Ductal Carcinoma *in Situ* (DCIS) of Breast. *Br J Cancer* (2021) 124:1150–9. doi: 10.1038/s41416-020-01216-6

102. Jackson HW, et al. The Single-Cell Pathology Landscape of Breast Cancer. *Nature* (2020) 578:615–20. doi: 10.1038/s41586-019-1876-x

103. Arnol D, Schapiro D, Bodenmiller B, Saez.-Rodriguez J, Stegle O. Modeling Cell-Cell Interactions From Spatial Molecular Data With Spatial Variance Component Analysis. *Cell Rep* (2019) 29:202–211.e6. doi: 10.1016/j.celrep.2019.08.077

104. Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. Explainable Multi-View Framework for Dissecting Intercellular Signaling From Highly Multiplexed Spatial Data. *Genome Biol* (2022) 23(1):97. doi: 10.1186/s13059-022-02663-5

105. Schapiro D, Sokolov A, Yapp C, Chen YA, Muhlich JL, Hess J, et al. MCMICRO: A Scalable, Modular Image-Processing Pipeline for Multiplexed Tissue Imaging. *Nat*

Methods (2022) 19(3):311–315. doi: 10.1038/s41592-021-01308-y

106. Nirmal AJ, Maliga Z, Vallius T, Quattrochi B, Chen AA, Jacobson CA, et al. The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single Cell Resolution. *Cancer Discov* (2022) 12(6):1518–1541. doi: 10.1158/2159-8290

107. Pell R, et al. The Use of Digital Pathology and Image Analysis in Clinical Trials. *J Pathol Clin Res* (2019) 5:81–90. doi: 10.1002/cjp2.127

108. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet* (2019) 10:317. doi: 10.3389/fgene.2019.00317

109. Guillot A, Kohlhepp MS, Bruneau A, Heymann F, Tacke F. Deciphering the Immune Microenvironment on A Single Archival Formalin-Fixed Paraffin-Embedded Tissue Section by An Immediately Implementable Multiplex Fluorescence Immunostaining Protocol. *Cancers (Basel)* (2020) 12:1–18. doi: 10.3390/cancers12092449

110. van Maldegem F, Valand K, Cole M, Patel H, Angelova M, Rana S, et al. Characterisation of Tumour Microenvironment Remodelling Following Oncogene Inhibition in Preclinical Studies With Imaging Mass Cytometry. *Nat Commun* (2021) 12(1):5906. doi: 10.1038/s41467-021-26214-x

111. Pombo Antunes AR, Scheyltjens I, Lodi F, Messiaen J, Antoranz A, Duerinck J, et al. Single-cell profiling of myeloid cells in glioblastoma across species and disease stage reveals

macrophage competition and specialization. *Nat Neurosci*. (2021) 24(4):595–610. doi: 10.1038/s41593-020-00789-y

112. Naulaerts S, Borras DM, Martinez AA, Messiaen J, Van Herck Y, Gelens L, et al. Immunogenomic, Single-Cell and Spatial Dissection of CD8+T Cell Exhaustion Reveals Critical Determinants of Cancer Immunotherapy. *bioRxiv* (2021), 468617. doi: 10.1101/2021.11.22.468617

113. Halse H, et al. Multiplex Immunohistochemistry Accurately Defines the Immune Context of Metastatic Melanoma. *Sci Rep* (2018) 8:11158. doi: 10.1038/s41598-018-28944-3

114. Bosisio F, Antoranz A, Herck YV, et al. Mapping the Immune Landscape in Metastatic Melanoma Reveals Localized Cell-Cell Interactions Correlating to Immunotherapy Responsiveness. *Research Square* (2022). doi: 10.21203/RS.3.RS-1236531/V1.

115. Sade-Feldman M, et al. Defining T Cell States Associated With Response to Checkpoint Immunotherapy in Melanoma. *Cell* (2018) 175:998–1013.e20. doi: 10.1016/j.cell.2018.10.038

116. Schürch CM, et al. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* (2020) 182:1341–1359.e19. doi: 10.1016/j.cell.2020.07.005

117. Rakaee M, et al. Prognostic Value of Macrophage Phenotypes in Resectable Non–Small Cell Lung Cancer Assessed by Multiplex Immunohistochemistry. *Neoplasia (United States)* (2019) 21:282–93. doi: 10.1016/j.neo.2019.01.005

118. Hofman P, Badoual C, Henderson F, Berland L, Hamila M, Long-Mira E, et al. Multiplexed Immunohistochemistry for Molecular and Immune Profiling in Lung Cancer—Just About Ready for Prime-Time? *Cancers (Basel)* (2019) 11(3):283. doi: 10.3390/cancers11030283

119. Färkkilä A, et al. Immunogenomic Profiling Determines Responses to Combined PARP and PD-1 Inhibition in Ovarian Cancer. *Nat Commun* (2020) 11:1459. doi: 10.1038/s41467-020-15315-8

120. Patil NS, et al. Intratumoral Plasma Cells Predict Outcomes to PD-L1 Blockade in non-Small Cell Lung Cancer. *Cancer Cell* (2022) 40:289–300.e4. doi: 10.1016/j.ccell.2022.02.002

121. Lewis SM, Asselin-Labat ML, Nguyen Q, Berthelet J, Tan X, Wimmer VC, et al. Spatial Omics and Multiplexed Imaging to Explore Cancer Biology. *Nat Methods*. 18(9):997–1012. doi: 10.1038/s41592-021-01203-6.

122. Hoch T, Schulz D, Eling N, Gómez JM, Levesque MP, Bodenmiller B. Multiplexed Imaging Mass Cytometry of the Chemokine Milieus in Melanoma Characterizes Features of the Response to Immunotherapy. *Sci Immunol* (2022) 7(70):eabk1692. doi: 10.1126/sciimmunol.abk1692

123. Hao Y, et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* (2021) 184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048

124. van Eck NJ, Waltman L. Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping. *Scientometrics* (2010) 84:523–38. doi: 10.1007/s11192-009-0146-3
125. Bolognesi MM, Manzoni M, Scalia CR, Zannella S, Bosisio FM, Faretta M, et al. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *J Histochem Cytochem*. (2017). 65(8):431–444. doi: 10.1369/0022155417719419

Keywords: multiplexed immunofluorescencence and immunohistochemistry, spatial profiling, single-cell ‘omics, tissue architecture analysis, methods for spatial profiling

Citation: Bosisio FM, Van Herck Y, Messiaen J, Bolognesi M, Marcelis L, Van Haele M, Cattoretti G, Antoranz A and De Smet F (2022) Next-Generation Pathology Using Multiplexed Immunohistochemistry: Mapping Tissue Architecture at Single-Cell Level. *Front. Oncol.* 12:918900. doi: 10.3389/fonc.2022.918900

Received: 12 April 2022; Accepted: 20 June 2022;

Published: 29 July 2022.

Chapter 5

BRAQUE: Bayesian Reduction for Amplified Quantization in UMAP Embedding

Lorenzo Dall'Olio 1∗†, Maddalena Bolognesi 2†, Simone Borghesi 3, Giorgio Cattoretti 2‡ and Gastone Castellani 4‡

1 Department of Physics and Astronomy, University of Bologna, Bologna, Italy; lorenzo.dallolio4@unibo.it

2 Department of Medicine and Surgery, University of Milano Bicocca, Monza, Italy

3 Department of Mathematics and Applications, University of Milano Bicocca, Milan, Italy

4 Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy

* Correspondence: lorenzo.dallolio4@unibo.it or lorenzo.dallolio@studio.unibo.it; (L.D.) † These authors contributed equally to this work.

‡ These authors contributed equally to this work.

Abstract: Single-cell biology has revolutionized the way we understand biological processes. In this paper, we provide a more tailored approach to cluster and analyze spatial single-cell data coming from immunofluorescence imaging techniques. We propose Bayesian Reduction for Amplified Quantization in UMAP Embedding (BRAQUE) as an integrative novel approach, from data preprocessing to phenotype classification. BRAQUE starts with an innovative preprocessing, named Lognormal Shrinkage, able to enhance input fragmentation by fitting a lognormal mixture model and shrinking each component towards its median, in order to help further clustering step in finding more separated and clear clusters. Then BRAQUE's pipeline consist of a dimensionality reduction step performed using UMAP, and a clustering performed using HDBSCAN on UMAP embedding. In the end, clusters are

assigned to a cell type by experts, using effects size measures to rank markers and identifying characterizing markers (Tier 1) and possibly characterizing markers (Tier 2). The number of total cell types in one lymph node detectable with these technologies is unknown and difficult to predict or estimate. Therefore, with BRAQUE we achieved a higher granularity than other similar algorithms such as PhenoGraph, following the idea that merging similar clusters is easier than splitting unclear ones into clear subclusters.

Keywords:
Multipleximmunostaining;GaussianMixture;Lognormal;Single-cell;Bayesian;UMAP; DBSCAN; Clustering; Lymphoid tissue; Effect size;

1. Introduction

Single-cell (1) biology has revolutionized the way we understand biological processes[1]. iPrior to 2013, very few biological assays were indicative of single-cells (e.g., FlowCytometry); the introduction of single-cell RNA sequencing changed the paradigm. In a few words it can be explained as the detection of the same traits in every single-cell in a sample, not just as the mean value of the bulk. This revolution has added a new level of resolution to what we can capture and makes us better understand the complexity of a sample. In medicine the single-cell approach is particularly helpful, revealing unknown mechanisms in healthy and pathological tissues, and improving healthcare[2] [3]. Characterization of single-cells has increased our understanding of cell phenotypes, the dynamics and

trajectory of their development, and their network.

All of this has been possible thanks to the fast development of a large group of technologies: “single-cell multi-omic technologies”. Each of them provides a very specific kind of information for every cell analyzed based on genomics, transcriptomics, proteomics, and metabolomics signals, not to mention spatial approaches that associate with cell identity its x and y localization within the tissue. This, together with information related to neighboring cells[4], is crucial information because cell identity and role are determined by the spatial context. Integration of them into “single-cell multimodalomics” [5][6][7] is the next step to fully characterize cells and cell types.

This impacted the way of doing research and new bioinformatics tools became more and more indispensable in order to explore single-cell data. Moreover, single-cell analysis was intrinsically born as a discipline that has to face big data sets, with thousands of entities (i.e., cells), and tens to thousands of columns. All of this had to be managed with unsupervised/supervised approaches, and required different expertise. Therefore, this kind of data is a complex kind of data to deal with.

While these technologies were flourishing, many global research efforts have been done to collect and share an enormous amount of single-cell data and to make them publicly available. Some large consortia have been created and represent now the standard and the repository to which refer to. Among them the most important are Human Cell Atlas

(HCA)[8], BRAIN Initiative Cell Census Network (BICCN)[9] and Human Biomolecular Atlas (HuBMAP)[10].

Together with all of these novelties, high-plex spatial proteomic represents a small but crucial niche for many reasons: it has a single-cell resolution associated with spatial localization, it evaluates the presence of protein and not RNA signal (bypassing post transcriptional modification), it analyses whole cells without losing any types of them [11] as often happens in tissue disaggregation [12] or other single-cell technologies, it allows retrospective study on Formalin-Fixed Paraffin-Embedded (FFPE) sample collection, which is easily integrated with other "-omics" technologies.

Thanks to global initiatives and to the commercialization of these technologies, the need to approach these big data from a computing point of view, in order to standardize analysis and make them available, grew exponentially. In the last few years more than 1000 bioinformatic tools have been developed, mainly in the two world-wide used interpreted languages, R and Python [13].

In this project we consider single-cell proteomics data extracted from the lymphoid tissue, which is a very dense type of tissue (approximately 1 to 2 · 106 cells per *mm*3 of tissue [14]).

Among the big repositories of single-cell data, it is not very often represented. If present it contained only a selected population of cells, such as stromal cells [15] or an organ, such as tonsils [16]. Unfortunately, with the exception of the HuBMAP project [10], there are no other lymph nodes data set from

“imaging technologies” available and, at the moment, shared data have not more than 30~40 Antibodies.

Moreover, In the single-cell field the majority of developed tools are specifically cre- ated for single-cell RNA sequencing data analysis [17][18]. One of the few method widely applied outside the original scRNAseq data is PhenoGraph [19]. The evolution of technologies [20] [21] has added modules for data integration, annotation, spatial distribution, and neighborhood analysis, however a dedicated part to approach data sets coming from single-cell imaging technologies is still missing or preliminary [22].

In short, many methods are tuned and perform clustering for high dimensional complex data, but very few of them are suited for spatial transcriptomics and the continuous property of its data. Moreover, the high number of markers included in most database presented during this analysis (7 out of 8 datasets have 70+ markers) adds a further level of information that can and should be analyzed together, and this is the target for which BRAQUE was developed.

1 Whenever we speak of “single-cell” acquired from imaging techniques, it would be appropriate to remind ourselves that, in reality, we are talking about computational approximations of cell-segmentation boundaries and spatial-mappings of features to DAPI stained nuclei.

2. Materials and Methods

2.1. Datasets and Data Acquisition

The data used for the analysis consisted of 8 different data set (per number of markers and cells) of normal lymphoid tissue,

which were obtained with the Multiple Iterative Labeling by Antibody Neodeposition (MILAN) technology [23] (seven samples) and CODEX (one sample). MILAN datasets consist of 3 tonsils and 3 lymph nodes cores (each of 2mm in diameter) and one whole lymph node, all of them were sections of 5 μm. The number of cells were ranging from ∼25k to ∼65k for the 6 cores, while the whole lymph node dataset had ∼730k cells. Each of these dataset counted 70+ markers after a marker selection step performed by the experts, all of them acquired with 8-bits channels. Cores belonged to 3 different tissue microarray (TMA) constructed with a Tissue Microarrayer Galileo CK4500 (Tissue Microarrayer Model TMA Galileo CK4500; Integrated Systems Engineering srl, Milan, Italy). Sections were stained by using MILAN technology [23] which consists of multiple stainings, imaging and stripping cycles in Immunofluorescence. Images were acquired with a NanoZoomer S60 slide scanner (Hamamatsu, Japan) at 20x magnifica- tion. This method has been shown to preserve tissue integrity and provide high stainings reproducibility (less than 10% variation) over 30 cycles [24].

Primary antibodies were validated for in-situ use on FFPE sections [25]. Highly ex- pressed, partially overlapping, lineage-defining markers were preferred, including nuclear transcription factors. Multiplex staining and image optimization was performed according to a published protocol (MILAN [23]). CODEX dataset was downloaded from the reposi- tory Globus (of the HubMAP project) and counted of ∼109k cells. This

dataset counted of only 28 markers acquired in 16-bits channels, and therefore there was no marker selection step, in order to not enlarge the already considerable difference with MILAN datasets in terms of dimensionality.

DAPI-based image segmentation was performed by means of the algorithm "CyBorgh" developed by S. Borghesi (available upon request to the same author). It is a Matlab code which can handle very large images (in excess of 250 megapixels) in a reasonable time (it took less than 3 hours on an Intel Xeon 6130 2.1 2.6GHz 16 core, 192 GB RAM machine to segment a 1 terapixel 8-bit gray scale image with 2.2 million cells found and to generate its dataset with 92 filters). CyBorgh operates in two steps: it initially searches for the boundaries of individual tissue cells in the DAPI image after having applied to it a series of filters which must be carefully tuned to specific image features such as noise, cell size, shape and inner structure details. Once the segmentation of the DAPI image is produced, the connected components of its complement (which are expected to correspond to tissue cells) are sorted out in a Matlab cell array, each of its objects containing the coordinates of the pixels of one component. If *n* is the number of filtered images of the same portion of tissue, a stack of *n* accurately registered images is used to read out the pixel values of each component in every image. Let *k* be the number of pixels of a component. This way the algorithm associates to that component *n* sets of *k* integers between 0 an+-d 255. By fixing a metric (in our case it was the mean), we can “merge” each of

the *n* sets of integers in a rational number, thereby producing a vector in the *n* dimensional Euclidean space. The data points comprise all such vectors, each representing one tissue cell/component in the DAPI image in such a space. Their coordinates are the "raw" data. Summing up, CyBorgh takes *n* registered images as input, one of which is declared to be DAPI, and outputs a .csv file with rows corresponding to coordinates of data points and columns to filter response on tissue cells.

It was at this point that data continuous nature came in play. When the pixel is acquired, for every marker, it has a usual 8-bit discrete value. But after the segmentation step, every identified cell was made of multiple pixels, and therefore their value needed to be averaged to extract a single value for a single-cell for each marker, giving the continuous nature to our data. Moreover, single-cell data from CODEX dataset have already been observed to behave as a continuum regarding protein expression [26], reinforcing the concept that we are not dealing anymore with discrete data as in single-cell RNA seq.

2.2. Data preprocessing and lognormal shrinkage

In the preprocessing we introduced a new method to improve the capability of identi- fying clusters in the successive steps of the pipeline.

The core idea is to look at each marker distribution within a dataset (e.g., considering only one dataset at a time) and find possible subpopulations with similar distribution but shifted in location (therefore having higher or lower marker expression).

This concept resembles quite well the working principle of distribution mixture models.

First we have to highlight that one of the main differences between transcriptomic and fluorescence data resides in the distribution followed by their values. In both cases, values are non-negative, but transcriptomic data follow a discrete distribution, while the data used for this study are continuous. Therefore, many algorithms and packages built for discrete data such as Poisson, or Negative-Binomial distribution, are not suited for continuous data.

Since mixture models are often computationally heavy, to deal with our data, we had to find a distribution exhibiting the following properties: continuous, non-negative, and fast to compute and model. Therefore we chose to use the lognormal distribution as the basic element for our mixture model. This also allowed the use of pre-built and optimized gaussian mixture models, still for computational reasons, instead of a slower mixture model with customizable distribution.

To perform this task, we used the Bayesian Gaussian Mixture algorithm from scikit learn python library. The reason behind such choice is that this algorithm uses the expecta- tion maximization (EM) technique to infer from the data the correct number of components to use, where each component is a gaussian distribution with its own independent mean and variance. The only precaution to consider is that such algorithm requires a number of components as input parameter. This number is not lowered, but the fitting algorithm can send some

components importance exactly to 0, excluding them from any kind of contribution. Therefore this parameter acts as an upper limit for the number of gaussians used to estimate subpopulations, and this property should be properly taken into account. On one side, the higher this number and the more accurate the algorithm estimates. On the other hand, the smaller the number, the faster the convergence of the algorithm.

We will further show values for this parameter, but in order to generalize, we suggest a procedure to properly tune a good trade-off between accuracy and computation time: if time is not a concern, choose the smallest number of components for which every feature (or almost every feature) ends up having at least 1 discarded component after the EM fit procedure. If instead time is a concern and the previous suggested value ends up making the algorithm too slow, we suggest to use the highest value that meets your time requirements, since the lower the value, the more subpopulation will risk to be merged with each other (making it harder for the next steps to split them). Such recommendation is not strict, using slightly bigger values for the number of components will simply imply a waste of time but equally good results, using slightly smaller values will risk to sacrifice a bit of quantization, but the core of this step is just to *guess* possible initial subpopulations, the successive pipeline is in charge for validating and tuning such guess.

The EM algorithm can be summarized in giving the set of tunable parameters some random values, and then recursively

apply two steps until convergence:

1. E-step, compute the probability of the data points to belong to the various gaussian distributions given by the most recent set of parameters values (and assign the data points to the most probable distribution);
2. M-step, use the updated assignments to compute the new best parameters for the distributions, obtained by maximizing the likelihood L of the assigned data given at the previous step;

therefore this method tunes the parameters maximizing the likelihood of the data, using the information carried in it.

When the EM algorithm converged, all points were assigned to a gaussian distribution by the algorithm. This procedure was possible considering for each value the probability of belonging to the i-th gaussian, computed as:

where *gi* is the i-th gaussian component, characterized by mean *mi* and standard deviation *σi*. The point x was assigned to the gaussian with the highest $P(x \in gi)$.

Once all points were assigned to a gaussian distribution, every point was then shrunk towards the mean of the belonging gaussian using

where *γ* is a properly tuned contraction factor. This step dismembers the original distribution, but preserves the order. Doing so values rankings are maintained, but gaps are created in correspondence of where we could have a good separation between two different subpopulations.

A small incompatibility could be that the lognormal distribution is

strictly positive instead of non-negative, but such a problem was easily solved by adding a very small constant before fitting the mixture model. Doing so, one of the gaussian components from the mixture model was always going to account for the values at the very small value, basically identifying as a separate subpopulation all those values that started as a zero in the original distribution.

A summary of the preprocessing algorithm for each single marker could be as follows:

1. Robustly scale the marker distribution, dividing it by the median absolute deviation (MAD). This step is suggested so all the parameters for the Bayesian gaussian mixture algorithm are going to be the same for every marker;
2. sum a small positive constant to every value to avoid taking the logarithm of 0;
3. compute the logarithm of the shifted and robustly scaled distribution;
4. perform a Bayesian gaussian mixture fit using EM algorithm;
5. once the final gaussians are identified, shrink every value towards its belonging

gaussian, then back transform the values by exponentiation;

6. subtract the minimum of the distribution and (optionally) scale robustly, dividing

again by the new MAD of the final distribution.

We named such preprocessing as Lognormal Shrinkage (LS). Two final considerations regarding the LS method:

· The shrinking factor was tuned based on the quality of

results, but no big difference in the range from 2 to 10, was observed, therefore a value of 5 was chosen;

- the base for the logarithm is not affecting the performances, only scales the log trans- formed distribution; therefore, we used base 2 logarithm, and in case of different choices the shrinking factor should be tuned accordingly (e.g. the base 10 logarithm should use a contraction factor of $5/\ln(10) \sim 2$.

An example of the effect of the LS procedure on the distribution of data is reported in Figure 1.

2.3. Dimensionality reduction step

The dimensionality reduction step has the aim to move data from a high-dimensional space to a lower-dimensional space, called embedding, in order to tackle data sparsity and other problems caused by the course of dimensionality.

For this step Uniform Manifold Approximation and Projection (UMAP) [27] was chosen due to its huge advantages in terms of memory and computation time with respect to t-SNE (t-distributed stochastic neighbor embedding) [28], and its capability of wrapping more information than older methods such as Multidimensional scaling (MDS) [29]. Two main parameters that must be fixed before running UMAP are: the number of nearest neighbors K and the metric used to compute the distance. For the first, higher values aim to produce a much more characterized embedding by the global structure. A value of 50 (instead of the default 15) was used to have a good trade-off between global structure and computational efficiency. Regarding the metric, an angular metric (i.e., correlation) was

used, where such a distance between vector **x** and **y** was computed as:

$$d_{ij} = 1 - corr(\mathbf{i}, \mathbf{j}) = 1 - \frac{(\mathbf{i} - \bar{\mathbf{i}}) \cdot (\mathbf{j} - \bar{\mathbf{j}})}{||(\mathbf{i} - \bar{\mathbf{i}})|| \cdot ||(\mathbf{j} - \bar{\mathbf{j}})||} \quad (3)$$

where the second equivalence assumes the usage of Pearson's correlation, but one could change it to Spearman's correlation in order to investigate more general trends. We preferred the first for computational reasons.

The UMAP algorithm can be summarized into 2 main phases. First a weighted directed graph is built to represent the data neighborhood relations in the starting high dimensional space. A graph is made of nodes (which represents data points) and edges which connect two nodes, where in our case such connection represents a proximity relation between the two connected nodes. The fact that the graph is weighted means that not all edges are equal, in fact they might be more or less strong according to an intensity value called *weight* which resembles higher or lower proximity among the connected nodes. Directed simply implies that the weight *wij* of the link connecting node i to node j can differ from the weight *wij* connecting j to i. In UMAP the weight *wij* between point i and j is computed using the equation:

$$w_{ij} = e^{-\max(0; d_{ij} - \rho_i)/\sigma_i} \quad (4)$$

where *σi* is tuned such that

$$\sum_{j=1}^{k} w_{ij} = log_2(k) \quad \text{and} \quad \rho_i = \min_{j \in \{1,\dots,k\}} (d_{ij} > 0) \tag{5}$$

Here *dij* is the distance according to the metric in the high-dimensional space, *σi* is called the scaling factor for point i and it is used to normalize the sum of the outgoing weighted edges, and finally *ρi* is the smallest strictly positive distance between point i and its neighbors $j \in \{1,...,k\}$ (and therefore $\rho i > 0$).

The second phase of the algorithm is the low-dimensional space (called embedding) optimization. This is needed to reproduce as correctly as possible the weighted directed graph structure. Such procedure is achieved by placing the data points in the embedding space (either with random starting positions or with coordinates initialized by some criteria, for example via spectral embedding or, if not possible, with PCA coordinates), and then alternatively apply to every point i attractive forces towards its nearest neighbors and repulsive forces towards its non-nearest neighbors. The intensity of both attractive and repulsive forces can be tuned among the initial set of UMAP parameters. The final output of the algorithm is a low-dimensional space which replicates the proximity observed in the original high-dimensional space among the data points, but with a considerable boost for clustering algorithms performances and with the advantage of possibly visualizing the new data space.

2.4. Clustering step

All clustering algorithms have advantages and disadvantages, but we will try to briefly explain the reason for our suggested choice. Hierarchical clustering brings the obvious advantage of producing a hierarchical structure that can be investigated, allows the plotting of dendrograms, and follows the very simple intuition of merging the 2 closest elements together iteratively.

One of its possible drawbacks is related to the linkage method, which affects the shapes of the resulting clusters [30]. In fact it is well known that single linkage produces for example narrow line-like clusters, while complete linkage produces more spherical clusters. And in general, even if more and more linkage methods can be found, all of them imply some kind of assumption regarding the resulting clusters' shape. One possible solution to this drawback is switching to a density-based clustering algorithm. The most renown of this kind is Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [31], which requires two input parameters: ε and min_samples. The first defines what is close, the second defines more or less the number of points needed to start the formation of a cluster.

Thanks to the usage of a distance parameter and a certain number of points that must be contained in such distance, DBSCAN uses density to form its clusters, allowing any possible shape with no restrictions. The big downside in this case becomes the fact that now the threshold is imposed on the density, and therefore identifying clusters of different densities could become a problem, where the bigger the difference

between densities and the less the chances to identify both clusters at the same run.

To briefly explain DBSCAN for further purposes, its procedure could be summarized in the following way:

- Given the two parameters epsilon and min_samples, a point x is defined as a “core point” if and only if at least min_samples points are within ε distance from x.
- Connect all “ε-reachable” points. Where two points are said to be “ε-reachable” if they are in each other’s neighborhood, and points are “density-connected” if they are directly or transitively “ε-reachable”.
- We conclude labeling as a cluster every maximal “density-connected” subset of the data, while the remaining unlabeled points are relabeled as noise at the end of the algorithm.

The HDBSCAN algorithm [32] is the hierarchical improvement of DBSCAN, and basically provides advantages from both approaches, while limiting their drawbacks. It allows detecting clusters with generic shapes and possibly very different densities, as long as such densities are higher than the respective neighborhood. Therefore this last algorithm is the suggested algorithm for the clustering phase. HDBSCAN’s working principle is similar to running DBSCAN for all $\varepsilon \in [0, +\text{inf}[$, then use the different values of ε (or more precisely, a new distance, called Mutual Reachability Distance and based on the centroid of the neighborhood) to connect points in a hierarchical dendrogram fashion. Finally clusters are extracted not by

cutting the dendrogram at a given depth, but instead considering cluster stability at varying depth. This last step is performed using $\lambda = \varepsilon - 1$, and then identifying for each candidate clusters their birth (the λ at which cluster forms) and their death (the λ at which the cluster splits into separate clusters).

Now for each point in the cluster we can define λp as the value at which that given point leaves the cluster. Therefore $\lambda p \in]\lambda birth, \lambda death]$, since each point either leaves the clusters before death or exactly when the cluster splits into minor clusters. Now we can calculate the stability of cluster C as follows:

$$\sum_{\forall p \in C} \lambda_p - \lambda_{death} \tag{6}$$

The cluster extraction is then performed by starting from leaves clusters, and replacing some of them with their roots cluster every time that root cluster stability is bigger than the sum of the leaves cluster stability.

Last but not least, one could choose to select much simpler methods such as DBSCAN for computational reasons, but turns out that for big databases the performances of HDB- SCAN are better in terms of computation time [33], and therefore HDBSCAN is preferable also under this point of view.

2.5. Cluster characterization step

After the clusters are identified, a method to describe which

markers are mostly expressed from the clusters is needed. A first step was to perform a Welch t-test for each marker, using as null hypothesis $H0 = (mC \leq m)$, where mC is the mean of the marker in the cluster and m is the mean of the marker in the remaining cells from the whole sample.

An important digression should be made regarding the p-values at this point. It is quite common to find in literature the usage of a standard threshold of 0.05, but we have to notice how p-value is always a function of the number of samples used in the test. In particular, the more the samples, the smaller the p-value (if a difference between the groups is present). Therefore, samples with a considerably high number of cells (e.g., 105 ∼ 106) would end up in emphasizing even very small differences, often producing p-values $< -10^{-100}$.

It seems indispensable to understand if a marker is significantly more expressed in a cluster with respect to the whole sample, but it is also important to properly quantify the magnitude of such difference. For this reason, together with a Welch's t-test with Bonferroni multiple test correction, a measure of effect size was used to quantify how much a marker is more expressed in the considered cluster. Thanks to this measure it was possible to rank for every cluster the markers from most expressed to least expressed.

A further advantage of this approach was the capability of looking for gaps in the ranked effect sizes, and defining markers probably expressed by the cluster, called tier 1 markers, and clusters possibly expressed by the cluster, called tier 2. The

tiers were identified by first looking for the two biggest gaps in the positive ranked effect sizes, and then calling tier 1 the markers whose effect size was higher than the first gap, and tier 2 the markers whose effect size was between the two identified gaps.

To calculate the effect size, a robust version of the Cohen's d was used [34], which is computed as:

$$d := \sqrt{\frac{(m_2 - m_1)^2}{\sigma_1^2 \cdot \frac{N_1+N_2}{N_1} + \sigma_2^2 \cdot \frac{N_1+N_2}{N_2}}} \tag{7}$$

where *mi*, *σi* and *Ni* indicate respectively the mean, the variance, and the number of samples from group i, having usually that the group 2 was the analyzed cluster, and the group 1 was the whole sample excluding the analyzed cluster. One last detail to notice is that the formula is symmetric with respect to the two groups, therefore from this version of d it was not possible to understand if a marker was more or less expressed than the whole sample, for such reason the final metric used the formula:

$$d_{signed} = sign(m_2 - m_1) \cdot \sqrt{\frac{(m_2 - m_1)^2}{\sigma_1^2 \cdot \frac{N_1+N_2}{N_1} + \sigma_2^2 \cdot \frac{N_1+N_2}{N_2}}} \tag{8}$$

allowing to have negative values of *dsigned* when *m*2 < *m*1.

2.6. Comparison with other existing algorithms and final validation

As last step, we compared the outcomes obtained by BRAQUE

with the ones obtained using PhenoGraph [19] R package, which was run on RStudio (R version 4.0.2).

Cell type assignment for each BRAQUE cluster was based on a list of its most significant markers in terms of ranked effect size. For PhenoGraph, a heatmap of marker expression by cluster was used to assign a cell type, as performed in previous work using the same algorithm [24]. Then the evaluation considered: number of noisy/unclear clusters, rare populations correctly clustered as separate clusters, and number of redundant clusters (where this was considered as a minor side effect as far as the clusters were explainable as a correct cell type).

3. Results

In this section we will show the results that can be achieved using BRAQUE's pipeline. For clarity purposes we will not show every produced plot (since in total we produced at least 3 plots for each of the 620 clusters), but we will focus on explaining the key steps and showing the results for: 1 preprocessed marker, 1 whole sample analysis, 1 cell type across different samples, and a comparison with PhenoGraph algorithm.

For the analysis, a server with Intel(R) Xeon(R) CPU E5-2620 v4 2.10GHz was used, with 32 cores, 252 GB of RAM and Python version 3.9.12. In terms of memory consumption, the average memory required for the analysis was around 30 GB for the biggest database (~ 730k rows x 80 columns). Moreover the usage of UMAP's parameter "low_memory" set equal to "true" allowed it to never exceed 100 GB of RAM usage,

required only for the nearest neighbor search phase of the algorithm (which took approximately 1 hour). In terms of time required to run the analysis on the biggest database, approximately 95% of the computation time of the pipeline resides in the Bayesian Gaussian Mixture fitting procedure, which took around 2 hours per column, resulting in a week of computation time for the whole analysis. It is important to underline that both, memory and time consumption are not linear with respect to the number of rows in the database, therefore the smallest databases (the cores with a number of rows between 20k and 70k) were computed with around 10 GB of memory, a peak of 30 GB, and with the whole analysis completed in approximately 3 ~ 8 hours.

Fine tuning of the parameters was obtained by a continuous and mutual feedback between bioinformatics and pathologists regarding computational needs, results clarity and interpretability. As in every machine learning tool, parameters should be tuned based on data nature, desired result quality, and computational efficiency. In Table 1 we briefly report suggested ranges and values that were tuned during this analysis.

Some small clarifications about Table 1 reported values:

- regarding BGM *covariance_type*, UMAP's *init*, and HDBSCAN *cluster_selection_method* should be fixed to the suggested value, since any alternative could only imply worse results in terms of correctness and generalization;
- regarding BGM *n_components* and *tol*, the best values

would be as high as possible for the first and as low as possible for the second, but what we observed is that such exaggeration has strong computational drawbacks, and therefore a suggestion is to use a relatively low *tol* and tune the number of gaussians following the criteria described in the methods section;

- all the other parameters should be tuned dataset-wise for better performances, but we can point out that, as long as values are in the suggested range, *contraction_factor*, UMAP's *metric*, *min_dist* and *nn* will not strongly affect the results;
- HDBSCAN *min_samples* could be tuned as a proportion of the considered dataset, we observed that for whole lymph node dataset the best performances were obtained using 0.05% of the total number of cells, while for CODEX and the cores, values between 0.1% and 0.2% were better performing.

Lastly, if time is again not a concern, we strongly recommend to fix BGM's *max_iter* as high as possible to always reach convergence criteria, otherwise compute for the maximum amount of time possible and set this parameter consequently.

3.1. Lognormal Shrinkage on a marker

Now, to show the way BRAQUE preprocessing works, we will show step-by-step the effect of Lognormal Shrinkage on the marker *vWF*, a marker used to identify "Endothelium" kind of cells, a rare, sparse but well-defined cell type.

As previously mentioned in the methods section, the procedure

is performed on every marker separately. For each marker, a robust scaling is performed, dividing by the MAD. This to have all the markers in similar ranges, and therefore having the fitting parameters to behave similarly among them. Then, a small constant is added (to avoid further logarithms of 0) the logarithm is taken for every value, the Bayesian gaussian mixture algorithm is fitted using Expectation Maximization algorithm (Figure 1a). When the fit converges, each cell is assigned to the component to which it is most likely it belongs to (Figure 1b), and its marker value is modified by shrinking it towards the mean of the assigned gaussian distribution (Figure 1c). The same effect can be observed on the back transformed marker distribution (Figures 2), reminding that every normal distribution in the log2 space becomes a lognormal distribution in the back transformed space, with new mean and standard deviation given by

$$\mu_{lognormal} = e^{(\mu_{normal} + \frac{1}{2}\sigma^2_{normal})} \tag{9}$$

$$\sigma_{lognormal} = \sqrt{(e^{\sigma^2_{normal}} - 1) \cdot e^{(2\mu_{normal} + \sigma^2_{normal})}} \tag{10}$$

where the pedix *normal* indicates the normal distribution in the log2 space, while the pedix *lognormal* indicates the lognormal distribution in the original data space.

3.2. Analysis of a sample

To give a broad view on the analysis, we will now show the step-by-step pipeline application to a whole sample, in this case specifically to the dataset L2. First of all, in Figure 3 we show that the number of gaussians used (i.e., 15), was sufficient,

given that no marker exceeds the 14 different gaussians with non-zero weight after the fitting procedure. The main results are, after the application of the Lognormal shrinkage preprocessing to every marker, the application of UMAP on the preprocessed dataset, and the application of HDBSCAN clustering on the UMAP embedding (Figure 4). For evaluation, we runned UMAP with the same identical parameters, but on the raw data (Figure 5) instead of the preprocessed data, and from the comparison of these two figures it seems clear how the Lognormal shrinkage procedure added a considerably higher separation in the UMAP embedding, that is not depending on UMAP's parameters.

Once the clusters are identified, it is possible to perform the validation for both, single clusters and global results. Since the single cluster step will be the focus of the next subsection, now we will focus on the global result. This step can be performed with a scatter plot on the real space, where every cluster has an assigned explanatory label. Such a label could be the list of the main expressed markers in terms of size effect (Figure 6).

The spatial reconstruction of the clusters can help clinicians in the global evaluation. For example, according to the experts Figure 6 highlights a dense lymphoid tissue, bisected by a stromal streak and containing four lymphoid follicles, surrounded by interfollicular spaces. Moreover, it is possible to observe that three of the follicles are centered by rounded Germinal Centers (proliferating B cells).

3.3. Clusters analysis

After the pipeline is performed, it is possible to analyze the single identified clusters. We will show the results for 4 clusters that were labeled as “Endothelium” considering 4 different samples: L2, whole lymph node, T2, and CODEX, in order to show different organs and different datasets in terms of data acquisition.

The usual cluster report consists of 3 plots for each cluster, comprising: the most expressed markers for each cluster(Fig. 7),the expression of some selective diagnostic markers including transcription factors (Fig 8), and cluster location on UMAP plot and on the tissue (Fig 9). The last one also contained a summary of Tier 1 and Tier 2 markers, and was used by experts for a fast classification, given the fact that contained both spatial information and phenotype information. The first two were used to adjust the classification by analyzing deeply markers expression, either to better visualize the gaps between Tier 1, Tier 2, and the rest, or to inspect possible incongruities such as CD4 and CD8 both expressed by the same cluster. Cell type assignment was based on all of them and performed by experts.

3.4. Algorithm comparison

After running the complete BRAQUE analysis on the 8 datasets, the experts evaluated every cluster interpretability and most suited cell type (where possible). Then the Pheno- Graph algorithm was runned, the clusters analyzed similarly, and the two algorithms were compared.

Among all the datasets, experts identified a total of 41 cell

types, 11 of which were considered as infrequent due to their low total clusters count (≤10, among 620 clusters in total). The properties considered in the algorithms comparison were: number of different infrequent populations identified, number of different common populations identified, percentage of unclear clusters cells, percentage of T cells that are unseparated between CD4+ and CD8+ (meaning that within the same cluster we have CD4 positive T cells mixed with CD8 positive T cells.). Obviously the first two properties should be as high as possible, since identifying more cell types after experts validation is better. Moreover, this properties were considered since validating and counting cluster-wise cell types is easier to check than verifying the correct cell type, cell-wise, for more than a million cells. On the other hand, the last two properties should be as low as possible, given that clearly separated T cells are better than merged ones, and identifying a cluster of potentially interesting cells should ideally always end up being recognized as a cell type.

On this last point, an important focus should be pointed to the noise cluster that density-based clustering methods always have, i.e., the cluster labelled as “-1”. This cluster collects all the *unclear* cells, acting as a proper bin.

The noise-cluster of HDBSCAN helps removing cells that would average cluster properties, making them less clear. As long as this cluster’s size is relatively not big (e.g., <20% of the sample cells), this property was preferred by the experts as a tradeoff for having clearer and faster interpretation. Given that all of the

important cell types were found during the analysis, the noise cluster was not considered as a strong negative downside, while a strong negative property was labelling a cluster as potentially interesting and then having the experts labelling it as junk. Therefore the percentage of unclear clusters cells ignores the “-1” cluster, since it does not steal time for a useless evaluation by the experts.

Clustering comparison results are reported in Table 2.

4. Discussion

In single-cell analysis many efforts have been made to standardize the process of data analysis. However, even if the majority of the tools available shares the same goals, each type of technique presents its own peculiarities and involves different approaches.

In this paper we addressed the need to have a more tailored approach to analyze single-cell data coming from imaging technologies revealed with immunofluorescence technology. We proposed BRAQUE, an integrated and novel approach spreading from data pre-processing to phenotype classification. We tested it on lymphoid tissue (tonsils and lymph nodes) which is one of the densest and most challenging tissues [14].

We introduced two innovations that we hope will become standards in the field of single-cell analysis: to “fragment” data input distribution in order to help clustering even small differences in separate clusters, and the usage of effect size measures to rank markers according to their importance in

characterizing every cluster. The former could be performed in different ways. The aim of the method we chose (i.e., Lognormal Shrinkage) is not to correctly find all the subpopulations in each marker, but to guess where most relevant differences could lie and make them more evident. For instance, if two subsets are slightly different by just one marker, they will probably end up in the same cluster anyway. But if these hypothetical subsets are slightly different on 30+ markers, then our approach helps increasing such discrepancies, making more likely to split those subsets into separate clusters. Therefore, the key point is not finding the perfect subpopulation split, but rather finding multiple reasonable guesses, that will end up confirmed or not by UMAP embedding.

Thanks to its preprocessing, BRAQUE detects more clusters than other established methods such as PhenoGraph. As comparison between Figure 4 and 5 highlights, feeding UMAP with our preprocessed data results in a higher number of clearly isolated data clusters.

Since the number of total cell types in one lymph node detectable with these tech- nologies is difficult to estimate, we suggest an over-fragmentation rather than a under-fragmentation. This following the principle that it is easier to merge similar clusters than to split unclear clusters into clear subsets. Therefore, BRAQUE may be the most suited available tool for the task of cell type clustering, with no contraindications stopping it from being extended to other types of tissues acquired with similar techniques. Therefore we also strongly

suggest to privilege granularity over its opposite, in the development of new method that will come.

Another key advantage of BRAQUE is that in some cases, groups who dealt with analogous databases [35] [26] relied on prior knowledge of markers known to have high or low expression in specific cell types, discard of non-contributory markers, neighboring cell definition, hand-gating, etc. BRAQUE instead proposes a marker-agnostic and spatial- agnostic approach with top level granularity, which we believe is more adapted to analyze very dense and complex spatial protein data with no strong assumption or predefined bias.

An interesting further work could be applying only BRAQUE preprocessing (i.e., Lognormal Shrinkage) to the input of other clustering methods and assess the effective gain that could come from it.

Nowadays we too often see the usage of just p-values without their respective effect size. This trend should be highly reconsidered, since p-values are good only for boolean/threshold-like answers regarding statistical significance of possible differences. The problem is that in medicine, biology, and many other fields, after the statistical significance is achieved the important question becomes “How much is it different?”, and p-values are not adequate to address such an issue [36] [37] [38]. For this reason, when it comes to define which markers mostly characterize a cluster, we strongly suggest using effect size measures, whether it be of different or same nature as the one used in BRAQUE’s pipeline.

Furthermore, the choice to classify each cluster not just by listing absolute value normalized markers, but by ranking the most significant ones, has produced a more precise classification with more cell types and more accurate phenotypes according to the experts.

Lastly, since most of the computation time comes from the Bayesian Gaussian Mixture fitting procedure, it is important to reduce the number of maximum gaussians as much as possible as previously remarked in the methods section. But, in order to maintain the correctness of the approach and the quality of the results, this number should never or almost never “saturate” (meaning that the final number of fitted gaussians with a weight different from 0 is equal to the starting number of gaussians, thus implying that all gaussian components are useful for the final fit). This could imply that the possible ideal number of subpopulations in a distribution is higher than the found value, and the procedure could not be as efficient as it was shown in our results.

Table 1

1 A value is reported only if it does not vary over different datasets, if that is not the case a brief explanation of how to tune it is reported in the result section.

2 A range is reported only if different value could be suggested for different datasets.

(a) Bayesian Gaussian Mixture fit.

(b) Assignment step.

(c) Shrinkage step.

Figure 1. This Figure shows the Lognormal Shrinkage effect on the log2 of the vWF distribution. As we can see, the algorithm automatically selects the best number of components instead of

using all of the available ones. Then a clear separation between different subpopulation is achieved, in order to help further UMAP embedding in creating a more fragmented output for an easier and more precise clusterization.

Figure 2. This Figure shows the Lognormal Shrinkage effect on the original distribution of the vWF marker in the dataset L2. As we can see, the algorithm automatically selects the best number of components instead of using all of the available ones. Then a clear separation between different subpopulation is achieved, in order to help further UMAP embedding in creating a more fragmented output for an easier and more precise clusterization.

Figure 3. Distribution of the number of components selected for a single marker after the Bayesian Gaussian Mixture fit on dataset L2. Please notice how no marker had 15 components, showing that the upper limit of 15 given as input to the algorithm was appropriate.

Figure 4. HDBSCAN clusters computed over UMAP's embedding of dataset L2 (left) and reported on the real spatial coordinates of the cells (right).

Figure 5. UMAP's embedding of raw data in dataset L2. This figure should be compared with Figure 4 to understand the higher separation that is clearly achieved at a visual level by

UMAP's embedding thanks to the insertion of BRAQUE's preprocessing.

Figure 6. Global validation plot for dataset L2, list of most expressed markers per cluster (according to robust effect size measure). In this case the validation is not in depth but at a global level, since the clinicians can look for known biological structures.

Figure 7. This Figure reports markers expression of 4 clusters whose cells were identified as Endothelium. The markers are ranked according to the robust effect size metric *d-signed*, and colored according to their Welch t-test p-value. In every plot it is possible to notice gaps in the descending order of markers, such gaps were used to suggest important (i.e., "Tier 1") and possibly useful (i.e., "Tier 2") markers for the experts following classifications.

Figure 8. This Figure reports the comparison between whole dataset marker distribution (blue) and the cluster's one (red). This plot is intended to help the experts in comparisons, to better show them where the actual difference comes from, and help them estimating/validating the cellular type of the cluster. The difference on the x-axis range for CODEX dataset is a consequence of that data being acquired over 16-bits channels, while MILAN data were acquired with 8-bits channels, therefore they can span from to 0 to respectively $216 - 1$ and $28 - 1$.

Plots are resized consequently.

Figure 9. This Figure reports spatial positions in UMAP embedding (left plots) and in real space (right plots). This plots are enriched by the subtitle, which reports Tier 1 and Tier 2 markers, together with their effect size. This kind of plot was considered as the main tool for the experts to rapidly assign a cell type to the cluster, using all of the available information in one plot.

Table 2. Clustering table summarizing performances for BRAQUE (BR) and PhenoGraph (PH) on the 8 different datasets. Where a comparison can clearly be better or worse, the bold value indicates the best algorithm.

2This publicly available dataset had only usable 20 markers.

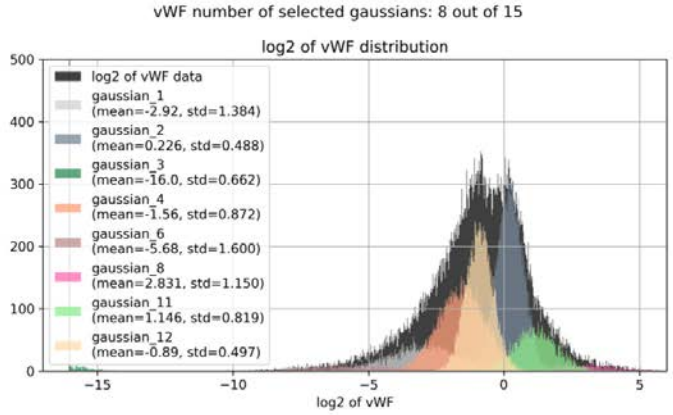
3This measure indicates the average of (BRAQUE metric - PhenoGraph metric) divided by the maximum possible for that specific metric, may it be 11 for Infrequent populations, 30 for Common populations, or 1 for percentages.

Table 1

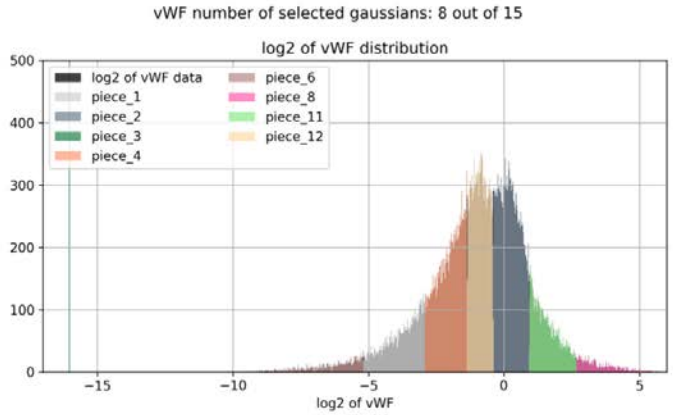
Algorithm	Parameter name	Suggested value[1]	Suggested range[2]
BGM	covariance_type	*full*	-
BGM	n_components	-	[10, 30]
BGM	tol	10^{-2}	$[10^{-5}, 10^{-1}]$
Lognormal Shrinkage	contraction factor	5	[2, 10]
UMAP	metric	*correlation*	angular metrics
UMAP	nn	50	[30, 500]
UMAP	min_dist	0.0	[0.0, 0.1]
UMAP	init	*spectral*	-
HDBSCAN	min_samples	-	[0.05%, 0.2%] of cells
HDBSCAN	cluster_selection_eps	-	[0.0, 0.3]
HDBSCAN	cluster_selection_met	*eom*	-

Table 1. .

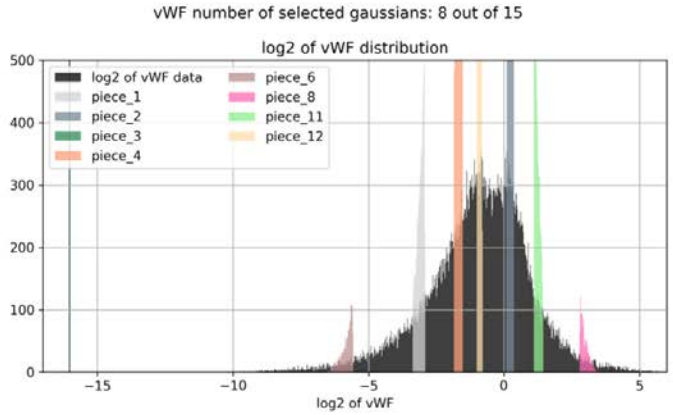
Fig.1



(a) Bayesian Gaussian Mixture fit.

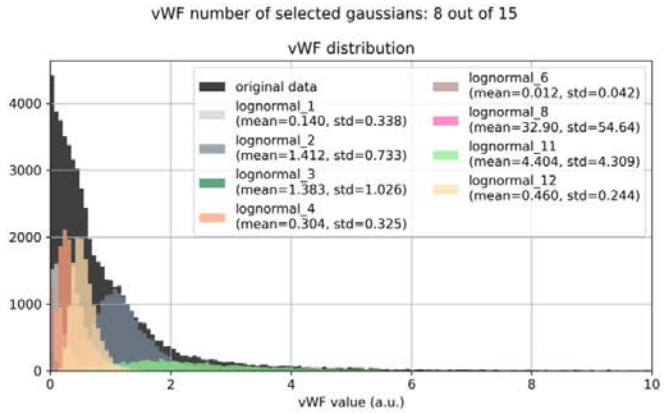


(b) Assignment step.

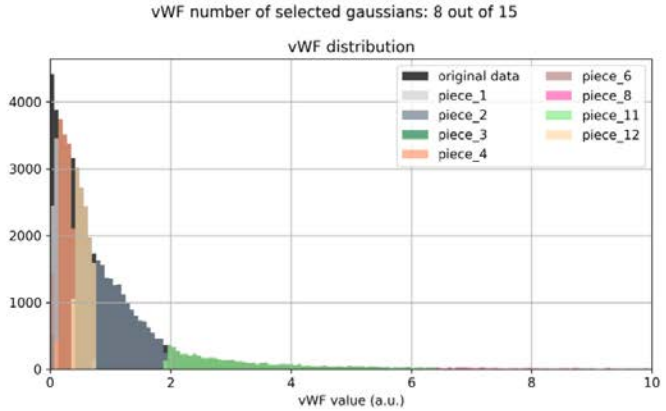


(c) Shrinkage step.

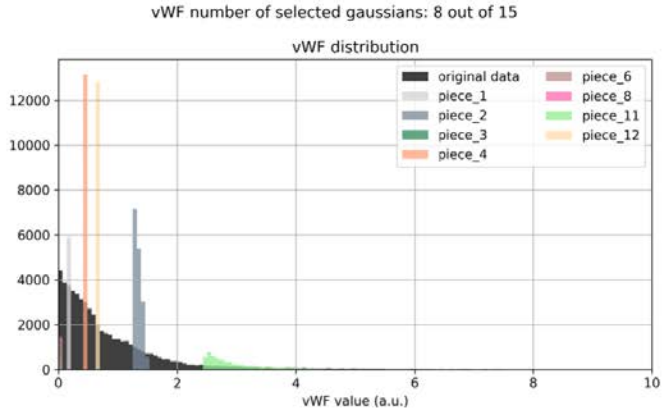
Fig.2



(a) Bayesian Gaussian Mixture fit.



(b) Assignment step.



(c) Shrinkage step.

Fig.3

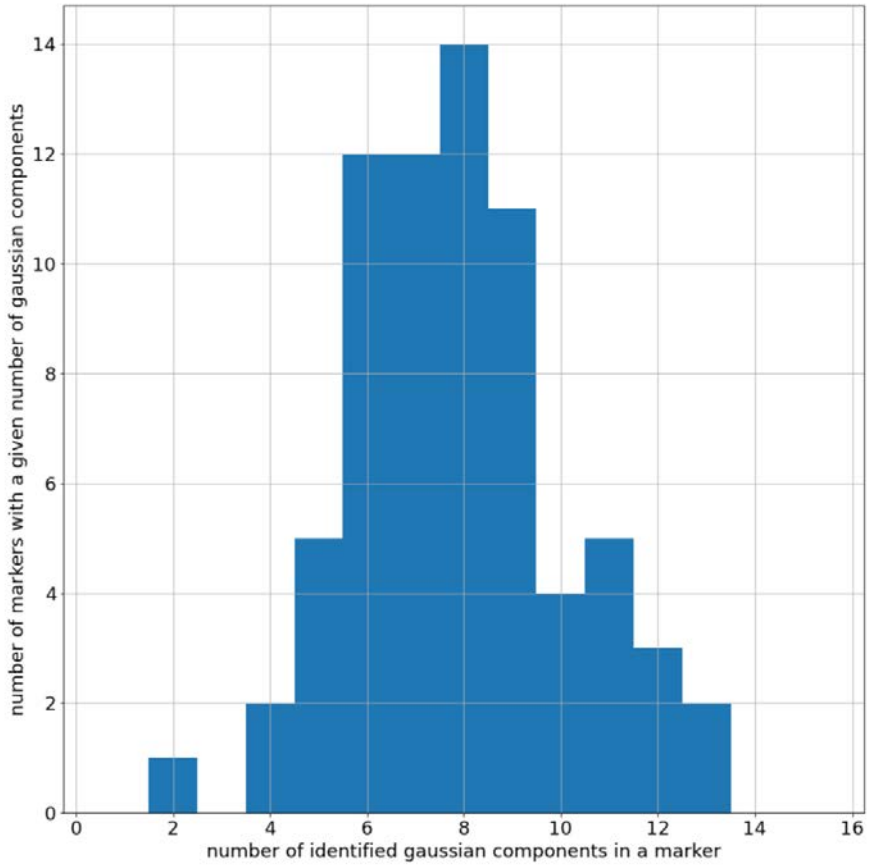

number of markers with a given number of gaussian components
number of identified gaussian components in a marker
0
2
4
6
8
10
12
14
16

Fig.4

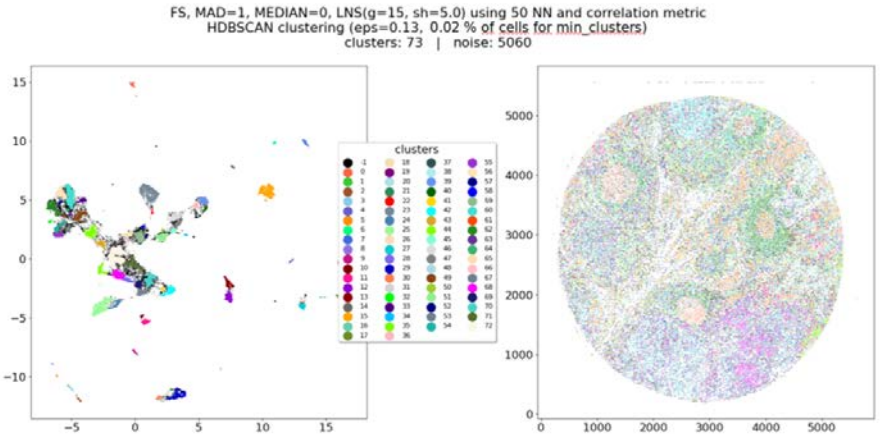

FS, MAD=1, MEDIAN=0, LNS(g=15, sh=5.0) using 50 NN and correlation metric
HDBSCAN clustering (eps=0.13, 0.02 % of cells for min_clusters)
clusters: 73 | noise: 5060
clusters

Fig.5

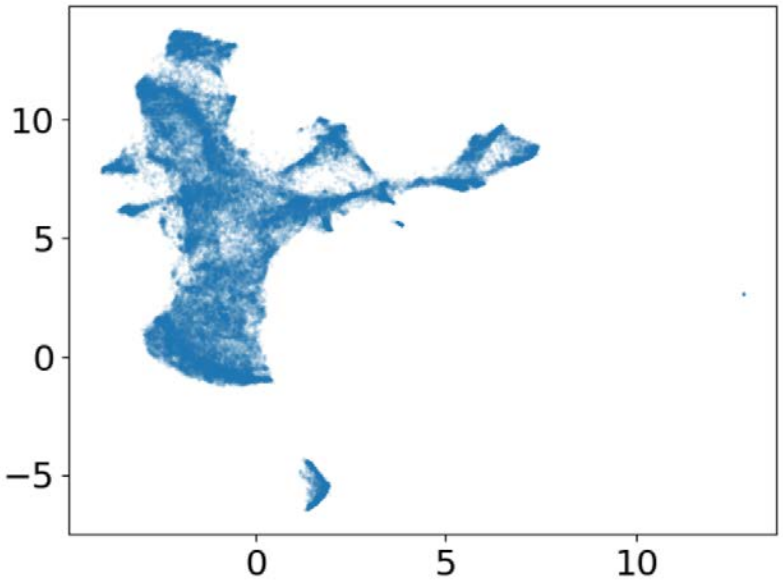

10
5
0
−5
0
5
10

Fig.6

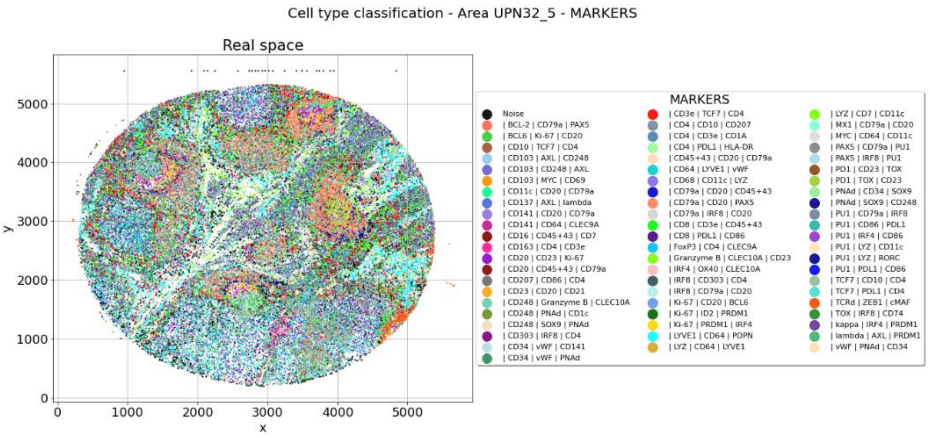
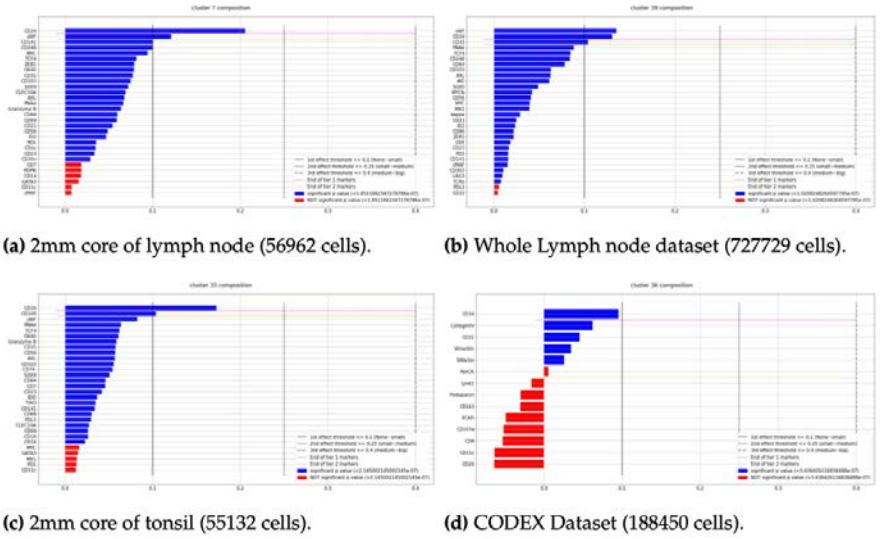

Cell type classification - Area UPN32_5 - MARKERS
Real space
5000
4000
3000
2000
1000
0
y
0
1000
2000
3000
4000
5000
x
MARKERS
Noise
| BCL-2 | CD79a | PAX5
| BCL6 | Ki-67 | CD20
| CD10 | TCF7 | CD4
| CD103 | AXL | CD248
| CD103 | CD248 | AXL
| CD103 | MYC | CD69
| CD11c | CD20 | CD79a
| CD137 | AXL | lambda
| CD141 | CD20 | CD79a
| CD141 | CD64 | CLEC9A
| CD16 | CD45+43 | CD7
| CD163 | CD4 | CD3e
| CD20 | CD23 | Ki-67
| CD20 | CD45+43 | CD79a
| CD207 | CD86 | CD4
| CD23 | CD20 | CD21
| CD248 | Granzyme B | CLEC10A
| CD248 | PNAd | CD1c
| CD248 | SOX9 | PNAd
| CD303 | IRF8 | CD4
| CD34 | vWF | CD141
| CD34 | vWF | PNAd
| CD3e | TCF7 | CD4
| CD4 | CD10 | CD207
| CD4 | CD3e | CD1A
| CD4 | PDL1 | HLA-DR
| CD45+43 | CD20 | CD79a
| CD64 | LYVE1 | vWF
| CD68 | CD11c | LYZ
| CD79a | CD20 | CD45+43
| CD79a | CD20 | PAX5
| CD79a | IRF8 | CD20
| CD8 | CD3e | CD45+43
| CD8 | PDL1 | CD86
| FoxP3 | CD4 | CLEC9A
| Granzyme B | CLEC10A | CD23
| IRF4 | OX40 | CLEC10A
| IRF8 | CD303 | CD4
| IRF8 | CD79a | CD20
| Ki-67 | CD20 | BCL6
| Ki-67 | ID2 | PRDM1
| Ki-67 | PRDM1 | IRF4
| LYVE1 | CD64 | PDPN
| LYZ | CD64 | LYVE1
| LYZ | CD7 | CD11c
| MX1 | CD79a | CD20
| MYC | CD64 | CD11c
| PAX5 | CD79a | PU1
| PAX5 | IRF8 | PU1
| PD1 | CD23 | TOX
| PD1 | TOX | CD23
| PNAd | CD34 | SOX9
| PNAd | SOX9 | CD248
| PU1 | CD79a | IRF8
| PU1 | CD86 | PDL1
| PU1 | IRF4 | CD86
| PU1 | LYZ | CD11c
| PU1 | LYZ | RORC
| PU1 | PDL1 | CD86
| TCF7 | CD10 | CD4
| TCF7 | PDL1 | CD4
| TCRd | ZEB1 | cMAF
| TOX | IRF8 | CD74
| kappa | IRF4 | PRDM1
| lambda | AXL | PRDM1
| vWF | PNAd | CD34

Fig.7



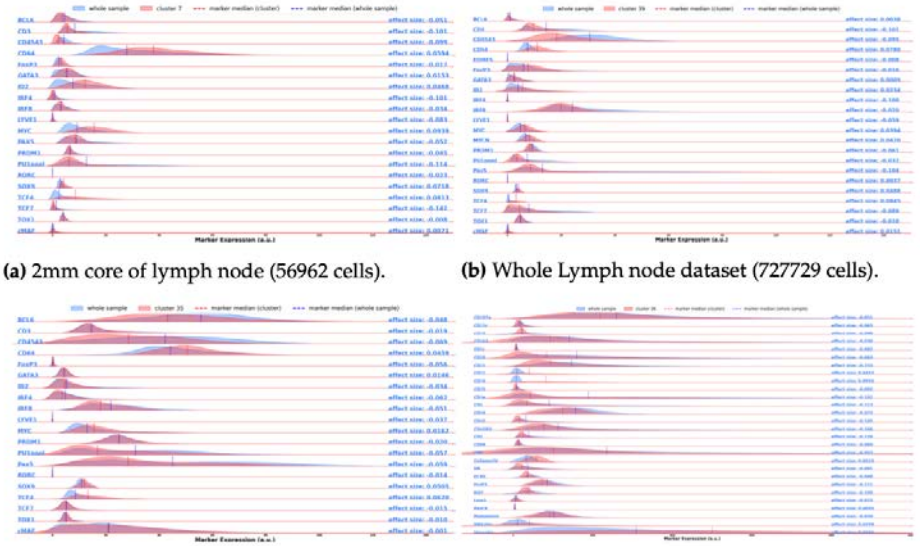
(a) 2mm core of lymph node (56962 cells).

(b) Whole Lymph node dataset (727729 cells).

(c) 2mm core of tonsil (55132 cells).

(d) CODEX Dataset (188450 cells).

Fig.8



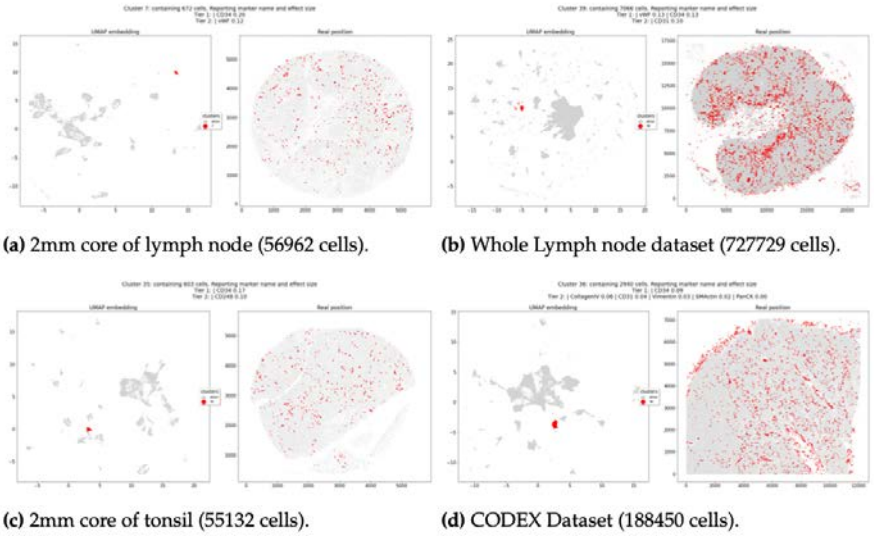
(a) 2mm core of lymph node (56962 cells).

(b) Whole Lymph node dataset (727729 cells).

(c) 2mm core of tonsil (55132 cells).

(d) CODEX Dataset (188450 cells).

Fig.9



(a) 2mm core of lymph node (56962 cells). **(b)** Whole Lymph node dataset (727729 cells).

(c) 2mm core of tonsil (55132 cells). **(d)** CODEX Dataset (188450 cells).

Table 2

Dataset	Whole Lymph node		L1		L2		L3		T1		T2		T3		CODEX[2]		Normalized
Algorithm	BR	PH	BR	PH	BR	PH	BR	PH	BR	PH	BR	PH	BR	PH	BR	PH	Average Difference (%)[3]
Infrequent populations (# out of 11)	**6**	3	0	**1**	0	**2**	2	2	0	**3**	**1**	0	**3**	1	1	**2**	-1.14%
Common populations (# out of 30)	**26**	17	**18**	14	**25**	16	**21**	14	**22**	16	**22**	20	**23**	15	**13**	12	+19.17%
% of unclear clusters cells	**0.59**	14.19	**9.21**	14.17	**0.43**	16.53	**1.09**	37.33	**0.25**	3.76	**6.25**	15.10	**0.00**	2.28	10.07	**0.00**	-9.44%
% of T cells not separated	**6.25**	31.82	39.85	**0.00**	0.00	0.00	2.94	**0.00**	**12.30**	23.36	**10.20**	89.99	**44.78**	82.63	4.56	**0.00**	-3.36%
number of clusters	120	48	55	32	73	29	54	27	80	33	62	35	82	33	96	22	X

Author Contributions: data acquisition, M.B. and G.Cat; cell-segmentation software, S.B.; BRAQUE software L.D.; results validation, M.B. and G.Cat; data curation, M.B.; writing—original draft prepa- ration, L.D., M.B., G.Cat., G.Cas.; writing—review and editing, L.D., M.B., S.B., G.Cat., G.Cas.; supervision, G.Cat. and G.Cas. . All authors have read and agreed to the published version of the manuscript.

Funding: This research was founded by Regione Lombardia

POR FESR 2014–2020, Call HUB Ricerca ed Innovazione: ImmunHUB to Giorgio Cattoretti.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Comitato Etico Brianza, N. 3204,

“High-dimensional single-cell classification of pathology (HDSSCP)”, October 2019.

Informed Consent Statement: Patients consent was obtained or waived according to article 89 of the EU general data protection regulation 2016/679 (GDPR) and decree N. 515, 12/19/2018 of the Italian Privacy Authority.

Data Availability Statement: CODEX Publicly available dataset analyzed in this study can be found here: https://portal.hubmapconsortium.org/browse/dataset/c95d9373d698faf60a66ffdc27499fe1. The dataset L1, L2, L3, T1, T2, and T3 presented in this study are available at: http://dx.doi.org/10.17632/j8xbwb93x9.1.

Acknowledgments: We Acknowledge for this study the EU Horizon 2020 programme (GenoMed4All project #101017549 to GC, HARMONY and HARMONY-PLUS project #116026 to GC), and the AIRC Foundation (Associazione Italiana per la Ricerca contro il Cancro; Milan, Italy; projects #26216 to GC.

Conflicts of Interest: The authors declare no conflict of interest. **Abbreviations**

The following abbreviations are used in this manuscript:

References

1. Schier, A.F. Single-cell biology: beyond the sum of its parts. *Nature Methods* **2020**, *17*, 17–20. https://doi.org/10.1038/s41592-019 -0693-3.
2. Savas, P.; Virassamy, B.; Ye, C.; Salim, A.; Mintoff, C.P.; Caramia, F.; Salgado, R.; Byrne, D.J.; Teo, Z.L.; Dushyanthen, S.; et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nature Medicine* **2018**, *24*, 986–993. Number: 7 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41591-018-0078-7.
3. Schiller, H.B.; Montoro, D.T.; Simon, L.M.; Rawlins, E.L.; Meyer, K.B.; Strunz, M.; Vieira Braga, F.A.; Timens, W.; Koppelman, G.H.; Budinger, G.R.S.; et al. The Human Lung Cell Atlas: A High-Resolution Reference Map of the Human Lung in Health and Disease. *American Journal of Respiratory Cell and Molecular Biology* **2019**, *61*, 31–41. Publisher: American Thoracic Society - AJRCMB, https://doi.org/10.1165/rcmb.2018-0416TR.
4. Cho, C.S.; Xi, J.; Si, Y.; Park, S.R.; Hsu, J.E.; Kim, M.; Jun, G.; Kang, H.M.; Lee, J.H. Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **2021**, *184*, 3559–3572.e22. https://doi.org/10.1016/j.cell.2021.05.010.
5. Zhu, C.; Preissl, S.; Ren, B. Single-cell multimodal omics: the power of many. *Nature Methods* **2020**, *17*, 11–14. Number: 1 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41592-019-0691-5.

6. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M.; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated analysis of multimodal single-cell data. *Cell* **2021**, *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
7. Jain, M.S.; Polanski, K.; Conde, C.D.; Chen, X.; Park, J.; Mamanova, L.; Knights, A.; Botting, R.A.; Stephenson, E.; Haniffa, M.; et al. MultiMAP: dimensionality reduction and integration of multimodal data. *Genome Biology* **2021**, *22*, 346. https://doi.org/10.1186/s13059-021-02565-y.
8. Rozenblatt-Rosen, O.; Stubbington, M.J.T.; Regev, A.; Teichmann, S.A. The Human Cell Atlas: from vision to reality. *Nature* **2017**, *550*, 451–453. https://doi.org/10.1038/550451a.
9. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **2021**, *598*, 86–102. https://doi.org/10.1038/s41586-021-03950-0.
10. Snyder, M.; Lin, S.; Posgai, A.; Atkinson, M.; Regev, A.; Rood, J.; Rozenblatt-Rosen, O.; Gaffney, L.; Hupalowska, A.; Satija, R.; et al. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **2019**, *574*, 187–192. https://doi.org/10.1038/s41586-019-1629-x.
11. Hor, J.L.; Germain, R.N. Intravital and high-content multiplex

imaging of the immune system. *Trends in Cell Biology* **2022**, *32*, 406–420. https://doi.org/10.1016/j.tcb.2021.11.007.

12. Denisenko, E.; Guo, B.B.; Jones, M.; Hou, R.; de Kock, L.; Lassmann, T.; Poppe, D.; Clément, O.; Simmons, R.K.; Lister, R.; et al.

Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biology* **2020**, *21*, 130. https://doi.org/10.1186/s13059-020-02048-6.

13. Zappia, L.; Theis, F.J. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. *Genome Biology* **2021**, *22*, 301. https://doi.org/10.1186/s13059-021-02519-4.

14. Cupedo, T.; Stroock, A.; Coles, M. Application of tissue engineering to the immune system: development of artificial lymph nodes. *Frontiers in Immunology* **2012**, *3*.

15. Krishnamurty, A.T.; Turley, S.J. Lymph node stromal cells: cartographers of the immune system. *Nature Immunology* **2020**, *21*, 369–380. https://doi.org/10.1038/s41590-020-0635-3.

16. Massoni-Badosa, R.; Soler-Vila, P.; Aguilar-Fernández, S.; Nieto, J.C.; Elosua-Bayes, M.; Marchese, D.; Kulis, M.; Vilas-Zornoza, A.; Bühler, M.M.; Rashmi, S.; et al. An Atlas of Cells in the Human Tonsil, 2022. Pages: 2022.06.24.497299 Section: New Results,

https://doi.org/10.1101/2022.06.24.497299.

17. Satija, R.; Farrell, J.A.; Gennert, D.; Schier, A.F.; Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature*

Biotechnology **2015**, *33*, 495–502. Number: 5 Publisher: Nature Publishing Group, https://doi.org/10.1038/nbt.3192.

18. Dries, R.; Zhu, Q.; Dong, R.; Eng, C.H.L.; Li, H.; Liu, K.; Fu, Y.; Zhao, T.; Sarkar, A.; Bao, F.; et al. Giotto, a toolbox for integrative analysis and visualization of spatial expression data, 2020. Pages: 701680 Section: New Results, https://doi.org/10.1101/701680.

19. Levine, J.H.; Simonds, E.F.; Bendall, S.C.; Davis, K.L.; Amir, E.a.D.; Tadmor, M.D.; Litvin, O.; Fienberg, H.G.; Jager, A.; Zunder, E.R.; et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **2015**, *162*, 184–197. https://doi.org/10.1016/j.cell.2015.05.047.

20. Efremova, M.; Teichmann, S.A. Computational methods for single-cell omics across modalities. *Nature Methods* **2020**, *17*, 14–17. https://doi.org/10.1038/s41592-019-0692-4.

21. Jovic, D.; Liang, X.; Zeng, H.; Lin, L.; Xu, F.; Luo, Y. Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and Translational Medicine* **2022**, *12*, e694. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ctm2.694, https://doi.org/10.1002/ctm2.694.

22. Harris, C.; Wrobel, J.; Vandekar, S. mxnorm: An R Package

to Normalize Multiplexed Imaging Data. *Journal of open source software* **2022**, *7*, 4180. https://doi.org/10.21105/joss.04180.

23. Bolognesi, M.M.; Manzoni, M.; Scalia, C.R.; Zannella, S.; Bosisio, F.M.; Faretta, M.; Cattoretti, G. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society* **2017**, *65*, 431–444. https://doi.org/10.1369/0022155417719419.
24. Manzoni, M.; Bolognesi, M.M.; Antoranz, A.; Mancari, R.; Carinelli, S.; Faretta, M.; Bosisio, F.M.; Cattoretti, G. The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas. *Scientific Reports* **2020**, *10*, 702. Number: 1 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41598-020-57627-1.
25. Bolognesi, M.M.; Mascadri, F.; Furia, L.; Faretta, M.; Bosisio, F.M.; Cattoretti, G. Antibodies validated for routinely processed tissues stain frozen sections unpredictably. *BioTechniques* **2021**, *70*, 137–148. Publisher: Future Science, https://doi.org/10.2144/ btn-2020-0149.
26. Hickey, J.W.; Tan, Y.; Nolan, G.P.; Goltsev, Y. Strategies for Accurate Cell Type Identification in CODEX Multiplexed Imaging Data. *Frontiers in Immunology* **2021**, *12*.
27. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, 2020. arXiv:1802.03426 [cs, stat].

28. Wu, D.; Poh Sheng, J.Y.; Su-En, G.T.; Chevrier, M.; Jie Hua, J.L.; Kiat Hon, T.L.; Chen, J. Comparison Between UMAP and t-SNE for Multiplex-Immunofluorescence Derived Single-Cell Data from Tissue Sections. preprint, Bioinformatics, 2019. https://doi.org/10.1101/549659.
29. Yang, Y.; Sun, H.; Zhang, Y.; Zhang, T.; Gong, J.; Wei, Y.; Duan, Y.G.; Shu, M.; Yang, Y.; Wu, D.; et al. Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *Cell Reports* **2021**, *36*, 109442. https://doi.org/10.1016/j.celrep.2021.109442.
30. Dogan, A.; Birant, D. K-centroid link: a novel hierarchical clustering linkage method. *Applied Intelligence* **2022**, *52*, 5537–5560. https://doi.org/10.1007/s10489-021-02624-8.
31. Schubert, E.; Sander, J.; Ester, M.; Kriegel, H.P.; Xu, X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems* **2017**, *42*, 1–21. https://doi.org/10.1145/3068335.
32. Campello, R.J.G.B.; Moulavi, D.; Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In Proceedings of the Advances in Knowledge Discovery and Data Mining; Pei, J.; Tseng, V.S.; Cao, L.; Motoda, H.; Xu, G., Eds.; Springer: Berlin, Heidelberg, 2013; Lecture Notes in Computer Science, pp. 160–172. https://doi.org/10.1007/978-3-642-37456-2_14.
33. Li, L.; Xi, Y. Research on Clustering Algorithm and Its Parallelization Strategy. In Proceedings of the 2011 International Conference on Computational and Information

Sciences, 2011, pp. 325–328. https://doi.org/10.1109/ICCIS.2011.223.

34. Vandekar, S.; Tao, R.; Blume, J. A Robust Effect Size Index. *Psychometrika* **2020**, *85*, 232–246. https://doi.org/10.1007/s11336-020 -09698-2.
35. Zhang, W.; Li, I.; Reticker-Flynn, N.E.; Good, Z.; Chang, S.; Samusik, N.; Saumyaa, S.; Li, Y.; Zhou, X.; Liang, R.; et al. Identification of cell types in multiplexed in situ images by combining protein expression and spatial information using CELESTA. *Nature Methods* **2022**, *19*, 759–769. Number: 6 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41592-022-01498-z.
36. Sullivan, G.M.; Feinn, R. Using Effect Size—or Why the P Value Is Not Enough. *Journal of Graduate Medical Education* **2012**, *4*, 279–282. https://doi.org/10.4300/JGME-D-12-00156.1.
37. Klein, D.F. Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research. *American Journal of Psychiatry* **2005**, *162*, 643–a. Num Pages: 644 Publisher: American Psychiatric Publishing, https://doi.org/10.1176/appi.ajp.162.3.643-a.
38. Cohen, J. Things I have learned (so far). In *Methodological issues & strategies in clinical research*; American Psychological Association: Washington, DC, US, 1992; pp. 315–333. https://doi.org/10.1037/10109-028.

Chapter 6

Spatial single cell high-plex imaging classifies normal lymphoid tissue.

Bolognesi M.M.[1], Dall'Olio L.[2], Manzoni M.[3] , Borghesi S.[4], Castellani G.[2], Bosisio F.M. [5], Mascadri F.[1], Faretta M.[6], Cattoretti G [1].

1 Department of Medicine and Surgery, University of Milano Bicocca, Monza, Italy

2 Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy

3 Department of Pathology, ASST Bergamo Est, Bergamo, Italy

4 Department of Mathematics and Applications, University of Milano Bicocca, Milan, Italy

5 Translational Cell and Tissue Research Unit, Department of Imaging and Pathology, KU Leuven, Leuven, Belgium

6 Department of Experimental Oncology, European Institute of Oncology IRCCS, Milan, Italy

manuscript in preparation

Introduction:

Lymph nodes are the districts where the immunological response is orchestrated. In the human body they are almost 500-600, distributed in the whole body. In standard condition they are composed of different cell types compartmentalized in lobulated and referenced areas (Willard-Mack, 2006).

Differently from any other organ, whose microscopic aspect, architecture and size are invariant across individuals at any given adult time, lymph nodes in the homeostatic state are bathed in a variety of stimuli (e.g. antigens) which cause a

variability in cell composition and architecture, even without changing the node size.

The cell types involved can be classified in two complementary and heterogeneous groups that continually interact and influence themselves reciprocally (Honan and Chen, 2021): the stromal part (Krishnamurty and Turley, 2020) (Chang and Turley, 2015) to which belong the endothelial and mesenchymal compartment and the hematopoietic one (Efremova et al., 2020) (Domínguez Conde et al., 2022) to which belong lymphoid (T cell, B cell, NK cell) and myeloid cells (Monocytes, Macrophages and Dendritic cells).

Stromal cells have a structural role and are involved in recruitment of immune cells and in their maintenance.

During the immune activity, the LN stromal compartment expands and reorganizes (Krishnamurty and Turley, 2020).

On the hematopoietic side instead (Efremova et al., 2020), cells are more heterogeneous and present different properties in order to face homeostasis from one side and to respond to stimuli in a specific and dynamic way to the other.

Defining “normal condition” of a lymph node and, in general, of a certain type of tissue, is complex. How many samples should be collected to encompass “normality”? According to HCA (Rozenblatt-Rosen et al., 2017) community (https://www.humancellatlas.org/wp-content/uploads/2021/10/25-OCT-2021-The-Commitment-of-the-Human-Cell-Atlas-to-Humanity-1.pdf) it is important to catalog biological and cellular variation, which is one of their goal.

Some global projects (HuBMAP Consortium et al., 2019) (Rozenblatt-Rosen et al., 2017) are already facing sample heterogeneity and their databases are continually expanding because of the acquisition of new dataset available to the scientific community.

Lymphoid tissue, is underrepresented in public databases (except tonsils, (Massoni-Badosa et al., 2022)) and has been characterized with sample tonsil (Massoni-Badosa et al., 2022) or few lymph nodes samples (The Tabula Sapiens Consortium and Quake, 2021) or with relatively few markers ((HuBMAP Consortium et al., 2019) and a deep and spatial characterization at single cell level is still missing.

Typically immune cell data come from blood and are not evaluated within their tissue context.

Lastly the clinical approach of pathologists exploring lymphoid tissue is usually focused on detection of abnormalities (e.g. metastasis in lymph node sentinel) and does not linger on the normal part.

The advent of single cell biology has conspired to deeply characterize cell types in many organs, in a precise condition or in a trajectory (Saelens et al., 2019).

Lymph nodes have an intrinsic density ($1mm^3$ of lymph node tissue approximately 1 to 2×10^6 cells) (Cupedo et al., 2012) and consent an easy disaggregation for single cell RNA sequencing for recirculating cells but not for sessile elements (stromal cells, some macrophages, dendritic cells etc.), making difficult the complete analysis of all cell types .

For this reason spatial approaches (Moffitt et al., 2022), which preserves tissue integrity, should be preferred. Most importantly, localizing cells within their environment adds a new level of comprehension of cell localization and cell-cell interaction.

In this study we classify at single cell level 38 human normal (i.e. non pathologic) lymph nodes and 3 tonsils. Samples were processed with immunofluorescence-based MILAN (Bolognesi et al., 2017) technology with more than 80 markers and analyzed with bioinformatic tools available in the public domain.

Populations have been annotated by evaluating expression of marker intensity within each cluster and their spatial distribution within the sample.

A subset of samples was analyzed using the BRAQUE algorithm (Manuscript submitted) to deeply characterized cell populations and distribution. The final goal is to create an annotated reference datasets to improve the phenotypic cell classification and, in the end, integrate population with other, where available, omics data.

As a result of this comprehensive approach new marker positivities were associated with novel cell subsets/clusters.

Limitations of study are due to the scarcity of specific markers for the detection of stromal cell varieties and other rare populations such as B cell precursor and innate lymphoid cell subsets.

Results:

Results of individual antibody staining (Table 1 and Supplemental Table1) of a representative lymph node can be visualized as correlation matrix versus all the other antibodies. Fig1 also contains information about the channel of acquisition of the image and relative number of the round.

The matrix (Fig.1) helps understanding co-expression or avoidance of markers used. Notably, T cell and B cell markers clusterize separately in the table. Another cluster, which contains among the others CD7, CD23 and GZMB, shows previously unknown correlations which we will be explored in detail in the single cell analysis.

More than 1,2 million cells have been globally analyzed. Lymph Nodes cells were analyzed separately from tonsils cells.

Mean cell number analyzed per core is 55,048 (n. =37 ; ± 12,835 SD) (Supplemental Table 2) A whole lymph node contains 727,549 cells.

Common known population and relative % were classified with Seurat in tonsil and lymph nodes. (Fig 2-3). Seurat is the package of choice to integrate multiple samples: 37 lymph nodes produced 53 clusters (Fig. 2), 3 tonsils 32 (Fig. 3), identifying the vast majority of lymphoid tissue constituents.

Lymph Nodes:

T cells and Dendritic cells were identified on average by one single cluster per cell type, however B cell, Monocytes-macrophages and Stromal cell subsets were represented by more than one, suggesting the presence of further final subdivision not easily evident by Seurat clustering.

Notably some novel cell populations were found: a MYC positive B cell subset , macrophages expressing CD7 and CD23 and 2 subsets of stromal cells expressing SOX9 and TCF4.

Some clusters were unclassified and some minute ones were not contributory (not shown).

Similar findings were found in tonsils. Differences, shown by abundance of B cell populations, reflect the usual B cell activation of this tissue.

Since Seurat pipeline of analysis is not a tailored algorithm for highly multiplexed imaging of fluorescent-based data sets, we applied BRAQUE (Bayesian Reduction for Amplified Quantization in Umap Embedding; manuscript in preparation), which is native to spatial analysis, to a subset of samples (3 cores of lymph nodes, 3 of tonsils and 1 whole lymph node).

This allowed us to improve the clustering and have more granular data based on spatial visualization and highlighting markers uniquely and statistically significantly expressed in each cluster (mean 68 cluster/cores ± 11SD, (Fig. 4) [see previous chapter BRAQUE]. Afterwards, multiple clusters with a similar phenotype were merged into a single cell type with the purpose of cell classification.

Across all samples 41 cell types have been identified (Supplemental Fig 1). 8 subsets of T cells, 11 subsets of B cells (including plasma cell), 9 subsets with myelo/mono signature, 5 subsets of Dendritic cells and 7 subsets of

Stromal/Endothelium populations. Many of these have been published extensively (Szabo et al., 2019a) (Crotty, 2014a).
In the end BRAQUE has allowed to find new populations, which will be described in more detail (see Table 2):

T Cells: In Lymph nodes we found CD8+ activated T cell bearing activation markers and evidence of immune checkpoint engagement (EOMES, TIM3, IDO) such as when responding to a viral insult. Activated T cells may associate with a novel macrophage subtype (see below). They were scattered throughout the lymph node and focally in correspondence of an area with blast, high endothelial venules and hemorrhagic effusion. (Fig 4D ans Supplemental Fig 3A).
CD4+, T reg, CD8+ (both S100+ and negative), extrafollicular and follicular T Helper cells were localized in tissue as published(Szabo et al., 2019b) (Crotty, 2014b).

B cells: The following populations were found:
virgin IgD+ Mantle Zone B cells,
A proliferating population characterized by ID2 and occasionally Pax5 is diffusely present interstitially and with focal colonization of Germinal Centers (Figure 4 F). This may be the earliest antigen activated B cell type colonizing the germinal center.
With a similar location but restricted to the outer cortex we found a scattered cell type bearing a phenotype of Centroblasts (Stoler-Barak et al., 2019) which also aggregates

in the germinal center. They show variable combinations for Ki-67, CD30, AID, BCL6, MYC as markers for early activation.
As expected we found centrocytes within the germinal center. A small portion was distributed in medullary coords.
Post-Germinal Center “products” found, outside the follicle, are Plasma cells at different stages of terminal differentiation in the medullary cords and marginal zone B cells closer to the follicle. The former (Plasma cells) show different levels of positivities for PRDM1 and IRF4 and a decrease of PAX5 and PU1 signals. We could identify 3 types of post Germinal center antibodies secreting cells: 1) Plasmablast (IRF4+, PRDM1+, light chain negative) interstitial, 2) Plasma Cells precursors (Light chain + but with an intermediate/low level of Bcell Marker PAX 5 and PU1 and Plasma Cell markers IRF4 and PRDM1) and 3) Plasma Cells. They are located in the medullary cords as expected. However a portion of plasmablast localized in T cell areas in the outer cortex (Supplemental Figure 2).
Spatial allocation of Mantle Zone and Marginal Zone subsets show the expected distribution of Marginal cells outside of the Mantle Zone but at the same time a frequent admixture in the follicle of the two (Figure 4E and Supplemental Figure 3).

Mono/Macrophages: 9 subsets were found (Supplemental Figure1). Conventional CD16lowCD14highLYZ+ monocytes and alternative CD16highCD14low were found largely colocalizing. TissueMacrophages/histiocytes were classified for the presence of mature markers (CD64, CD68, CD163) and can be further

subdivided according to the presence of, LYVE 1 and CD68 as unique marker (phagocytes). A new population CD23+ CD7+ with TIM3 e IDO (partially independent from CD16) ,was found associated with activated CD8 T Cells. In the marginal and medullary sinuses there are some cell populations characterized by close contact between endothelial cells and medullary sinus macrophages with an unclear phenotype.

DC: classical DC1 and DC2 and pDC (Supplemental Figure 2) were identified by the canonical markers.
cDC1 were divided in two subsets and distributed in 2 distinct areas partially overlapping, one more peripheral and the other more central. This probably reflects the different intensities of the panel components (Figure 4A).
A poorly defined population, non T and non B cells, characterized by PU1 and with a variable combination of other dendritic cell markers at low level, IRF4, IRF8, ID2 and Ki-67± could be assigned to the early dendritic lineage. We confirmed this hipotesis because this population is distributed in areas closer to classical dendritic cells (Figure 4 B).
AXL, putatively expressed on DC precursors, was significantly observed only in macrophages subsets.
cDC2 are distributed in part with cDC1 but they are in touch with histiocytes but not phagocytes (Figure 4C).

Endothelial cells are divided in Vascular (vW+) and lymphatic (LYVE1 +) clusters (Figure 4G and Supplemental Figure 2).

Stromal cells (SOX9+, TCF4+, PDPN), divided in 2 subsets by the expression of MYC or NMYC, are localized in all lymph nodes and in particular closer to large and medium size vessels.

Epithelial cells were detected, even if restricted epithelial markers (keratin) were not included in the panel, because of MYC and PRDM1 positivity, absence of CD45 and the spatial distribution.

Besides cells discarded upfront by BRAQUE based on the specifics of the algorithm, 1-8 % of cells were excluded because of technical reasons (artifacts, scratches..). An example of population that BRAQUE allows to identify and exclude is in Figure 4 H.

Discussion:

In this paper we presented the first spatial single cell classification of 2 different types of “normal” lymphoid tissue (tonsil and lymph node) based on high-plex immunofluorescence data.

The high number of samples (41) and markers (>80) has allowed us to identify and localize all commonly known cell types plus new subsets which need to be further investigated. In addition, plotting the populations on the tissue has let us infer some relationships between cells in spatial context.

The analysis has been performed with two different bioinformatic tools. Seurat was chosen for its data integration approach, in order to verify data integrability and major population presence. As said before, it is a tool for the exploration of single cell RNA-seq data and not for IF based high-plex data analysis. However we obtained from the integration of samples (cores) all known major cell types.

Clusters obtained appeared robust as suggested by precedent comparison of protein and single cell RNA high dimensional analysis where T cells were better separated/segregated (Hao et al., 2021). However, because of insufficient granularity of results (53 clusters in lymph nodes and 32 in tonsils) a deeper approach was preferred in order to have a more detailed description of the populations .

BRAQUE privileges a single sample analysis. This is computationally more affordable and to be preferred for samples where the numerosity of cells greatly exceeds the numerosity of markers and because of this, integration is not required to increment data set numerosity.

In this way we also skip batch effects during the analysis and data flattening to respect sample integrity.

Thus, merging clusters after annotation (Meta-Clustering) is more appropriate and consent to respect data variability.

Lastly BRAQUE layout (Suppl. Fig. 4) has the possibility to explore each cluster by showing together in the same box 1) cluster key markers expression (statistically significant compared to others), 2) spatial distribution of cluster within the

sample, 3) cluster position on Umap plot and 4) lineage specific and Transcription factor markers expression of cluster compared to sample (Supl. fig 4) .

This helps the pathologist or the researcher in superived cluster assignment.

The procedure, in fact, if compared to single cell RNA sequencing data, is different. In the high plex IF data field there are not yet dedicated tools such as Cell Typist (Domínguez Conde et al., 2022) or Azimuth (Hao et al., 2021) for labeling by querying an annotated repository; in our case cluster data should be compared with literature and repositories such as https://www.proteinatlas.org/ , https://tabula-sapiens-portal.ds.czbiohub.org or published catalogs such as (Kalina et al., 2019) for the assignment.

This scarcity of tools depends on the availability of an annotated data set for the comparison, obtained with a similar technology and on the same organ or cell type (Liu and Zhang, 2022).

Results with BRAQUE, even if the cluster assignment is more tedious than Seurat because of the higher numerosity, appear more granular and show new populations.

In the B cell subset some centroblasts were found not in the germinal center but outside, differently from the expected. They could be precursors or eventually the subset, already described in mice (Stoler-Barak et al., 2019), associated with germinal center activity also evident in tonsils where germinal centers are very active..

Marginal zone B cells and plasma cells are distributed in largely not overlapping areas.

Unexpectedly, marginal zone and mantle zone B cells do overlap, to a greater extent than expected (Fig.4 E). In the past, marginal zone B cell benign expansions have been termed monocytoid B cells (Stein et al., 1999), because of the characteristic cell morphology (clear cytoplasm, slightly irregular nucleus etc.); we looked at a serial H&E section of areas where MZ and MargZ commix and indeed the two distinctive cell types can be identified by an expert's eye.

We identified plasma blasts (characterized by the absence of light chain and the presence of plasma cell differentiation TF) although not in all samples. Notably a large portion was localized in the T cell zone, differently from the mature plasma cell and other maturing B cells. The significance of this new finding is still unclear. It may represent extrafollicular B cell activation.

Some of the markers in our panel have been often incorrectly used to subdivide macrophages into M1 and M2 subsets (Mantovani et al., 2004), however we didn't reproduce this subdivision. We found different populations, because of phenotypes and also because of spatial distribution (phagocytes, histiocytes, macrophages. Table 2).

Among them a new population with a precise phenotype, CD23+ and CD7+, was found. It probably reflects selective stimuli (i. e. bacteria, IL4, etc.) which can induce one but not others of the markers within the subset. Evidence of this

selective induction can be found in the experiment databases upon which are based RNA repositories suche the one BioGPS data.

CD23 is induced upon IL4 treatment of monocytes (Vercelli et al., n.d.).

CD7 has been previously described in tissue myeloid cells(Krop et al., 2022).

Two classical populations of Dendritic cells were found in different portions (peripheral and central) of the Lymph Node. They showed a different repertoire of C-lectins and expression of molecules of the immunological synapse which interact with T cells (IDO, HLA-DR and CD4). Further neighborhood analysis of the other DC subsets we have identified will clarify their relationship with the rest of the immune system.

Despite the restricted number of specific markers (Keratin, Smooth muscle actin, Vimentin and Collagen were absent), in particular for non-vascular stromal, surprisingly we identified the LN scaffold (Fig. 4G) and a subset of stroma with a peculiar phenotype showing SOX9, TCF4 and MYC family members.

Some clusters remained unclear, despite unique spatial distribution, and this can be explained by different factors. Some limitations are due to segmentation. It's common to detect signals from neighbor cells, because of the high density in lymphoid tissue. This can confound the identification of cell types but at the same time can help in the identification of close neighbors cells.

We tested other segmentation approach (i.e. CellPose (Stringer et al., 2021) not shown) but the algorithm we are currently using seems superior.

Panel choice also can affect the results and some populations can be more finely subdivided than others.

In fact, even if 80 antibodies per cell are a large number, single cell RNA seq data contain a much larger number of variables in comparison from which one can easily identify mutually exclusive signals per cluster. Our panel instead contains lineage oriented membrane markers, helpful to separate distinct populations, but also TF, which are broadly expressed therefore and it is not possible to find a combination of unique markers per cluster. For example, because of the absence of markers for functional characterization (except PD1, TOX, TIM3 for exhaustion or activation), a finer subdivision for T cell wasn't possible.

Another limitation is due to the lack of neighborhood analysis.

One approach has been to group into cell communities similar cells in close contact (Bhate et al., 2022). However this approach will not work with a population with a scattered distribution among other dissimilar cells.

Knowing this limitation we will explore other approaches.

At the moment we are not aware of similar studies with a larger number of markers on several samples of lymphoid tissue.

With these limitations we identified cell types reproducible on the whole data set but, as already noticed, defining "normalcy "

is difficult due to the intrinsic and dynamic nature of lymphoid tissue analyzed in such fine detail.

Data integration with other omic-technology is the next step we are going to explore.

Data integration became important because not only match protein signals to its RNA counterpart but identify populations directly in the tissue. Localization is crucial to locate cell populations within their context, to analyze neighbor cells and relation with tissue architecture.

In this landscape data integration (Luecken et al., 2022) and cell type classification represent the real challenges of the single cell biology era.

Materials and methods:

Data set:

Anonymous normal lymph nodes and tonsils were obtained from the San Gerardo Hospital Pathology Department. They were diagnosed as “healthy” by the pathologist. Lymph Nodes belong to 6 sites: axillary, inguinal, submandibular, laterocervical, iliac, mesenteric and were represented either from selective staging for a solid tumor (e.g. breast, cervix) or as an incidental finding during surgery. They measured 1.75 cm ± 0.76 in the greatest diameter.

After sample collection and routine processing (FFPE) they were organized in 3 different TMA (Tissue Microarray), with cores of 2 mm of diameter or as whole sections.

The TMA was constructed with a Tissue Microarrayer Galileo CK4500 (Tissue Microarrayer Model TMA Galileo CK4500; Integrated Systems Engineering srl, Milano, Italy).

MILAN technology (Bolognesi et al., 2017) was applied with more than 80 antibodies in different rounds of immunofluorescence stainings. The antibody panel was validated in multipleTh ways (see Table 1).

Antibodies used for the panel are listed in Table 1. It is organized as follows: a detailed list of antibodies used with relative info about their localization expression within the cell (Nuclear, Cytoplasmic or Membrane), their significance, their abundance, a link to Human Protein Atlas (to check their distribution across tissues).

Images were acquired with a Nanozoomer S60 scanner at 20x (final magnification 200x, 45µm/pixel) . Virtual slides in .svs file extension were converted in .tiff and their alignment was performed with a dedicated Fiji pipeline ((Furia et al., 2013)

A segmentation mask was created on DAPI staining with CyBorgh algorithm as published (see BRAQUE, manuscript submitted). Contextually mean values of all markers were collected from cells segmented on all coregistered images and data were organized in a .csv file together with spatial coordinates of cells.

Data were explored with the Seurat package (Hao et al., 2021) by adapting the csv as a Seurat Object. They were scaled, PCA was run. Anchors were found on “rpca” reduction. “Find cluster” function was applied with “resolution= 1”.

Supervised assignment of each cluster was based on the global heatmaps produced.

BRAQUE algorithm (manuscript submitted) was applied on a restricted data set, 3 cores of lymph nodes, 1 whole Lymphnode and 3 cores of tonsils, in order to deeply explore cell population. Supervised assignment of clusters was based on BRAQUE results which include 1) Cluster location in the UMAP plot 2) Cluster distribution on tissue 3) a bar plot ranking the most statistically significant markers within that cluster, 4) an overview of the distribution of a selection of markers (Lineage and relevant Transcription Factors) comparing the signal cluster versus the one of the whole sample. The list of cell types were produced at the end of the classification after evaluating all cell populations identified.

R (version 4.0.2) and Python (version 3.9) were used for data analysis.

Fig.1 Correlation matrix of all markers. The markers have been reordered with the corplot function in R with the angular order of the eigenvectors. Notably, T cell and B cell markers clusterize separately in the table (top left corner). The coloured bars (top and left) represent the IF channel of acquisition of each marker, green= FITC, yellow= TRITC, RED= Cy5, BLU= BV480. Numbers inside bars indicate the round of staining with MILAN. For the precise identification of secondary antibodies and fluorochromes refer to table 1.

Fig.2 Lymph nodes data analyzed with Seurat of 37 samples. The heatmap obtained has been rearranged per cell types for clarity. Each row contains the heatmap belonging to a cluster with the cell identification on the left. Clusters are grouped for cell types. Vertical red boxes help in identifying key markers magnified at the bottom of each group.

Fig.3 Tonsil data analyzed with Seurat of 3 samples. The heatmap obtained has been rearranged per cell types for clarity. Each row contains the heatmap belonging to a cluster with the cell identification on the left. Clusters are grouped for cell types. Vertical red boxes help in identifying key markers magnified at the bottom of each group.

Fig.4 The figure shows the section outline of the entire lymph node UPN107 (gray) on top of which 8 different cell type combinations are shown in order to illustrate spatial relationships. Name of cell types are color coded according to the lymph node image content. A portion of each image (black square) is magnified in the top right corner. Each cell is represented by a magnified dot for visualization.A) two types of CDc1 occupy largely non overlapping regions, B) putative DCs precursors (pre DC) occupy an area where classical DC are found. C) cDC2, which occupy the paracortical area, intermingle with histiocytes but not phagocytes in the medullary cords D)

scattered or focal T cell activated ar associated with a population of CD7+CD23+ monocytes/macrophages E) Mantel zone and marginal zone B cell occupy distinct and occasionally overlapping areas F) proliferating population ID2+ is diffusely and scattered interstitially and can be seen in Germinal Centers G) Stromal cells are seen in a adventitial position in relation to the vascular and lymphatics (LYVE1) endothelium. Marginal sinus cells not shown. H) An example of excluded cells (merged) as identified by BRAQUE. They represent examples of artifacts.

Table 1 List of Antibodies used for the panel, in red Transcription factors.

Table 2 List of Cell Types identified across all samples, in detail elective markers and others further identified.

Fig.1

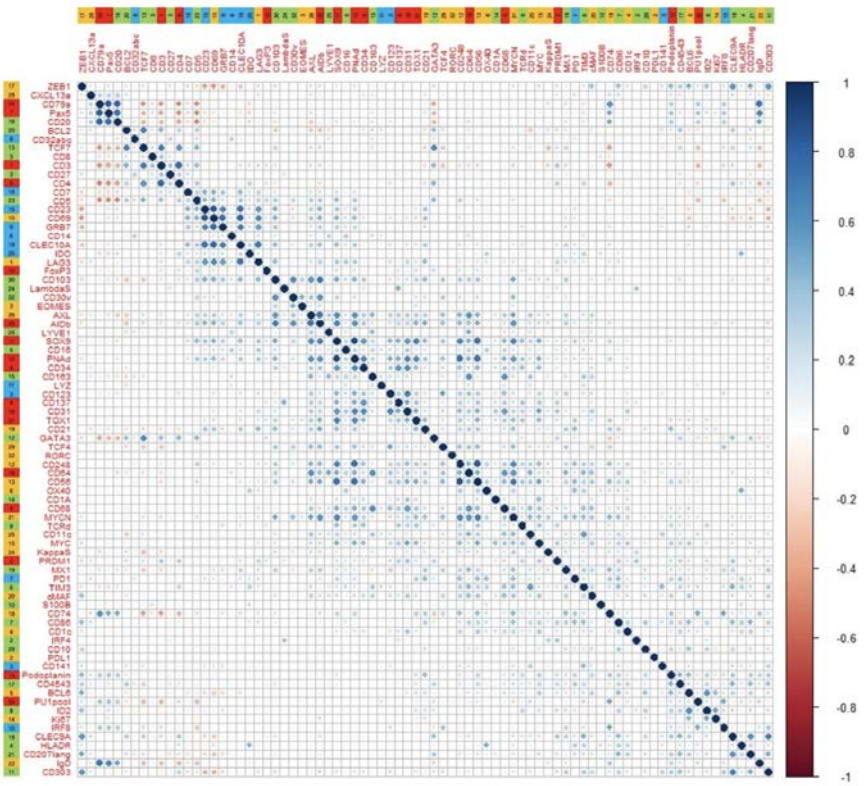

ZEB1
CXCL13e
CD79a
Pax5
CD20
BCL2
CD32abc
TCF7
CD8
CD3
CD27
CD4
CD7
CD5
CD25
CD69
GRB7
CD14
CLEC10A
IDO
LAG3
FoxP3
CD103
LambdaS
CD30r
EOMES
AXL
AIOI
LYVE1
SOX9
CD16
PNAd
CD34
CD163
LYZ
CD123
CD133
CD31
TdT
TBX1
CD21
GATA3
TCF4
RORC
CD248
CD64
CD56
CXA0
CD1A
CD68
MYCN
TCRd
CD11c
MYC
KappaS
PROM1
MX1
PD1
TIM3
dsMAF
S100B
CD74
CD86
CD1c
IRF4
CD19
PDL1
CD141
Podoplanin
CD4543
BCL6
PU1phos
IDO
KIT
IRF8
CLECSA
HLADR
CD207lang
IgD
CD303
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Fig.2 Cell Types Seurat 37 LN

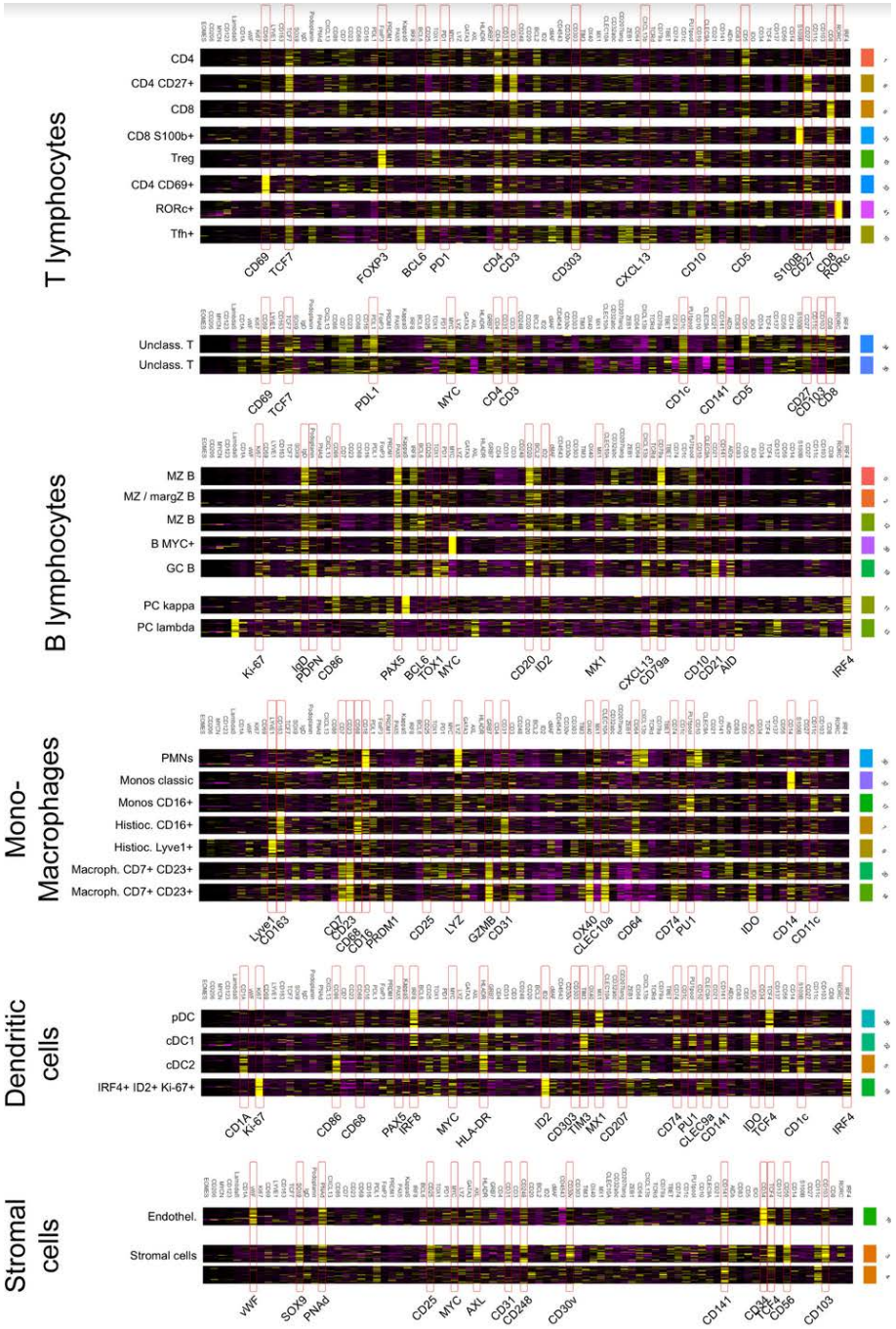


Fig.33 Tonsils Cell Type Seurat

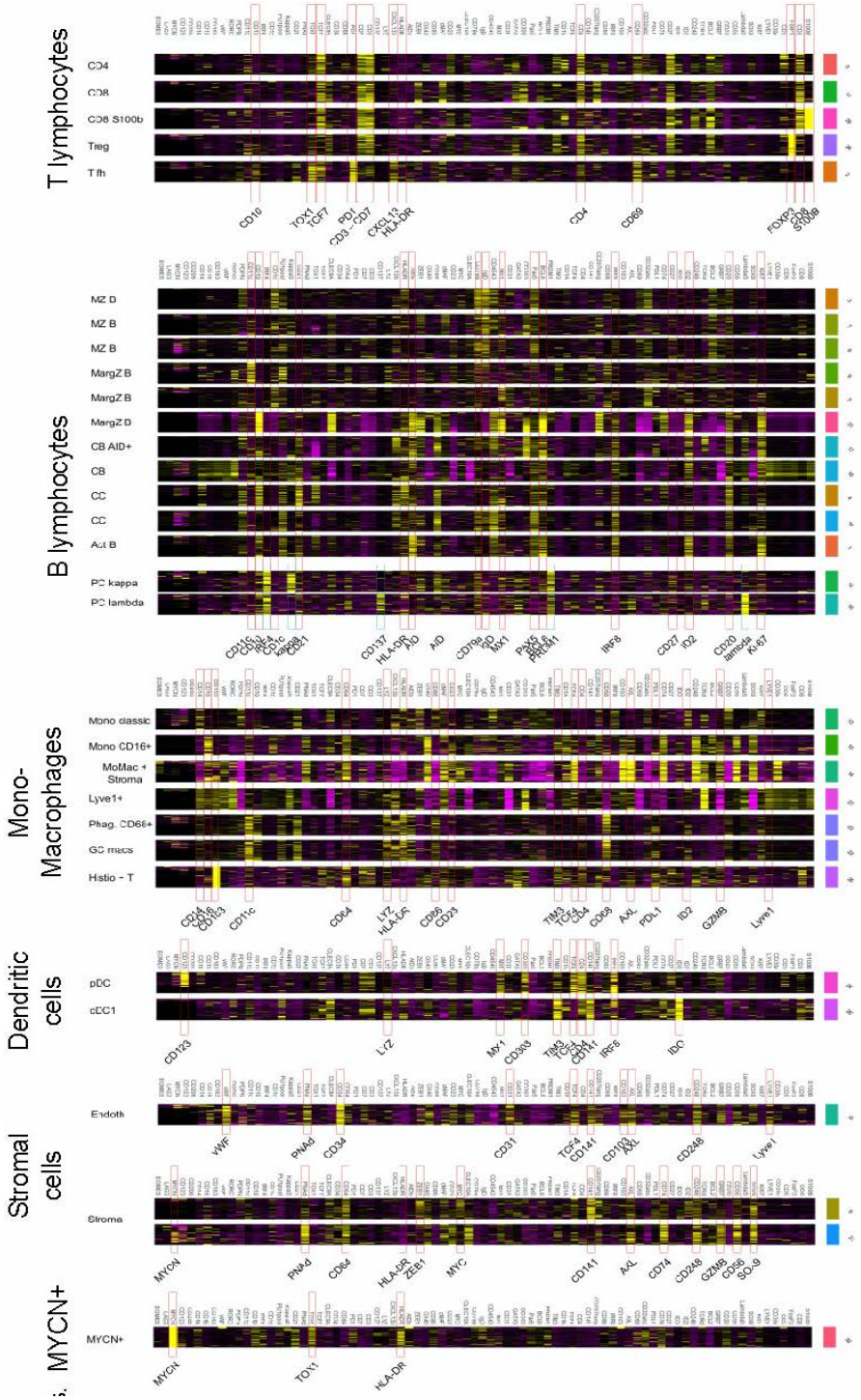


Fig. 4

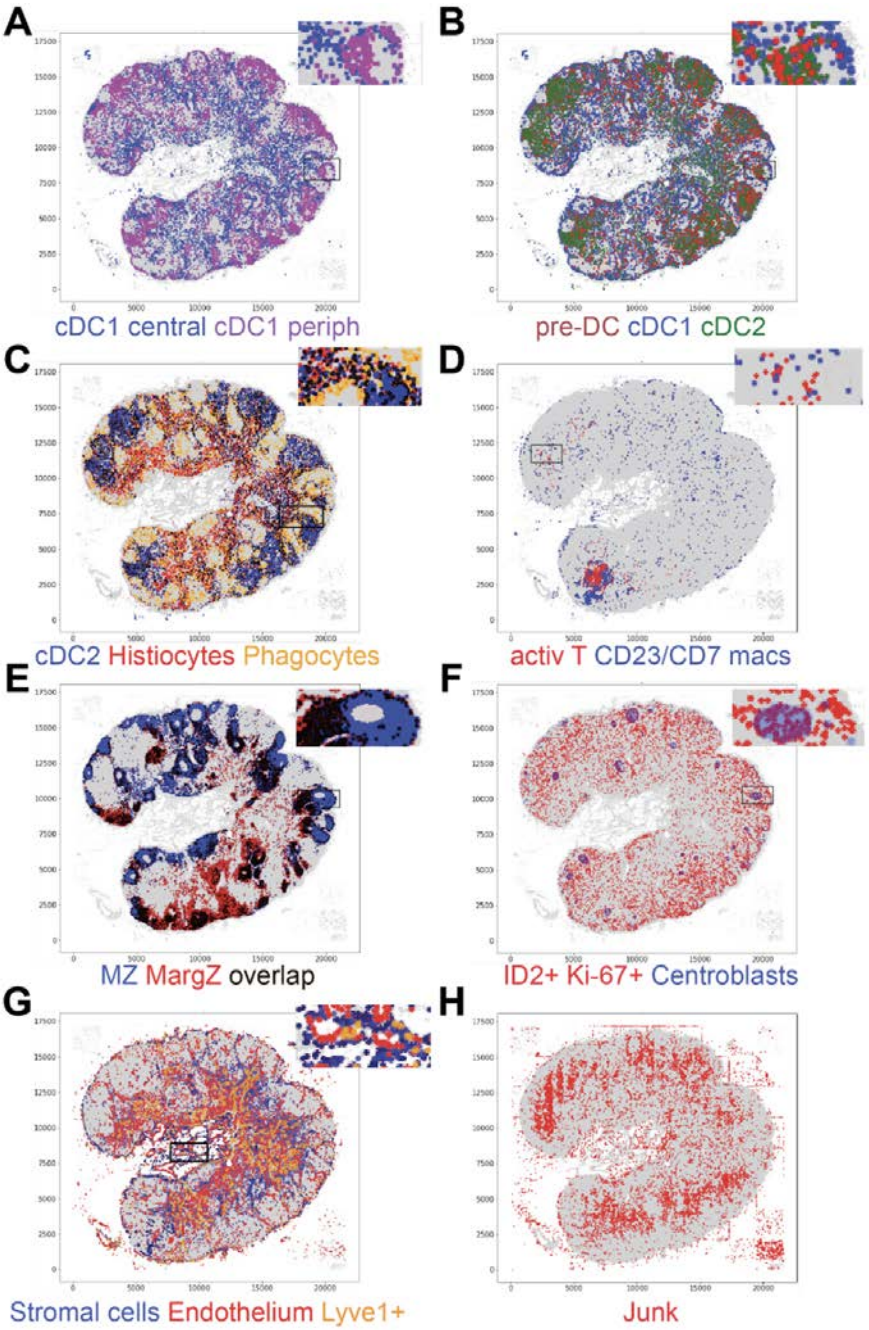

A
cDC1 central cDC1 periph
B
pre-DC cDC1 cDC2
C
cDC2 Histiocytes Phagocytes
D
activ T CD23/CD7 macs
E
MZ MargZ overlap
F
ID2+ Ki-67+ Centroblasts
G
Stromal cells Endothelium Lyve1+
H
Junk

Table 1

AID	CD25	CTLA4	JUN	PU1
AXL	CD27	CTNNB1	kappa	RORC
BCL-2	CD30v	CXCL13	Ki-67	S100A9
BCL6	CD303	CXCR5	LAG3	S100B
cMAF	CD31	EBNA1	lambda	SOX9
CD10	CD34	EBNA2	HHV8	TCF4
CD103	CD36	EOMES	CD207	TCF7
CD11c	CD39	FABP4	LEF1	TCRd
CD123	CD3e	FCeR1	LMP1	TIM3
CD127	CD4	EBER	LYVE1	TOX
CD137	CD45+43	FoxP3	LYZ	TP53
CD138	CD45RB	GATA3	MX1	VISTA
CD14	CD5	GCSF1R	MYC	vWF
CD141	CD56	gH2AX	MYCN	ZEB1
CD16	CD64	Granzyme B	cREL	
CD163	CD68	Granzyme K	p50	
CD1A	CD69	HHV8	p52	
CD1c	CD7	HIV p24	OX40	
CD2	CD74	HLA-DR	PAX5	
CD20	CD79a	ID2	PDL1	
CD206	CD8	IDO	PD1	
CD21	CD86	IgD	PNAd	
CD22	CDKN1B	IRF4	PDPN	
CD23	CLEC10A	IRF5	PRDM1	
CD248	CLEC9A	IRF8	pSTAT1	

Table 2

		Elective markers					Facultative markers				
Cell type def.	**features**										
activated T cells		EOMES	TIM3	IDO			OX40	CD30	IRF4		
T cells	poor CD4-CD8 distinction	CD3	CD4	CD8							
CD4		CD3	CD4	TCF7	bcl2	GATA3	PRDM1	CD5	CD7		
CD8		CD3	CD8	TCF7	CD7	CD5	bcl2	PRDM 1			
CD8 S100+		CD3	CD8	S100b	CD7	CD5					
T reg		CD3	FOXP3	CD4	TCF7		TOX1				
extrafoll T fh	spatially not follicular	PD1	CD4	HLA-DR	OX40	GATA3	TOX1				
Tfh (GC based) (1)		PD1	TOX1	CD69	CD3	CD4	TCF7				
B cells	no spatial/follicular info	Pax5	CD20	CD79a							
activated B cells		ID2	CD20				MX1	BCL6			
ID2+ Ki-67+ B cells		ID2	Ki-67	PAX5lo	PU1lo						
Centroblasts	GC microenv signature	Ki-67	ID2	BCL6	PAX5	CXCL13	AID	CD10	IRF8	TOX1	PU1
Centrocytes	GC microenv signature	CD21	TOX1	CD20	PAX5	CXCL13	CD86	CD10	IRF4	PRDM 1	
margZ B cells (2)	B cell markers IgDlo	PU1	CD74				CD11c				
MZ B cells	B cell markers IgD+ and/or CD1c	CD74	HLA-DR	CD1c	IgD						
PC kappa		kappa	IRF4	PRDM 1			CD79a	CD86			
PC lambda (3)		lambda	IRF4	PRDM 1	CD137		CD79a	CD86			
pre-PC	light chains+	PAX5lo	PU1lo	IRF4lo	PRDM1lo						
Plasmablasts (4)	no light chains	IRF4	PRDM 1								
monocytes	CD68/CD163neg	LYZ	CD14	CD64	CD11c						
CD16+ macs (5)	CD16 hi CD14 low +histio	CD16	CD11c	TIM3			CD14				
CD16+ monos	CD16 hi CD14 low NOhistio	CD16	LYZ				CD23				
CD23 CD7 Macs (6)	CD23+	CD23	CD7	GZMB	CD14	CLEC10a	CD16	TIM3	IDO	LYZ	
phagocytes CD68+	low/absent CD163	CD68	CD11c	CD64			LYZ	CD163			
GC macs	follicle-based CD68+										
Histiocytes		CD163	TIM3	CD11c	CD68		AXL	cMAF			
sinusoidal macs (7)		LYZ	CD23								
Lyve1+ sinus Macs		CD163	Lyve1	AXL	CD64		CD16				
Neutrophils (CD10)		CD10	LYZ								
CXCL13											
cDC1		TIM3	CD141	CLEC9a	LYZ	CD11c	S100b	HLA-DR	CD86	IDO	
cDC2 (8)		CD1a	CD207	CD1c	CLEC9a		S100b	IRF4			
DC precursor		IRF8	ID2	PU1							
ID2+ Ki-67+ (9)		ID2	Ki-67				BCL6	PRDM 1	FOXP 3		
pDC		TCF4	CD303	MX1	IRF8	CD4	CD68	CD74			
LymphoEpith complex (10)		CD45/43neg									
FDC subpop											
endothelium		vWF	CD31	CD34	TCF4	PNAd	CD103	CD248	OX40	CD141	
Lyve1+ endothelium		Lyve1	vWF	CD64	CD141						
marginal sinus stroma (11)	stromal and macrophage marker	CD69	Zeb1	CD34	CD23	AXL	CD103	AID	BCL6		
stromal cells (12)		SOX9	PNAd	CD74	GZMB	CD248	MYCN	CD30	AID	OX40	
stromal cell MYC+		SOX9	PNAd	CD74	GZMB	MYC					

Notes

1 no BCL6 detected (at variance with literature data) **2** no CD1c, no IgD **3** CD137: spurious reaction with PC lambda only **4** if withing the GC, name centrocytes **5** infrequent in tonsils, often sinusoidal **6** CD23: shown on activated Macs; CD7: not prev. shown **7** Not applicable to tonsils (don't have sinuses) **8** should have CLEC10a **9** unclear if DC or GC precursors **10** Tightly intertwined epithelial cells & hematopoietic cells **11** Not applicable to tonsils (don't have sinuses).**12** co-espressione di SOX9 e TCF4: to be proven

Supplemental Table 2

Supplemental Figure 1

Supplemental Figure 2

Supplemental Figure 3

Supplemental Figure 4

The project has obtained an IRB approval by the Institutional Review Board Comitato Etico Brianza, N. 3204, 'High-dimensional single cell classification of pathology (HDSSCP)', October 2019. Patient consent was obtained or waived according to article 89 of the EU General Data Protection Regulation 2016/679 and decree N.515 of the Italian Privacy Authority (19 December 2018).

References

Bhate, S.S., Barlow, G.L., Schürch, C.M., Nolan, G.P., 2022. Tissue schematics map the specialization of immune tissue motifs and their appropriation by tumors. Cell Syst. 13, 109-130.e6. https://doi.org/10.1016/j.cels.2021.09.012

Bolognesi, M.M., Manzoni, M., Scalia, C.R., Zannella, S., Bosisio, F.M., Faretta, M., Cattoretti, G., 2017. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. J. Histochem. Cytochem. 65, 431–444. https://doi.org/10.1369/0022155417719419

Chang, J.E., Turley, S.J., 2015. Stromal infrastructure of the lymph node and coordination of immunity. Trends Immunol. 36, 30–39. https://doi.org/10.1016/j.it.2014.11.003

Crotty, S., 2014a. T Follicular Helper Cell Differentiation, Function, and Roles in Disease. Immunity 41, 529–542. https://doi.org/10.1016/j.immuni.2014.10.004

Cupedo, T., Stroock, A., Coles, M., 2012. Application of tissue engineering to the immune system: development of artificial lymph nodes. Front. Immunol. 3. https://doi.org/10.3389/fimmu.2012.00343

Domínguez Conde, C., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., Mamanova, L., Huang, N., Szabo, P.A., Richardson, L., Bolt, L., Fasouli, E.S., Mahbubani, K.T., Prete, M., Tuck, L., Richoz, N., Tuong, Z.K., Campos, L., Mousa, H.S., Needham, E.J., Pritchard, S., Li, T., Elmentaite, R., Park, J., Rahmani, E., Chen, D., Menon, D.K., Bayraktar, O.A., James, L.K., Meyer, K.B., Yosef, N., Clatworthy, M.R., Sims, P.A., Farber, D.L., Saeb-Parsy, K., Jones, J.L., Teichmann, S.A., 2022. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science 376, eabl5197. https://doi.org/10.1126/science.abl5197

Efremova, M., Vento-Tormo, R., Park, J.-E., Teichmann, S.A., James, K.R., 2020. Immunology in the Era of Single-Cell Technologies. Annu. Rev. Immunol. 38, 727–757. https://doi.org/10.1146/annurev-immunol-090419-020340

Furia, L., Pelicci, P.G., Faretta, M., 2013. A computational platform for robotized fluorescence microscopy (I): High-content image-based cell-cycle analysis. Cytometry A 83A, 333–343. https://doi.org/10.1002/cyto.a.22266

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E.P., Jain, J., Srivastava, A., Stuart, T., Fleming, L.M., Yeung, B., Rogers, A.J., McElrath, J.M., Blish, C.A., Gottardo, R., Smibert, P., Satija, R., 2021. Integrated

analysis of multimodal single-cell data. Cell 184, 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Honan, A.M., Chen, Z., 2021. Stromal Cells Underlining the Paths From Autoimmunity, Inflammation to Cancer With Roles Beyond Structural and Nutritional Support. Front. Cell Dev. Biol. 9, 658984. https://doi.org/10.3389/fcell.2021.658984

HuBMAP Consortium, Writing Group, Snyder, M.P., Lin, S., Posgai, A., Atkinson, M., Regev, A., Rood, J., Rozenblatt-Rosen, O., Gaffney, L., Hupalowska, A., Satija, R., Gehlenborg, N., Shendure, J., Laskin, J., Harbury, P., Nystrom, N.A., Silverstein, J.C., Bar-Joseph, Z., Zhang, K., Börner, K., Lin, Y., Conroy, R., Procaccini, D., Roy, A.L., Pillai, A., Brown, M., Galis, Z.S., Caltech-UW TMC, Cai, L., Shendure, J., Trapnell, C., Lin, S., Jackson, D., Stanford-WashU TMC, Snyder, M.P., Nolan, G., Greenleaf, W.J., Lin, Y., Plevritis, S., Ahadi, S., Nevins, S.A., Lee, H., Schuerch, C.M., Black, S., Venkataraaman, V.G., Esplin, E., Horning, A., Bahmani, A., UCSD TMC, Zhang, K., Sun, X., Jain, S., Hagood, J., Pryhuber, G., Kharchenko, P., University of Florida TMC, Atkinson, M., Bodenmiller, B., Brusko, T., Clare-Salzler, M., Nick, H., Otto, K., Posgai, A., Wasserfall, C., Jorgensen, M., Brusko, M., Maffioletti, S., Vanderbilt University TMC, Caprioli, R.M., Spraggins, J.M., Gutierrez, D., Patterson, N.H., Neumann, E.K., Harris, R., deCaestecker, M., Fogo, A.B., van de Plas, R., Lau, K., California Institute of Technology TTD, Cai, L., Yuan, G.-C., Zhu, Q., Dries, R., Harvard TTD, Yin, P., Saka, S.K., Kishi, J.Y., Wang, Y., Goldaracena, I., Purdue TTD, Laskin, J., Ye, D., Burnum-Johnson, K.E., Piehowski, P.D., Ansong, C., Zhu, Y., Stanford TTD, Harbury, P., Desai, T., Mulye, J., Chou, P., Nagendran, M., HuBMAP Integration, Visualization, and Engagement (HIVE) Collaboratory: Carnegie Mellon, Tools Component, Bar-Joseph, Z., Teichmann, S.A., Paten, B., Murphy, R.F., Ma, J., Kiselev, V.Yu.,

Kingsford, C., Ricarte, A., Keays, M., Akoju, S.A., Ruffalo, M., Harvard Medical School, Tools Component, Gehlenborg, N., Kharchenko, P., Vella, M., McCallum, C., Indiana University Bloomington, Mapping Component, Börner, K., Cross, L.E., Friedman, S.H., Heiland, R., Herr, B., Macklin, P., Quardokus, E.M., Record, L., Sluka, J.P., Weber, G.M., Pittsburgh Supercomputing Center and University of Pittsburgh, Infrastructure and Engagement Component, Nystrom, N.A., Silverstein, J.C., Blood, P.D., Ropelewski, A.J., Shirey, W.E., Scibek, R.M., University of South Dakota, Collaboration Core, Mabee, P., Lenhardt, W.C., Robasky, K., Michailidis, S., New York Genome Center, Mapping Component, Satija, R., Marioni, J., Regev, A., Butler, A., Stuart, T., Fisher, E., Ghazanfar, S., Rood, J., Gaffney, L., Eraslan, G., Biancalani, T., Vaishnav, E.D., NIH HuBMAP Working Group, Conroy, R., Procaccini, D., Roy, A., Pillai, A., Brown, M., Galis, Z., Srinivas, P., Pawlyk, A., Sechi, S., Wilder, E., Anderson, J., 2019. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. Nature 574, 187–192. https://doi.org/10.1038/s41586-019-1629-x

Kalina, T., Fišer, K., Pérez-Andrés, M., Kuzílková, D., Cuenca, M., Bartol, S.J.W., Blanco, E., Engel, P., van Zelm, M.C., 2019. CD Maps-Dynamic Profiling of CD1-CD100 Surface Expression on Human Leukocyte and Lymphocyte Subsets. Front. Immunol. 10, 2434. https://doi.org/10.3389/fimmu.2019.02434

Krishnamurty, A.T., Turley, S.J., 2020. Lymph node stromal cells: cartographers of the immune system. Nat. Immunol. 21, 369–380. https://doi.org/10.1038/s41590-020-0635-3

Krop, J., van der Zwan, A., Ijsselsteijn, M.E., Kapsenberg, H., Luk, S.J., Hendriks, S.H., van der Keur, C., Verleng, L.J., Somarakis, A., van der Meeren, L., Haasnoot, G., Bos, M., de Miranda, N.F.C.C., Chuva de Sousa Lopes, S.M., van der Hoorn, M.-L.P., Koning, F.,

Claas, F.H.J., Heidt, S., Eikmans, M., 2022. Imaging mass cytometry reveals the prominent role of myeloid cells at the maternal-fetal interface. iScience 25, 104648. https://doi.org/10.1016/j.isci.2022.104648

Liu, Z., Zhang, Z., 2022. Mapping cell types across human tissues. Science 376, 695–696. https://doi.org/10.1126/science.abq2116

Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., Theis, F.J., 2022. Benchmarking atlas-level data integration in single-cell genomics. Nat. Methods 19, 41–50. https://doi.org/10.1038/s41592-021-01336-8

Mantovani, A., Sica, A., Sozzani, S., Allavena, P., Vecchi, A., Locati, M., 2004. The chemokine system in diverse forms of macrophage activation and polarization. Trends Immunol. 25, 677–686. https://doi.org/10.1016/j.it.2004.09.015

Massoni-Badosa, R., Soler-Vila, P., Aguilar-Fernández, S., Nieto, J.C., Elosua-Bayes, M., Marchese, D., Kulis, M., Vilas-Zornoza, A., Bühler, M.M., Rashmi, S., Alsinet, C., Caratù, G., Moutinho, C., Ruiz, S., Lorden, P., Lunazzi, G., Colomer, D., Frigola, G., Blevins, W., Palomino, S., Gomez-Cabrero, D., Agirre, X., Weniger, M.A., Marini, F., Cervera-Paz, F.J., Baptista, P.M., Vilaseca, I., Prosper, F., Küppers, R., Gut, I.G., Campo, E., Martin-Subero, J.I., Heyn, H., 2022. An Atlas of Cells in the Human Tonsil. https://doi.org/10.1101/2022.06.24.497299

Moffitt, J.R., Lundberg, E., Heyn, H., 2022. The emerging landscape of spatial profiling technologies. Nat. Rev. Genet. https://doi.org/10.1038/s41576-022-00515-3

Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A., Teichmann, S.A., 2017. The Human Cell Atlas: from vision to reality. Nature 550, 451–453. https://doi.org/10.1038/550451a

Saelens, W., Cannoodt, R., Todorov, H., Saeys, Y., 2019. A comparison of single-cell trajectory inference methods. Nat. Biotechnol. 37, 547–554. https://doi.org/10.1038/s41587-019-0071-9

Stein, K., Hummel, M., Korbjuhn, P., Foss, H.-D., Anagnostopoulos, I., Marafioti, T., Stein, H., 1999. Monocytoid B Cells Are Distinct From Splenic Marginal Zone Cells and Commonly Derive From Unmutated Naive B Cells and Less Frequently From Postgerminal Center B Cells by Polyclonal Transformation. Blood 94, 2800–2808. https://doi.org/10.1182/blood.V94.8.2800.420a01_2800_2808

Stoler-Barak, L., Biram, A., Davidzohn, N., Addadi, Y., Golani, O., Shulman, Z., 2019. B cell dissemination patterns during the germinal center reaction revealed by whole-organ imaging. J. Exp. Med. 216, 2515–2530. https://doi.org/10.1084/jem.20190789

Stringer, C., Wang, T., Michaelos, M., Pachitariu, M., 2021. Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods 18, 100–106. https://doi.org/10.1038/s41592-020-01018-x

Szabo, P.A., Levitin, H.M., Miron, M., Snyder, M.E., Senda, T., Yuan, J., Cheng, Y.L., Bush, E.C., Dogra, P., Thapa, P., Farber, D.L., Sims, P.A., 2019a. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat. Commun. 10, 4706. https://doi.org/10.1038/s41467-019-12464-3

Szabo, P.A., Levitin, H.M., Miron, M., Snyder, M.E., Senda, T., Yuan, J., Cheng, Y.L., Bush, E.C., Dogra, P., Thapa, P., Farber, D.L., Sims, P.A., 2019b. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat. Commun. 10, 4706. https://doi.org/10.1038/s41467-019-12464-3

The Tabula Sapiens Consortium, Quake, S.R., 2021. The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans (preprint). Cell Biology. https://doi.org/10.1101/2021.07.19.452956

Vercelli, D., Jabara, H.H., Lee, B.-W., Woodland, N., Geha, R.S.,

Leung, D.Y.M., n.d. HUMAN RECOMBINANT INTERLEUKIN 4 INDUCES FcER2/CD23 ON NORMAL HUMAN MONOCYTES 11.

Willard-Mack, C.L., 2006. Normal Structure, Function, and Histology of Lymph Nodes. Toxicol. Pathol. 34, 409–424. https://doi.org/10.1080/01926230600867727

Chapter 7

Summary:

This PhD thesis collects a selection of results of the last 3 years of research after the publication of the MILAN[1] multiplex staining technology method. The first part is related to the application of MILAN[2] methods to different types of tissue. This has allowed us to characterize the immune infiltrate presence in a rare dataset of Uterine leiomyosarcomas[3] and to identify cell population in a mouse model of IPF by measuring the percentage of lung cell types and their attitude to express a drug transporter[4]. These papers represent the application of the method to two different scenarios, one clinical and one experimental and it's used to answer specific scientific questions.

Then an overview of the available technologies for multiplex immunohistochemistry is presented, "their advantages and challenges, the comparison with MILAN, and provide the basic principles on how to interpret high-dimensional data in a spatial context."[5]

Subsequently, because of the availability of a large lymphoid tissue image data set, for which there were no present specific algorithms, we created BRAQUE, a tailored pipeline of analysis for data coming from highly multiplexed imaging data sets (such as Codex, MILAN..).

The last part of the project is dedicated to the classification at single cell level of normal human lymph nodes thanks to the previously created bioinformatic tool BRAQUE. This has been possible thanks to a substantial effort of sample collection and

analysis of almost 100 samples stained with more than 80 antibodies.

Even if the project is still ongoing, we have already obtained some interesting results presented in chapter 5-6 (manuscripts in preparations). The database of lymph nodes cell populations will be subsequently integrated with scRNAseq data.

Conclusions and Future perspective

Many improvements occurred in recent years in single cell biology. Different technologies have been developed and implemented but a standard approach in using them has not been reached yet.

Spatial transcriptomics [6] have expanded the interest towards data coming directly from tissue and the possibility of data-integration is adding a new level of comprehension, related to the cell microenvironment.

MILAN technology and its pipeline of analysis become of interest within this context. Having the possibility to integrate “single cell spatial data clusters”, coming from a panel of 100 markers (antibodies), with single cell RNA seq data will produce results more accurately than ever and help in localizing very precisely population and signal.

Data interpretation and, as consequence, classification of clusters, still remains a crucial point that needs to be explored.

By providing a new tool tailored for high-plex immunofluorescence data, BRAQUE, we are going to create an

encoder for automatic cell typing by comparing the signature of the cell with clusters already generated .

This work has been possible thanks to a close collaboration with Bologna DIMES Department (prof. Castellani).

Despite some specific results related to the type of tissue analyzed, this approach can be also translated from research to clinic.

From a technical/translational point of view MILAN protocol can be easily exported to every wet lab, including Pathology hospital departments. It's a scalable technology that can be exploited both for small and precious samples (small biopsies) or for sample collection and retrospective studies by TMA construction. In fact only one section will be used and there's no waste of materials.

For this reason we would like to develop a more user friendly GUI (graphical user interface) to apply the BRAQUE algorithm to generic tissue samples and the pipeline needs to be expanded with other modules such as neighborhood analysis (for clinical and research purposes).

References

1. Bolognesi MM, Manzoni M, Scalia CR, et al. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. :14.

2. G C, Fm B, L M, Mm B. Multiple Iterative Labeling by Antibody Neodeposition (MILAN). Published online September 25, 2019. doi:10.21203/rs.2.1646/v5

3. Manzoni M, Bolognesi MM, Antoranz A, et al. The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas. *Sci Rep*. 2020;10(1):702. doi:10.1038/s41598-020-57627-1

4. Ciccimarra R, Bolognesi MM, Zoboli M, Cattoretti G, Stellari FF, Ravanetti F. The normal and fibrotic mouse lung classified by spatial proteomic analysis. *Sci Rep*. 2022;12(1):8742. doi:10.1038/s41598-022-12738-9

5. Bosisio FM, Van Herck Y, Messiaen J, et al. Next-Generation Pathology Using Multiplexed Immunohistochemistry: Mapping Tissue Architecture at Single-Cell Level. *Front Oncol*. 2022;12:918900. doi:10.3389/fonc.2022.918900

6. Marx V. Method of the Year: spatially resolved transcriptomics. *Nat Methods*. 2021;18(1):9-14. doi:10.1038/s41592-020-01033-y

Publications

1. Furia, L. *et al.* Automated multimodal fluorescence microscopy for hyperplex spatial-proteomics: Coupling microfluidic-based immunofluorescence to high resolution, high sensitivity, three-dimensional analysis of histological slides. *Front. Oncol.* **12**, 960734 (2022).
2. Antoranz, A. *et al.* Mapping the Immune Landscape in Metastatic Melanoma Reveals Localized Cell–Cell Interactions That Predict Immunotherapy Response. *Cancer Res.* **82**, 3275–3290 (2022).
3. Bolognesi, M. M. *et al.* Dissecting the histological features of lupus nephritis highlights new common patterns of injury in class III/IV. *Ann. Rheum. Dis.* (2022) doi:10.1136/ard-2022-222620.
4. Grasselli, C. *et al.* DNA Damage in Circulating Hematopoietic Progenitor Stem Cells as Promising Biological Sensor of Frailty. *J. Gerontol. Ser. A* **77**, 1279–1286 (2022).
5. Ciccimarra, R. The normal and fibrotic mouse lung classified by spatial proteomic analysis. *Sci. Rep.* 11 (2022).
6. Denti, V. *et al.* Lipidomic Typing of Colorectal Cancer Tissue Containing Tumour-Infiltrating Lymphocytes by MALDI Mass Spectrometry Imaging. *Metabolites* **11**, 599 (2021).
7. Bolognesi, M. M. *et al.* Antibodies validated for routinely

processed tissues stain frozen sections unpredictably. *BioTechniques* **70**, 137–148 (2021).

8. Gadeyne, L. *et al.* A Multi-Omics Analysis of Metastatic Melanoma Identifies a Germinal Center-Like Tumor Microenvironment in HLA-DR-Positive Tumor Areas. *Front. Oncol.* **11**, (2021).

9. Background-free Detection of Mouse Antibodies on Mouse Tissue by Anti-isotype Secondary Antibodies - Francesco Mascadri, Roberta Ciccimarra, Maddalena M. Bolognesi, Fabio Stellari, Francesca Ravanetti, Giorgio Cattoretti, 2021. https://journals.sagepub.com/doi/10.1369/00221554211033239.

10. Rossi, F. *et al.* Proteomics for the study of new biomarkers in Fabry disease: State of the art. *Mol. Genet. Metab.* **132**, 86–93 (2021).

11. Rejuvenated Vintage Tissue Sections Highlight Individual Antigen Fate During Processing and Long-term Storage - Francesco Mascadri, Maddalena M. Bolognesi, Daniela Pilla, Giorgio Cattoretti, 2021. https://journals.sagepub.com/doi/10.1369/00221554211047287.

12. Manzoni, M. *et al.* The Adaptive and Innate Immune Cell Landscape of Uterine Leiomyosarcomas. *Sci. Rep.* **10**, 702 (2020).

13. Bosisio, F. M. *et al.* Functional heterogeneity of lymphocytic patterns in primary melanoma dissected

through single-cell multiplexing. *eLife* **9**, e53008 (2020).

14. Bosisio, F. *et al.* High Resolution Multiplexing of Melanoma Microenvironment in Responders/Non-Responders to Checkpoint Therapy. *Mod. Pathol.* **33**, 457–458 (2020).

15. Bombelli, S. *et al.* PKHhigh/CD133+/CD24− Renal Stem-Like Cells Isolated from Human Nephrospheres Exhibit In Vitro Multipotency. *Cells* **9**, 1805 (2020).

16. Bolognesi, M. M., Bosisio, F. M. & Cattoretti, G. Unidentified Variables May Account for Variability in Multiplexing Results. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **68**, 351–353 (2020).

17. Ferrario, F. *et al.* Kidney Function, CKD Causes, and Histological Classification. in *Integration of Omics Approaches and Systems Biology for Clinical Applications* 301–318 (John Wiley & Sons, Ltd, 2018). doi:10.1002/9781119183952.ch18.

18. Scalia, C. R. *et al.* Antigen Masking During Fixation and Embedding, Dissected. *J. Histochem. Cytochem.* **65**, 5–20 (2017).

19. Bolognesi, M. M. *et al.* Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *J. Histochem. Cytochem.* **65**, 431–444 (2017).

20. Scalia, C. R. *et al.* Epitope Recognition in the Human–Pig Comparison Model on Fixed and Embedded Material. *J. Histochem. Cytochem.* **63**, 805–822 (2015).