THE UNIVERSITY OF CHICAGO

MULTI-OMIC URINE CFDNA APPROACHES FOR DETECTING MOLECULAR RESIDUAL DISEASE IN BLADDER CANCER AND SINGLE-CELL MECHANISMS OF RADIATION-INDUCED METASTASIS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

IMMUNOLOGY

BY

ARPIT PANDA

CHICAGO, ILLINOIS

JUNE 2024

To my family.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

As I write this, I realize that I truly stand on the shoulders of giants. None of this would have been possible without the many amazing people in my life. First and foremost, I must thank my mentor, Dr. Chuan He. His scientific pedigree is unmatched and there is not much more I can add to what others haven't already said and written about him. I feel lucky that he took me on a graduate student and gave me the freedom to choose translational projects that truly matched my scientific interests. From Chuan, I learned the importance of doing rigorous and impactful science. He provided the perfect environment for me to succeed where I was given the freedom to make mistakes but with the knowledge that there was always a safety cushion to prevent me from completely losing my way. In addition, he encouraged collaboration and this led to the development of the two collaborator-driven projects that ended up being the bulwark of my PhD thesis work.

The urine cfDNA biomarker project was made possible by Dr. Aadel Chaudhuri. As a leader in the cfDNA diagnostics space, he provided tremendous domain expertise and his lab provided all the human samples. The support of Chaudhuri lab members, Dr. Pradeep Chauhan and Dr. Irfan Alahi, was absolutely critical for the success of the project.

The radiation oncology project studying the effects of AREG was another opportunity for me to learn about bridging clinical practice with basic science. It was made possible by the excellent mentorship and leadership of Dr. Ralph Weichselbaum. He has exacting standards which helped push me to improve myself as scientist and a communicator. In addition, I worked closely with Dr. Andras Piffko on the AREG project. His scientific insights, and good humor made him an absolute pleasure to work with. I'm deeply grateful for all his help.

I want to give a special thanks to current and former He lab members like Dr. Diana West-Szymanski, Krissana Kowitwanich and Dr. Xiaolong Cui. I am grateful that I was surrounded by such thoughtful and kind people in the lab.

The thesis committee also deserves my thanks. Dr. Mengjie Chen is a computational powerhouse and was able to provide the critical mentorship I needed to carry out the computational aspects of the projects. Dr. Andrew Koh's immunological and epigenetic expertise helped me make sense of the hydroxymethylation landscapes in healthy and cancer patients. Dr. Brian Chiu had previously developed 5hmC biomarkers for patients with B-cell lymphomas and gave me important feedback on the unique challenges of working with patient data.

I am grateful to the Medical Scientist Training Program and the Committee on Immunology, which provided a rich academic environment for me to succeed.

Finally, I have to thank the most important people in my life. My parents, Akshaya and Anima, have made tremendous sacrifices to provide me with the life and opportunities I have today. They have inculcated values in me like kindness, resilience, and social responsibility. They inspired me to pick a career that was focused on helping make the lives of those around me better. My sister, Apekshya, is one of the most intelligent people I've ever met in my life and is one of my best friends and scientific collaborators. Finally, nothing would be possible with my fiancée, Christine. Thanks to her own experiences as a dual-degree trainee, she has been able to support and guide me through the toughest portions of my PhD training. These people add meaning to everything I do.

ABSTRACT

In Chapter 1, we developed cell-free DNA (cfDNA) biomarkers. Standard-of-care for patients with localized muscle-invasive bladder cancer is neoadjuvant chemotherapy followed by radical cystectomy. Bladder-sparing treatments are limited by our current inability to sensitively detect minimal residual disease (MRD). To address this, we focused on the biofluid most proximal to disease, urine, and analyzed cfDNA from 84 localized bladder cancer patients. We applied two orthogonal molecular approaches: low-pass whole genome sequencing (LP-WGS) and genome-wide 5-hydroxymethylcytosines (5hmC) profiling in order to achieve sensitive MRD detection. Using LP-WGS, we computed copy-number derived tumor fraction (TFx) and found that patients who had not achieved complete response (no pCR), as determined by gold-standard surgical pathology, had significantly higher TFx levels than those with complete response. A classifier for predicting residual disease based on TFx achieved a receiver operating curve area under the curve (ROC AUC) of 0.78 but did not predict patient overall survival. We then developed a 5hmC-based model based on a 25-gene marker panel for predicting MRD which achieved an ROC AUC of 0.91. The panel contained a number of independently prognostic genes implicated in carcinogenesis. Patients predicted to have MRD based on the 5hmC model had worse progression-free survival and overall survival. In summary, we provided proof-of-principle that urine cfDNA omics approaches can noninvasively detect MRD and reveal potentially new onco-relevant targets.

In the second chapter, we used a murine preclinical metastasis model to identify mechanisms of radioresistance. Previous work from Piffko et al. had identified amphiregulin (AREG), an EGFR ligand, as a key factor in driving metastatic growth in non-irradiated lesions in patients receiving stereotactic body radiotherapy (SBRT). To identify immune mechanisms of

AREG-mediated metastasis, we performed single-cell RNA sequencing (scRNA-seq) of $CD45^+$ cells in lung tissues following radiotherapy of flank tumors. Our analysis revealed that AREG triggered distinct transcriptional changes in mononuclear phagocytes(MNPs), especially inflammatory monocytes. These changes included upregulation of immunosuppressive and downregulation of inflammatory and antigen-presenting pathways. Moreover, there was a marked increase in EGFR signaling in inflammatory monocytes following irradiation, providing a link between AREG-mediated EGFR signaling activation and the immunosuppressive phenotype. A monocyte trajectory analysis unveiled that AREG led to divergent differentiation trajectories in monocytes, with some populations arrested at an immature, immune-suppressive state while others followed a tumor-tolerogenic trajectory. Building on the findings from Piffko et al., we suggest a model where radiation-induced soluble growth factors modulate the host immune response, potentially offering therapeutic targets to significantly improve patient outcomes in the future.

Chapter 1

INTRODUCTION

Circulating tumor DNA as biomarkers for cancer

Cell-free DNA(cfDNA) was first described in 1948 by Mandel and Metais and refers to extracellular DNA found in plasma as well as other biofluids like urine, cerebrospinal fluid and saliva[1,2]. The majority of cell-free DNA (cfDNA) is generated by apoptosis, necrosis, and phagocytosis, and is either bound to nucleosomes or packed in lipid microvesicles[2,3,4]. In healthy individuals, most of the cfDNA originates from hematopoietic cells such as erythrocytes and immune cells like circulating leukocytes[5,6]. Pathological processes like disease or a high degree of inflammation lead to excessive cell death and an accumulation of cfDNA[7,8,9].

In 1987, Stroun et al. discovered a correlation between cfDNA and cancer where they discovered a substantial portion of patients they were studying had a high fraction of cfDNA compared to healthy patients who had very limited detectable cfDNA[10]. Since then, a vast number of studies have confirmed the increased presence of tumor-derived cfDNA, also known as circulating tumor DNA (ctDNA), in cancer patients[11,12,13]. Plasma-based ctDNA detection has now been used in diagnosis, therapy monitoring, molecular disease detection and a host of other applications[14,15,16,17,18,19]. ctDNA is reflective of the mutational status of the underlying tumor and a number of detection methods have been employed to capture mutational states in patients with cancer which include high-depth whole genome sequencing as well as targeted approaches[15]. One potential confounder with mutational approaches is clonal hematopoiesis of indeterminate potential (CHIP), an acquisition of somatic mutations in subpopulations of blood cells in older patients who are cancer-free[20,21]. This underscores the necessity for additional approaches for cfDNA analysis including copy-number based methods and epigenetic assays.

Hydroxymethlation approaches to cell-free DNA analysis

Epigenetic modifications play a crucial role in mammalian development, cellular differentiation, and disease, with 5-methylcytosine (5mC) being the most well-studied epigenetic modification. 5mC globally represses gene expression and plays a significant gene-regulatory role in various diseases[22]. The ten-eleven translocation (TET) family of enzymes can oxidize 5mC to 5-hydroxymethylcytosine (5hmC), a stable marker of global transcriptional activation[23,24]. 5hmC is frequently used as a mark to investigate dynamic DNA methylation conversion during development and in human disease[22,25].

Previous work from our group extensively characterized how 5hmC modifications play an important role in tissue-specific gene and enhancer modulation by profiling 19 healthy human tissues[22]. Compared to 5mC, 5hmC is more specific and a potentially more relevant marker of dynamic processes like disease[22,25]. With the increasing need for minimally-invasive liquid biopsy solutions, our group and others have derived 5hmc-signatures from plasma cfDNA to detect several cancers including colon, pancreatic, hepatocellular carcinoma, diffuse large B-cell lymphoma, among others[26,27,28,29]. However, in the context of genitourinary (GU) diseases like bladder cancer, plasma cfDNA currently remains insufficiently sensitive for either diagnosing early-stage disease.

5hmC for molecular residual disease detection in bladder cancer

In the United States, bladder cancer is the sixth most commonly diagnosed cancer and the fourth most common cancer among men[30]. The stage of the disease at presentation largely influences treatment options[31]. About 70% of cases are non–muscle-invasive bladder cancer (NMIBC), typically treated with transurethral resection and possibly intravesical therapy[32]. These

patients often face a risk of recurrence and must undergo frequent, invasive cystoscopic monitoring, which contributes to high lifetime treatment costs[32].

Around 25% of cases involve muscle-invasive bladder cancer (MIBC), where the tumor penetrates into the detrusor muscle[33]. Treatment usually involves radical cystectomy with urinary diversion, often after neoadjuvant chemotherapy. Despite aggressive treatment, 5-year survival rates for MIBC are only around 50%, with significant recurrence risk[34,35].

Pathologic response is a key predictor of survival after radical cystectomy[36]. About 35% of patients achieve a pathologic complete response (pCR) following neoadjuvant treatment[36]. Identifying these patients beforehand could lead to more personalized and less invasive care, avoiding radical surgery in some cases. Cystoscopy is an invasive method for assessing tumor response to therapy, but it can underestimate disease burden[37,38]. This complicates the use of personalized, bladder-sparing treatment approaches, as these patients still face significant recurrence risk and require invasive monitoring.

Liquid biopsy solutions via urine cfDNA analysis offer an intriguing alternative for management of localized bladder cancer management. Urine poses a number of advantages over other sample types – it is noninvasive, can be easily collected serially outside a clinic/hospital setting to monitor disease and conduct treatment surveillance, and is the most proximal biofluid to bladder cancer. Springer et al. developed UroSEEK, a non-invasive method for detecting bladder cancer by analyzing genetic material in urine sediment[38]. In high-risk patients, UroSEEK showed significantly higher sensitivity compared to conventional cytology[38]. It detected bladder cancer earlier than diagnosis, and even earlier than standard-of-care methods in post-surgery patients[38]. In contrast, surveillance cytology had a much lower sensitivity[38]. While this pellet-based approach demonstrates the tremendous utility of urine as an analyte in bladder cancer, urine cfDNA remains

understudied. The objective of this current work is to pilot novel 5hmC-based epigenetic approaches in urine cfDNA.

5hmC-based approaches in plasma and urine cfDNA might provide a sensitive and specific solution for MRD detection in bladder cancer. Furthermore, by leveraging the unique ability of 5hmC to capture dynamic transcriptional regulation, we hope to study treatment course in bladder cancer patients receiving neoadjuvant therapy and develop biomarkers of treatment response.

A group of collaborators has previously piloted a urine cfDNA-based, mutation-based approach (uCAPP-Seq) for MRD detection[39]. This approach utilized blood and urine samples collected from the same patients, employing a hybrid capture panel targeting genes commonly mutated in bladder cancer to identify cfDNA[39]. By comparing the mutation profiles between urine and plasma, researchers were able to infer the presence of MRD effectively. This approach was reasonably successful and 81% sensitive with an ROCAUC of 0.78[39].

Our current work significantly advances the MRD diagnostics field by building on these findings with a much larger cohort of localized bladder cancer patients, with longer follow-up and introducing two independent, molecular methods. In order to avoid some of the confounders in mutation-based analyses, we apply a copy number-based approach to infer tumor fraction in cfDNA. Second, we leverage genome-wide 5hmC profiling as a novel epigenetic biomarker strategy. These methods offer a clear advantage as they only require urine samples, eliminating the need for matching blood samples and thus providing a less invasive option for patients. In this work, we assess the diagnostic and predictive capabilities of each approach, benchmarking them against uCAPP-Seq.

METHODS

Patient recruitment and sample collection

Chauhan et al. recruited 84 patients with localized bladder cancer who underwent curative-intent radical cystectomy at the Washington University Siteman Cancer Center[40]. Patients were required to be 18 years or older with a confirmed bladder cancer diagnosis through histologic or cytologic assessment. Urine and blood samples were collected at enrollment, and we also collected samples from 40 healthy adult volunteers for comparison. All procedures followed relevant guidelines and were approved by the institutional review board at Washington University in St. Louis School of Medicine. Both patients and healthy donors participated in study NCT04354064 (ClinicalTrials.gov), providing written informed consent in line with the Declaration of Helsinki. This study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for observational research.

Pathologic response assessment

Surgical resection specimens from radical cystectomy were managed according to a consistent protocol, including collection, handling, and submission to the Pathology Department at Washington University School of Medicine. Experienced board-certified genitourinary surgical pathologists, blinded to clinical details, carried out microscopic examination of the specimens. The study defined pathologic complete response (pCR) as pathologic stage T0, Tis, or Ta according to the AJCC 8th edition classification. Non-pCR was categorized as stages T1 through T4, possibly including nodal involvement (N1–N2) or metastatic disease.

Urine cell-free DNA extraction

This protocol was performed as described in Chauhan et al[40]. Urine samples were collected in cups containing 1-2 mL of 0.5 M EDTA. Immediately after collection, cfDNA was isolated from 22 to 90 mL of urine using Q-sepharose resin slurry (GE Healthcare, Chicago, Illinois). The process involved adding Q-sepharose resin to urine at a ratio of 10 μL slurry per mL of urine and mixing for 30 minutes. After centrifuging at 1800 × g for 10 minutes, the supernatant was discarded. The resin was then washed twice with 0.3 M LiCl/10 mM sodium acetate (pH 5.5), transferred to a Micro Bio-Spin column (Bio-Rad, Hercules, California, USA), and bound DNA was eluted using 70% ethanol before being passed through a QIAquick column (Qiagen, Hilden, Germany). Columns were rinsed with 2 M LiCl in 70% ethanol, followed by 75 mM potassium acetate (pH 5.5) in 80% ethanol. Finally, DNA was eluted in either nuclease-free water or 10 mM Tris-Cl (pH 8.5). The concentration of urine cfDNA was measured using the Qubit dsDNA High Sensitivity Assay kit (Thermo Fisher Scientific, Waltham, Massachusetts), and its quality was analyzed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California).

Low-pass whole genome sequencing (LP-WGS)

ULP-WGS libraries were prepared from 32 to 50 ng of sheared urine cfDNA using the Kapa HyperPrep kit (Roche, Basel, Switzerland). Libraries were balanced, pooled, and sequenced on a HiSeq 4000 (Illumina, San Diego, California) to a median deduplicated depth of 5-10x (Supplementary Data 6). FASTQ files were demultiplexed and raw reads were quality-filtered using fastq v.0.20.0. Quality-filtered reads were then aligned to the hg38 human genome assembly using BWA v.0.7.17. Aligned reads were deduplicated with Samtools v.1.13.

ichorCNA Tumor Fraction

ichorCNA v0.3.2 was utilized to estimate tumor fractions in each urine cfDNA sample. In essence, the process involves summing reads in 1 mb non-overlapping bins and adjusting the local read depth for GC bias and known low mappability regions. Artifacts were eliminated by referencing ichorCNA's built-in healthy control database. Using parameters designed for low tumor fractions in cfDNA samples, the software predicted copy number alterations (CNAs) throughout the genome, excluding the X and Y chromosomes. ichorCNA then employed these binned, bias-corrected copy number data to model a two-component mixture of tumor-derived and non-tumor-derived fragments. From this model, the software inferred the proportion of tumor-originating reads in each sample, known as the tumor fraction (TFx). The Youden index is a measure of the overall diagnostic effectiveness of a medical test. It is defined as the sensitivity + specificity – 1. Our optimal TFx threshold, equal to the Youden index, was 6.2%. We calculated sensitivity and specificity for this index. Furthermore, we calculated positive predictive value and negative predictive value by using this threshold to categorize all patients as MRD+ (above 6.2%) or MRD-(below 6.2%). The ground truth was actual pCR status obtained from pathological examination of the surgical specimens.

5hmC sequencing

Prior to 5hmC analysis, cfDNA was acoustically sheared to approximately 200-bp fragments using LE220 focused ultrasonicator (Covaris, Woburn, Massachussetts, USA). The Nano-hmC-Seal procedure was performed as outlined in Han et al[ref]. In brief, 10 ng of urine cfDNA was fragmented in Tagmentation buffer at 55 °C and then purified using the Zymo DNA Clean and Concentration Kit. The selective 5hmC chemical labeling process occurred in

glucosylation buffer (50 mM HEPES buffer pH 8.0, 25 mM MgCl2), containing the fragmented DNA, βGT, and N3-UDP-Glc, which was incubated at 37 °C for 2 hours. After purification in ddH2O, DBCO-PEG4-Biotin (Click Chemistry Tools) was added and incubated at 37 °C for 2 hours. Biotin-labeled DNA was then pulled down using C1 Streptavidin beads (Life Technologies) for 15 minutes at room temperature. The captured DNA fragments were amplified by PCR using the Nextera DNA sample preparation kit, and the amplified product was purified with 1.0X AMPure XP beads. An input library was created by directly performing PCR on fragmented DNA without labeling and pull-down. The resulting libraries were quantified using a Qubit fluorometer (Life Technologies) and sequenced on Illumina NovaSeq-X-Plus with an average of 25 million reads per sample.

5hmC analysis pipeline

Paired-end reads underwent initial trimming using Trim_Galore v.0.6.10 to remove adapter sequences and low-quality nucleotides. The high-quality reads were aligned to the hg38 reference genome using Bowtie v2.4.1, retaining only uniquely mapped reads for subsequent analysis. PCR duplicates were removed using mapq filtering. 5hmC peaks were identified using MACS2 v.2.2.91 with the *–extsize* parameter set to 150 and the *–nomodel* set to true. Peaks located in ENCODE blacklist regions were excluded. Annotation and enrichment analysis of 5hmC peaks were conducted using the ChIPseeker toolkit[41]. Promoters were defined as regions from −1 kb to +100 bp relative to the transcription start site (TSS). Metagene profiles were also generated using the ChIPseeker package with default parameters. 5hmC gene body counts were inferred using the *Rsubread()* function from the FeatureCounts package. Mapping to the gene body was consistent across all samples (60-70%). These raw counts were then normalized using the variance stabilizing

transformation using PyDEseq2 v.0.4.8[42]. Differential gene expression was performed using the PyDEseq2 package as well and top 300 differential genes were visualized on a heatmap generated using the Seaborn package ad clustermap function (z-normalization was performed across all rows). Gene set enrichment analysis (GSEA) was performed on the global ranking of all protein coding regions using gseapy v1.1.2[43]. The Reactome_2022, Kegg_2021_Human, ENCODE_Histone_modifications_2015,GO_Biological_Process_2023,MSigDB_Hallmark_2020, and MsigDB_Oncogenic_Signatures gene sets were used as inputs into GSEA[44,45,46,47,48].

Integrated 5hmC-model development

We began with a training set of 124 patients total (pCR = 35, no pCR = 49, healthy = 40). We first conducted differential expression analysis to identify differentially hydroxymethylated regions (DhMRs) in patients with pCR (n=35) and those without pCR (n=49), as well as between presumed cancer-free patients (n=75) and those without pCR (n=49). This gave us a list of 500 candidate genes on which we went on to perform feature selection. These candidate genes were further refined using elastic net regularization in a multivariate logistic regression model, as applied in the *LogisticRegressionCV()*function in the sklearn package. The model underwent cross-validation across a range of parameters for C and l1 ratio (C range: 10^{-5}-1 with logarithmic increments, l1 ratio: 0.1 – 1 in increments of 0.1) where C represents the overall penalty strength and l1 ratio controls the balance between Ridge and Lasso Penalties. This selection process was repeated 200 times, retaining a panel of 25 5hmC marker genes that validated in at least 90% of the iterations.

A weighted diagnostic score (WD-score) was calculated as the sum of the gene-specific product of logistic model coefficients and corresponding 5hmC marker value for each participant.

ROCAUC analysis along with leave-one-out cross-validation was performed on the entire cohort of patients who had bladder cancer at some point(pCR = 35, no pCR = 49) to evaluate diagnostic performance.

The Cancer Genome Atlas (TCGA) Validation

Validation of candidate markers in TCGA was performed using the gene expression profiling interactive analysis (GEPIA) database[49]. Bar plots summarizing median gene expression in TPM in all cancer types (tumor and matched normal) were generated using the online GEPIA platform.

Survival Analysis and Cox Regression

The lifelines package was used to perform Kaplan-Meier analyses to assess progression-free survival and overall survival. The *plot_surival_function*() was used to generate the all K-M plots and p-values were calculated using the *logrank_test()* test. Multivariate cox regression analysis on overall survival was performed with the *CoxPHFitter()* function with the following factors: age, gender, smoking history, TFx (MRD + or MRD-), WD-score (WD+ or WD-). Coefficients and p-values were reported.

RESULTS

Cohort characters and sample selection

A total of 84 patients at Washington University in St. Louis medical center with localized bladder cancer received a physician-choice neoadjuvant treatment regimen followed by curative-intent radical cystectomy. Cystectomy specimens were then analyzed by pathologists and pathologic complete response status (pCR) status was determined based on whether any residual cancer cells were found in the specimen. Urine was collected prospectively from these patients on the day of surgery as well as from an additional forty healthy volunteers and cell-free DNA(cfDNA) was extracted. Two orthogonal methods were employed – nano-5hmC-seal paired with next generation sequencing as well low-pass whole genome sequencing (LP-WGS) on all samples. Each assay was then correlated with residual tumor in the surgical resection specimen and with patient outcomes to evaluate whether we could discover novel, noninvasive molecular biomarkers for residual disease (Figure 1.1).

Seventy-seven percent of the study cohort had urothelial carcinoma, while the remainder had variant histologies. Seventy-seven percent of the patients were men while the remaining patients were women. Ninety-three percent of patients were white with the remaining being non-white. Seven percent of the patients had a smoking history. According to the TNM staging system for bladder cancer, Ta and Tis stages are classified as non-muscle-invasive bladder cancer (NMIBC), while T2, T3, and T4 stages are classified as muscle-invasive bladder cancer (MIBC)[31]. This classification is based on the extent of tumor invasion within the bladder wall, with MIBC indicating that the cancer has invaded the muscular layer of the bladder. Sixty-four percent of our cohort was MIBC. A full description of the cohort is displayed in Figure 1.2.

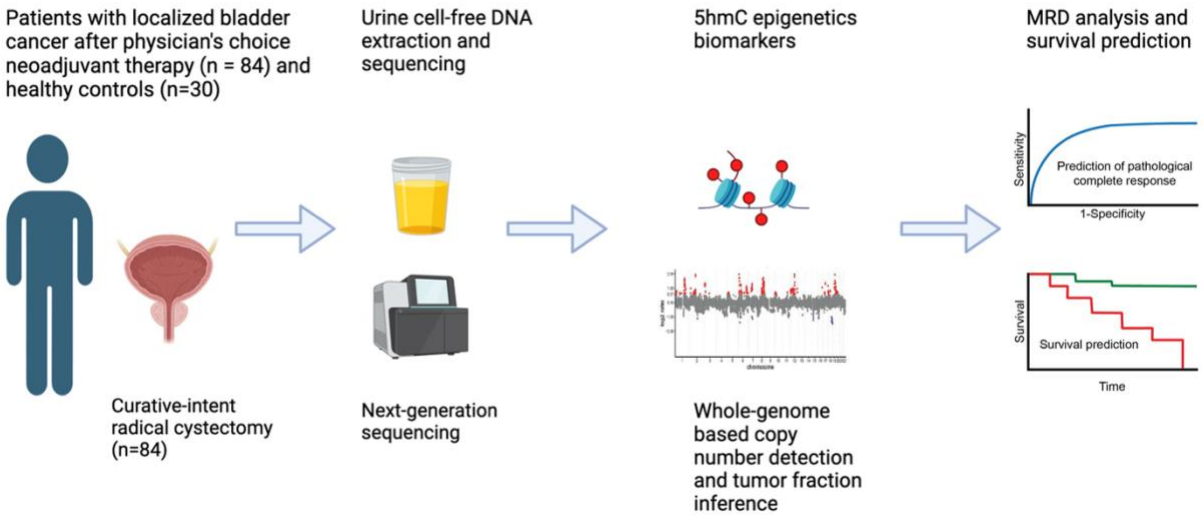

Figure 1.1 Study Schema.

Schematic of the experimental workflow. Urine was collected prospectively from 84 localized bladder cancer patients pre-operatively on the day of curative-intent radical cystectomy following physician's-choice neoadjuvant chemotherapy as well as from 30 healthy patients. Urine cell-free DNA was extracted and sequenced using nano-5hmC-Seal at 2-3x depth and Low-Pass Whole Genome Sequencing(LP-WGS) at 5-10x depth. Diagnostic and prognostic capability of each method was evaluated.

Patient Characteristics (n = 86)	**No. (%)**
Gender	
Male	66 (77)
Female	20 (23)
Median Age (years)	68
Median Follow-up (months)	
Ethnicity	
White	80 (93)
Non-White	6 (7)
Smoking History	
Yes	60[a] (70)
No	26 (30)
T-stage (pre-treatment)	
Ta	5 (6)
Tis	2 (2)
T1	23 (27)
T2	49 (57)
T3	5 (6)
T4	1 (1)
Neoadjuvant Chemotherapy	
Yes	42 (49)
ddMVAC	12 (29)
gemcitabine/cisplatin	22 (53)
gemcitabine/carboplatin	1 (2)
cisplatin/etoposide	1 (2)
carboplatin/paclitaxel	1 (2)
treatment change[b]	1 (2)
unknown regimen	4 (10)
No	43 (50)
Pathologic Complete Response	37 (43)
Histology	
Urothelial	66 (77)
Other	20 (23)

[a]Median 30 pack-years.

[b]Patient received 1 cycle of carboplatin/paclitaxel and 3 cycles of carboplatin/gemcitabine (switched due to cutaneous reaction to paclitaxel).
ddMVAC – dose-dense methotrexate, vinblastine, doxorubicin, and cisplatin

Figure 1.2. Patient Demographics

A subset of patients who did not have available LP-WGS or nano-5hmC-Seal sequencing data were excluded from the respective downstream analyses.

Tumor fraction inference and survival analysis

ichorCNA (Infer Copy Number Aberrations from Tumor DNA) is a computational tool used to infer copy number alterations (CNAs) from sequencing data[50]. It is particularly useful for analyzing low-frequency tumor DNA found in blood samples (such as circulating tumor DNA) or other bodily fluids. ichorCNA was developed to help identify and quantify the presence of CNAs in such samples, which can be valuable for monitoring cancer progression, treatment response, and detecting minimal residual disease (MRD). While ichorCNA has been successfully applied to plasma samples, its application to urine cfDNA has been relatively unexplored.

After extensively optimizing the tool to accommodate the unique properties of urine cfDNA, we established a set of optimal, computational parameters for inferring tumor fraction. Copy number-derived tumor fraction (TFx) levels, estimated from LP-WGS of urine cfDNA of our cohort, ranged from 0 to 62% (Figure 1.3). Bladder cancer patients who had achieved complete pathologic response (pCR) had significantly lower TFx than patients who had no pCR (median 2.1% for pCR vs median 11.82% for no pCR, $p = 0.00005$) (Figure 1.3). Similarly, healthy controls had significantly lower TFx levels than patients with no pCR (median 1.61% for healthy vs median 11.82% for no pCR, $p = 0.00007$) (Figure 1.3). As expected, there was no difference in TFx between healthy and patients with pCR (median 1.61% for healthy vs median 2.1% for pCR, $p = 0.24$) (Figure 1.3).

We aimed to establish an optimal TFx threshold to predict which patients had molecular residual disease (MRD). To do this, we compared TFx values between presumed cancer-free patients (those who were healthy or had achieved pCR) and patients who had not achieved pCR. We tested the ability to distinguish these groups using various TFx cutoffs and used the Youden Index to determine the cutoff point that offered the best diagnostic accuracy. This analysis

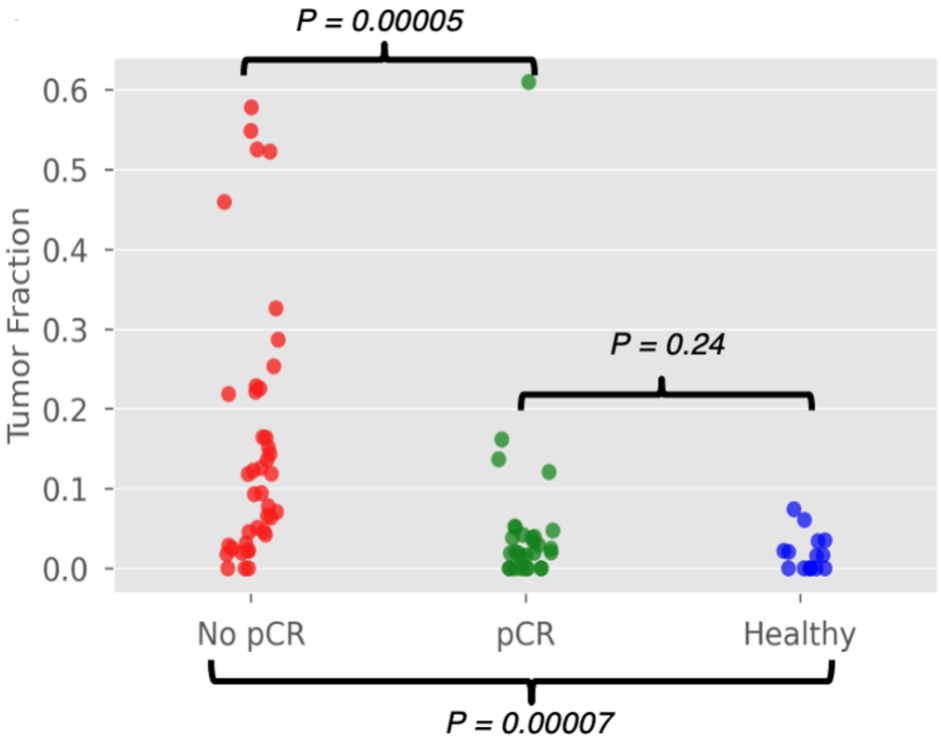

Figure 1.3. Urine tumor fraction is a significant biomarker for determining pCR status.

IchorCNA v0.3.2 was performed using low-tumor fraction parameters and using 1 mb bins on n=84 samples (41 pCR, 29 no pCR and 14 healthy patients). Significance based on Wilcoxon rank-sum test. pCR pathologic complete response.

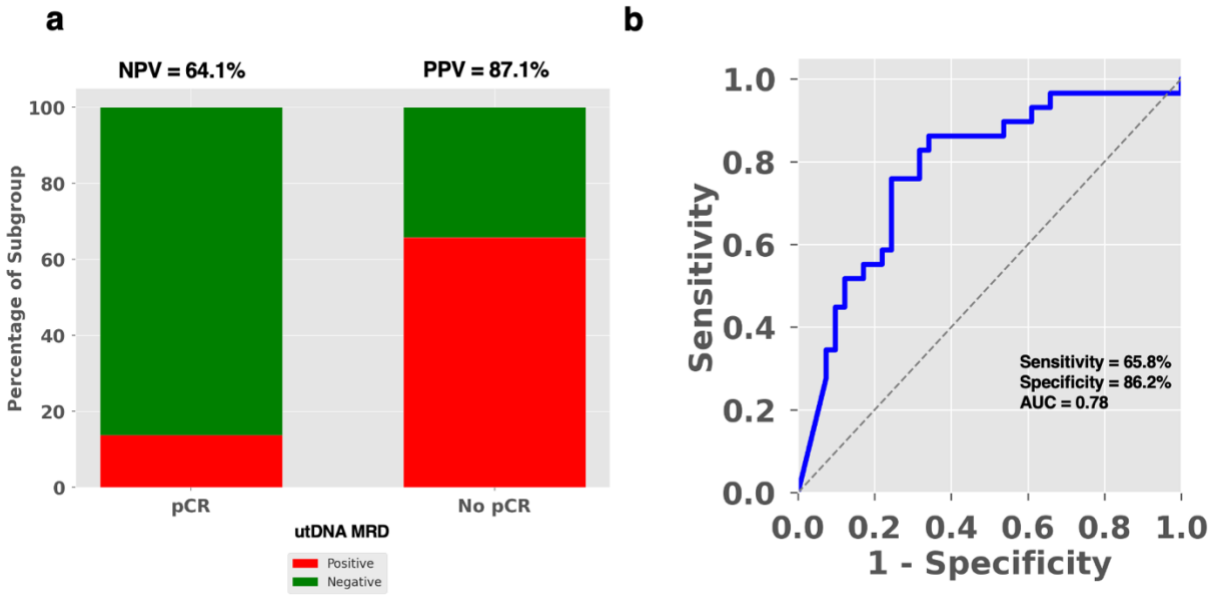

Figure 1.4. An optimal urine tumor fraction threshold is determined to discriminate pCR status.

a, An optimal tumor threshold was developed using the Youden index to determine minimal residual disease. Stacked bar plot depicting NPV and PPV for detecting MRD based on optimal threshold amongst patients with pCR (n=35) and no pCR(n=49). **b,** ROC analysis was performed to discriminate between pCR and no pCR across a range of tumor fraction thresholds and best performing sensitivity and specificity are identified. NPV negative predictive value, PPV positive predictive value, pcR pathological complete response, MRD molecular residual disease, ROC receiver operating characteristic, AUC area under the curve.

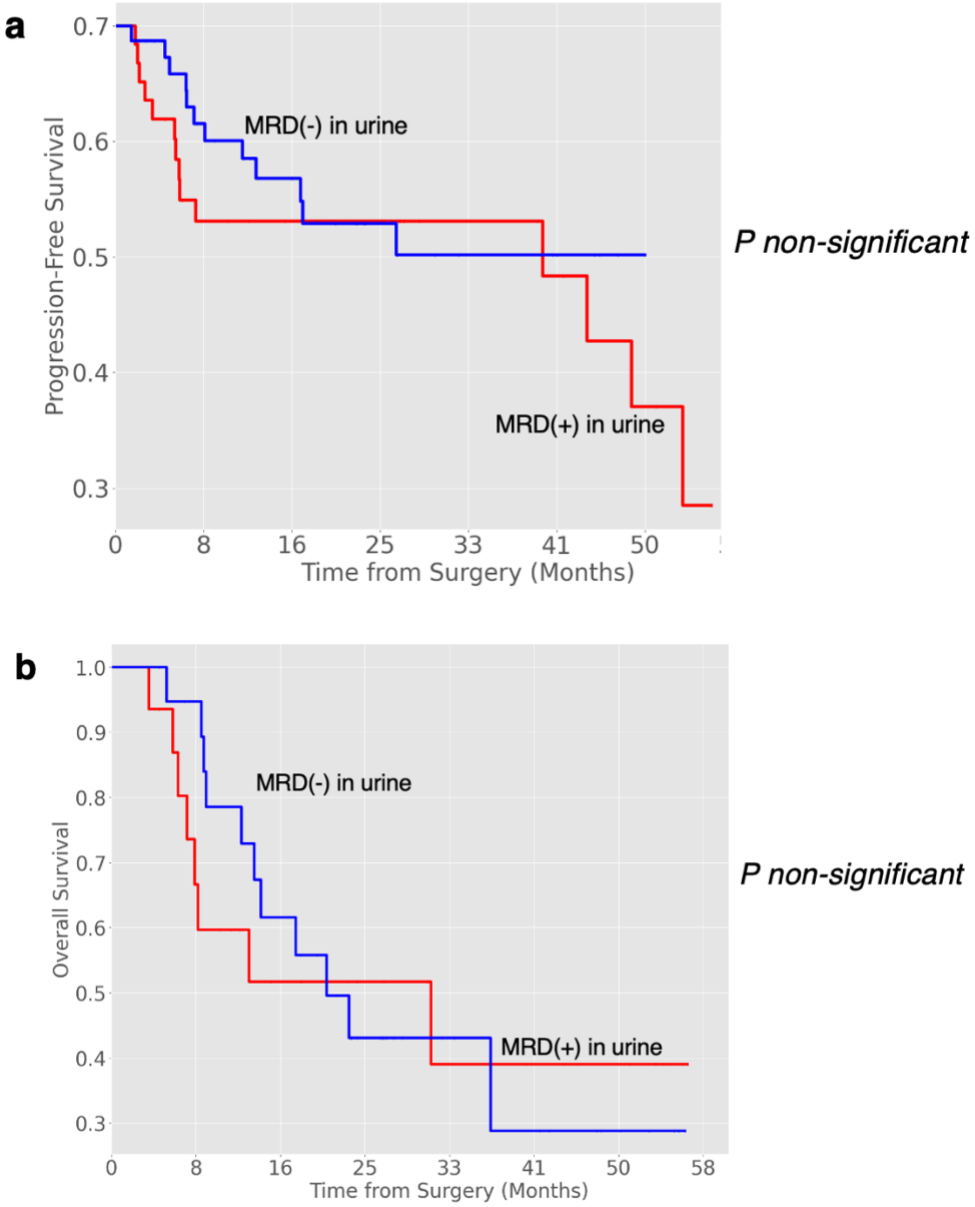

Figure 1.5. Survival analysis using tumor fraction alone is not predictive of patient outcomes.

Kaplan-Meier plots showing **a)** progression-free survival and **b)** overall survival stratified by MRD status determined by the optimal tumor fraction threshold. *p* values were calculated by the log-rank test. MRD molecular residual disease. N =70 patients.

identified a TFx cutoff value of 6.2%. Using this cutoff, we found that the positive predictive value for identifying MRD was 87.1% and the negative predictive value was 64.1% (Figure 1.4a). We went on to perform ROC analysis across a range of TFx cut-off values and the area under the receiver operating characteristic curve (AUROC) was 0.78 (Figure 1.4b). The diagnostic performance of the optimal TFx threshold provided a sensitivity of 65.8% and a specificity of 86.2% (Figure 1.4b).

To evaluate the prognostic power of MRD detection using TFx alone, we categorized patients into MRD+ or MRD- groups based on whether their TFx levels were above or below the 6.2% threshold, respectively. Then, we performed Kaplan-Meier analyses starting from the time of surgery. Progression-free survival and overall survival was not significantly different between the two groups (Figure 1.5). While TFx demonstrated diagnostic utility, its lack of significant prognostic value highlights the need for further exploration of additional molecular approaches.

5hmC modifies biologically relevant genes in localized bladder cancer

To systematically investigate the dynamics of DNA hydroxymethlation across our patient cohort, we performed nano-5hmC-Seal and next generation sequencing (average of 25 million reads). Duplication rates were low and the unique mapping ratio was consistently across the cohort (80-90%). Following peak calling, we assigned peaks to the nearest genomic feature. It has been previously shown that the 5hmC signal is enriched at promoters and gene bodies of individual genes[22]. We confirmed this finding in our dataset where the dominant 5hmC signal arose from intronic regions (55%) and promoters (12%) (Figure 1.6.a). Additionally, a metagene profile of 5hmC signal confirmed similar findings – hyperenrichment in the gene body and promoter regions with a depletion of signal at the transcription start site (TSS) (Figure 1.6.b).

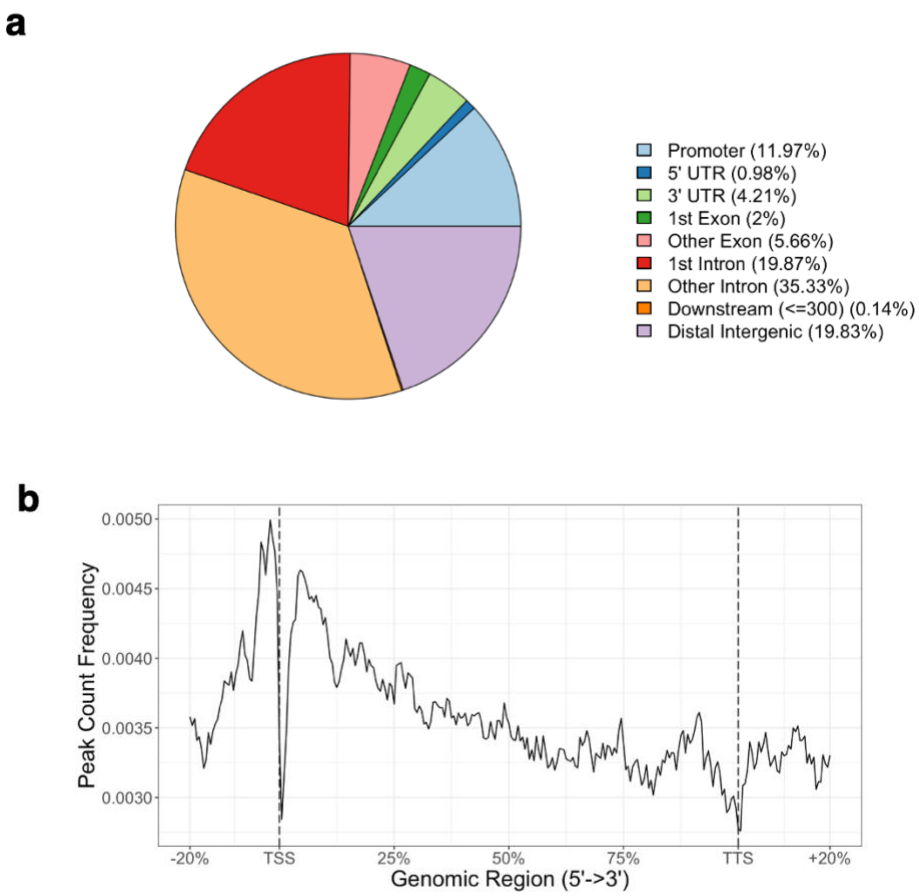

Figure 1.6. 5hmC signal is enriched in the gene body.

a) Representative pie chart of genomic feature representation of 5hmC peaks from human samples. Promoters were defined as regions within 1kb upstream of Transcription Start Site. **b)** Representative metagene plot of 5hmC profiles averaged across all genes in the genome showing under-representation outside the gene body and over-representation at promoter regions and in the gene body. UTR untranslated region. 5hmC 5-hydroxymethylation.

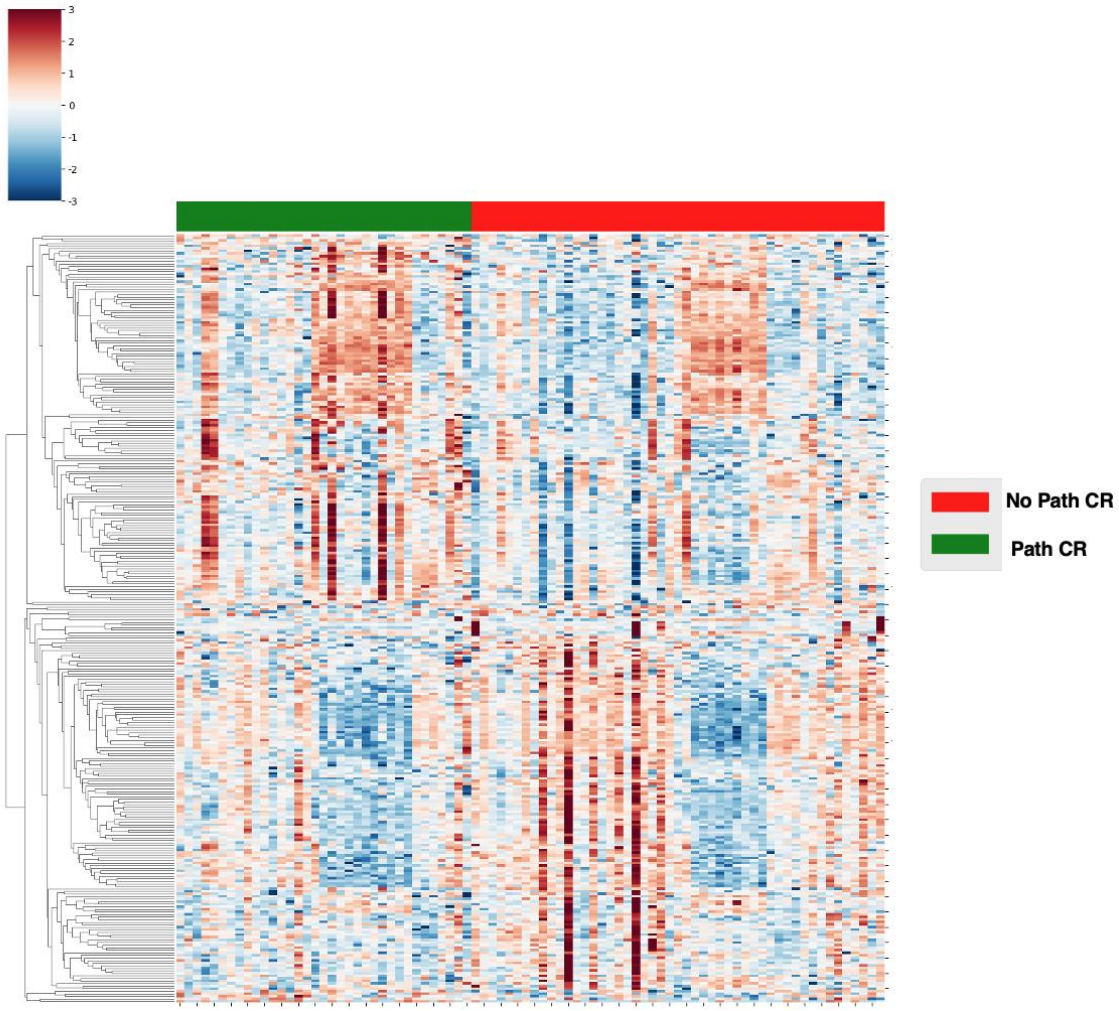

Figure 1.7. Heatmap of differentially hydroxymethylated genes between pCR and no pCR.

Heatmap of the top 300 differentially hydroxymethylated genes between pCR and no pCR. Genes identified using DEseq2 software. Rows represent genes and columns represent individual patient samples. Path CR pathological complete response. All rows were z-score normalized.

Next we visualized the top 300 most differentially hydroxymethylated (DhMR) genes between patients with no PCR and pCR using a heatmap (Figure 1.7). Using a genome-wide ranked list of DhMR genes, we performed Gene Set Enrichment Analysis(GSEA). When we ranked the top pathways based on normalized enrichment score (NES), we found a number of biologically relevant pathways that were upregulated in no pCR patients compared to pCR patients (Figure 1.8a). Notably, patients who did not achieve pCR showed significant enrichment in oncogenic pathways like KEGG Bladder Cancer and Hallmark 2020 Myc Targets (Figure 1.8b). The classical oncogene, MYC, which is crucial in regulating glycolysis, is amplified and activated in up to 30% of patients with bladder cancer[51]. Histone H3K36me3 is an epigenetic mark associated with gene activation. A pathway associated with activation in kidney epithelial cells and a second pathway associated with myocytes were also enriched (Figure 1.8b). These findings suggest increased shedding of kidney epithelial cells into urine, something that has been previously observed in inflammatory conditions[52]. Additionally, muscle invasion in MIBC patients may account for the presence of myocyte pathways, reflecting further specificity of signal.

Our GSEA analysis also identified a set of the most downregulated pathways which were primarily associated with immune signaling and cells from the hematopoietic lineage (Figure 1.9a-b). This observation aligns with previous studies that have demonstrated that cfDNA in healthy individuals mainly originates from immune cells[37]. Specifically, we observed gene sets specific for JAK-STAT signaling, chemokine receptor binding, cGMP-PKG and the hematopoietic lineage that were downregulated in patients with no pCR and conversely up in patients with pCR (presumed cancer-free) (Fig 1.9b).

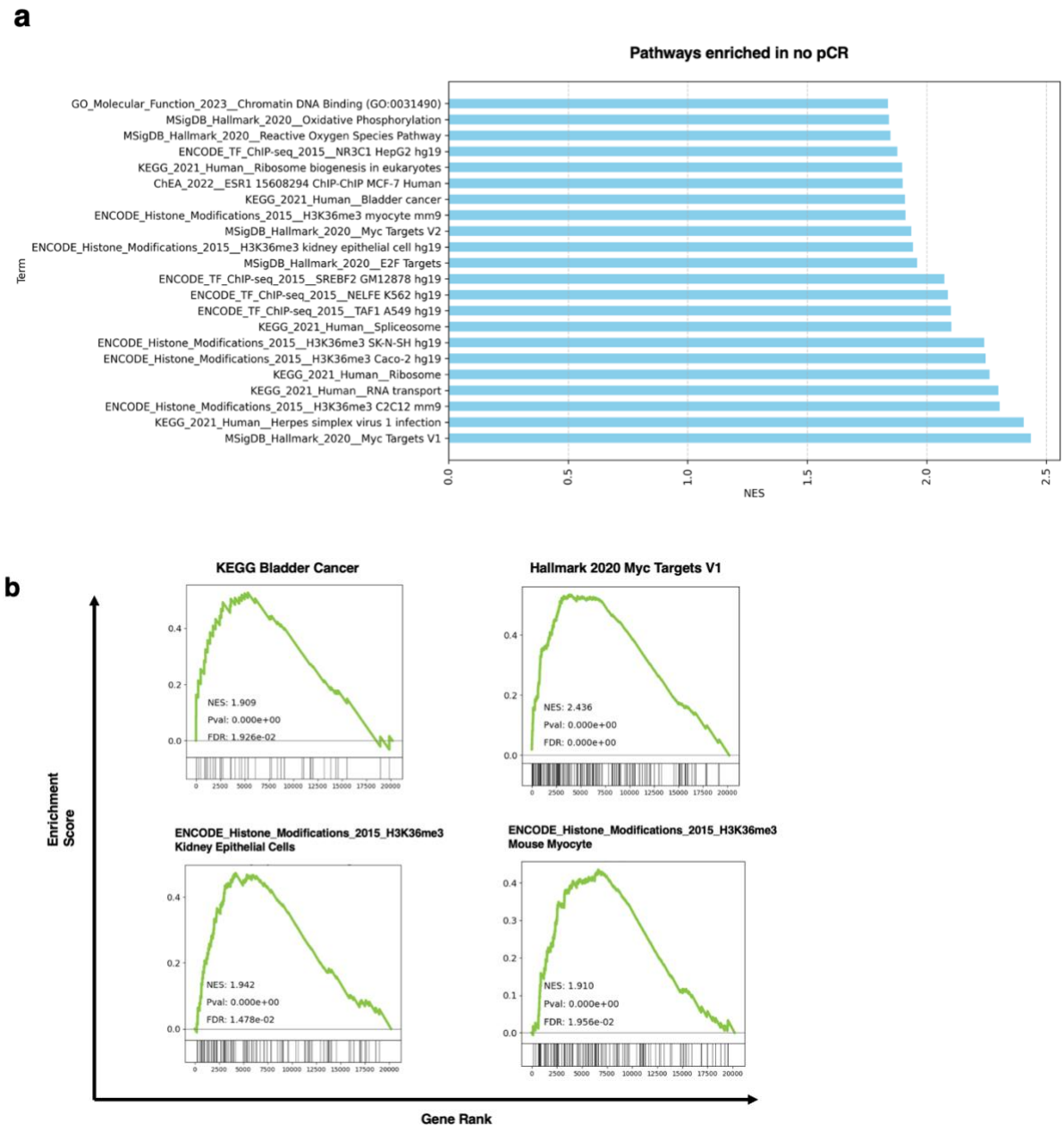

Figure 1.8. Gene Set Enrichment Analysis shows specific oncogenic, epithelial and myocyte pathways upregulated in no pCR.

a) Gene set enrichment analysis was performed on a genome-wide ranked list based on the differential hydroxymethylation of all genes between no pCR and pCR. Bar plot of top pathways from KEGG, Hallmark 2020, ENCODE_Histone_Modifications_2015_H3K36m3 and GO Molecular Function gene sets are displayed. **b)** Enrichment plots of four selected pathways.

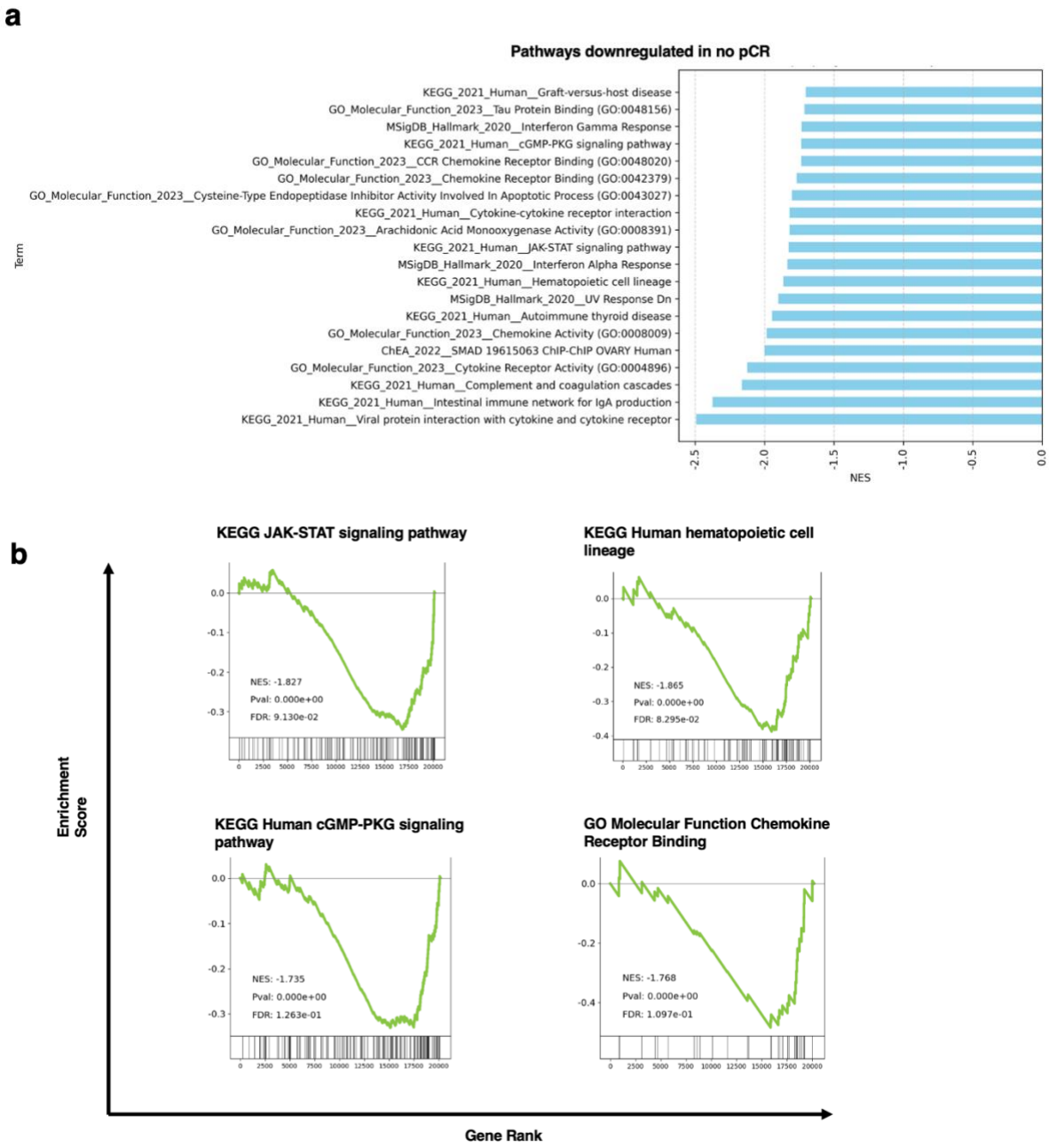

Figure 1.9. Gene Set Enrichment Analysis shows specific immune and homeostatic pathways downregulated in no pCR.

a) Gene set enrichment analysis was performed on a genome-wide ranked list based on the differential hydroxymethylation of all genes between no pCR and pCR. Bar plot of top pathways from KEGG, Hallmark 2020, ENCODE_Histone_Modifications_2015_H3K36m3 and GO Molecular Function gene sets are displayed. **b)** Enrichment plots of four selected pathways.

Integrated 5hmC-based classifier has diagnostic and prognostic utility

Our main goal was to create an integrated diagnostic model using urine 5hmC cfDNA profiles to distinguish between patients who achieved pathological complete response (pCR) and those who did not. To achieve this, we first conducted differential expression analysis to identify differentially hydroxymethylated regions (DhMRs) in patients with pCR (n=35) and those without pCR (n=49), as well as between presumed cancer-free patients (n=75) and those without pCR (n=49). This approach provided a wide range of relevant biological genes for investigation. We then selected 500 genes from both comparisons and applied elastic net regularization for feature selection to identify the most important markers (Fig. 1.10). This process consistently identified a final panel of 25 candidate marker genes, which were chosen for further validation (Fig. 1.11a). The list included several genes known from the literature for their diagnostic and prognostic value in genitourinary cancers[53,54]. A logistic regression coefficient was assigned to each gene: a positive coefficient indicated a higher likelihood of the gene contributing to a pCR prediction, while a negative coefficient suggested a higher likelihood of contributing to a no pCR prediction (Fig 1.11a). We performed ROC analysis using leave-one-out cross-validation to assess the diagnostic utility of our model. The analysis revealed an excellent discriminative ability, with a ROC AUC of 0.91 (Fig. 1.12b), significantly surpassing the performance of the TFx-based classifier (Fig. 1.4).

We then turned to The Cancer Genome Atlas (TCGA) to evaluate the potential of the candidate marker genes as novel biomarkers. Notably, we found that one of our candidate diagnostic model genes, BCAT2, exhibited the highest median expression levels in bladder cancer patients compared to other cancer types (Fig 1.12a). Another gene of interest, GSTM1,

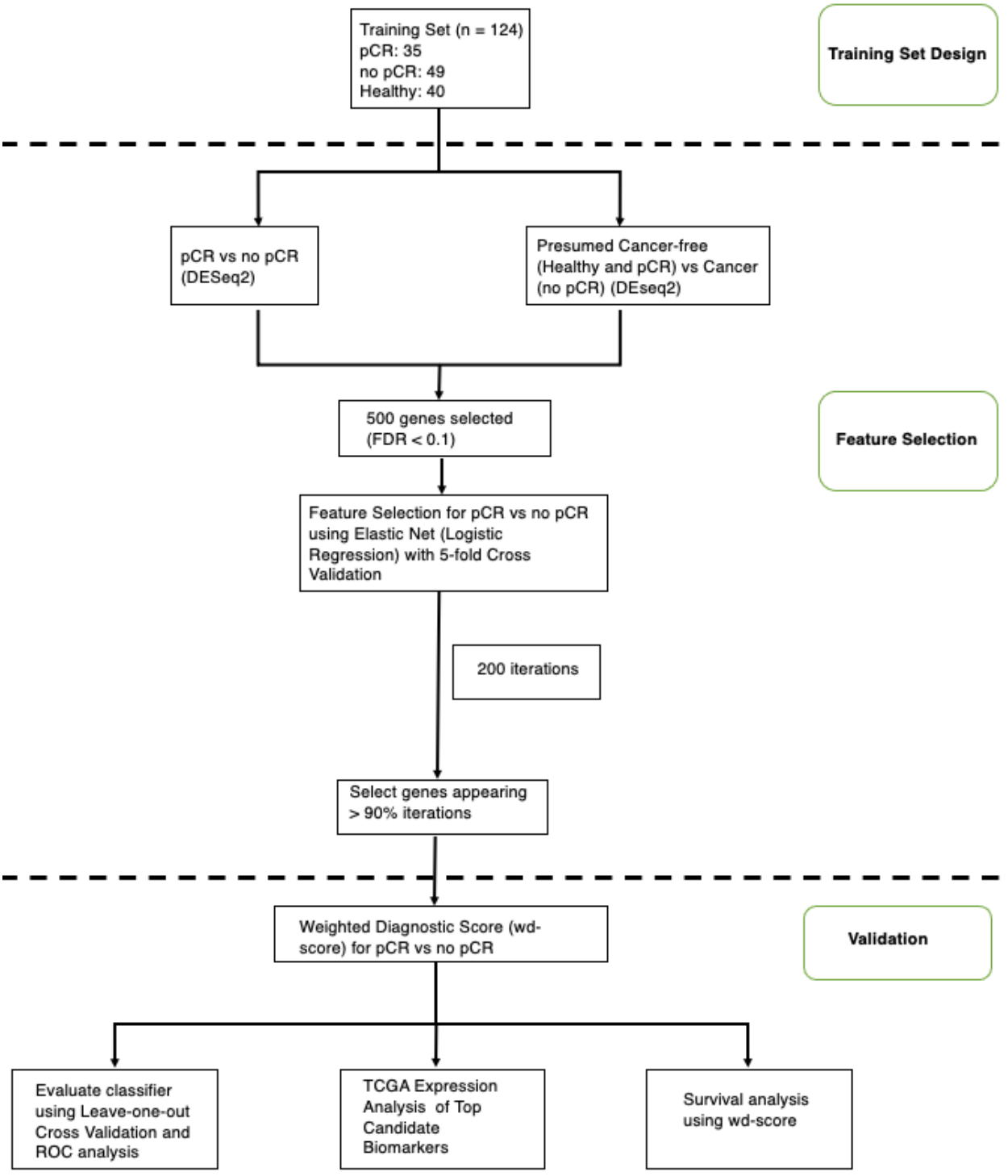

Figure 1.10. 5hmC-classifier model schema.

Schema for developing a 5hmC-based classifier model to distinguish pCR status. While healthy patients were used for initial feature selection, the final model was tested only on pCR and no pCR patients(n=84). pCR pathological complete response.

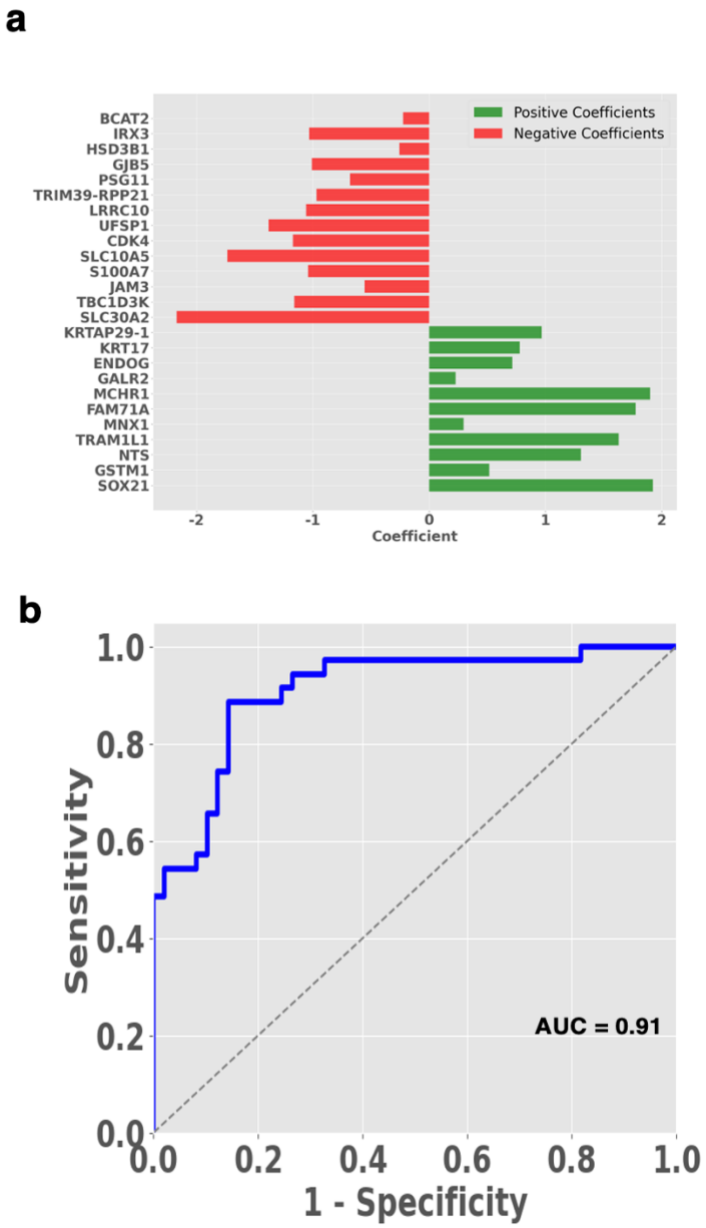

Figure 1.11. 5hmC model identifies informative genes and successfully predicts residual disease.

a) Bar plot showing logistic regression coefficients associated with each gene based on the optimized logistic regression model developed based on Figure 1.10. Positive coefficients increase probability of pCR and negative coefficients decrease probability of pCR. **b)** ROC analysis depicting performance of this final logistic regression model to predict pCR status using leave-one-out cross validation. AUC area under the curve.

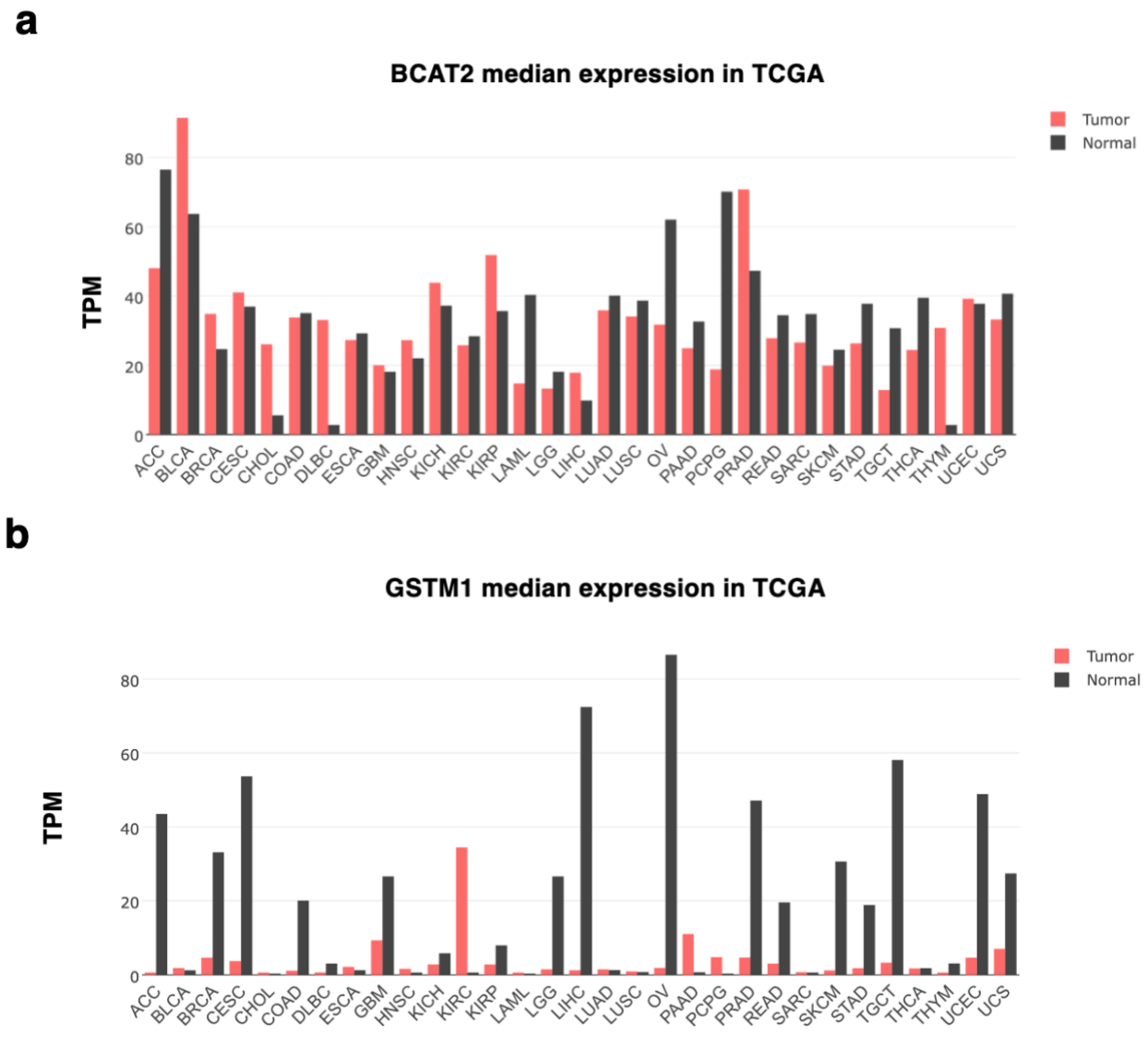

Figure 1.12. 5hmC model identifies genes that are independently validated in the Cancer Genome Atlas.

a) Bar plot of median expression of BCAT2 gene across various cancer types and their matched normal tissue. **b)** Bar plot of median expression of GSTM1 across various cancer types and their matched normal tissue. BLCA bladder cancer.

was consistently upregulated in normal tissues across various cancers in the TCGA dataset, supporting the positive coefficient assigned to it in our diagnostic model. GSTM1 has also been highlighted for its prognostic significance in bladder cancer in multiple independent studies[55].

We calculated Weighted-Diagnostic (WD) scores based on this panel of 25 marker genes and the logistic coefficients identified in Figure 1.11a. We categorized patients into WD+ and WD- groups based on whether their WD-score was above or below 0, respectively. Then, we performed Kaplan-Meier analyses starting from the time of surgery. Progression-free survival ($p = 0.016$) and overall survival ($p = 0.03$) were both significantly different between the two groups (Figure 1.13). Thus, our 5hmC classifier showed greater prognostic utility than the TFx-based approach (Figure 1.5 a-b).

Finally, we performed multivariate Cox regression with demographic factors (age, sex, smoking history) alongside key molecular markers(binarized MRD status, binarized WD-score). Only, WD-score had a significant p-value ($p = 0.03$), demonstrating that our 5hmC-diagnostic model was more effective than other known factors in predicting patient outcomes (Figure 1.14).

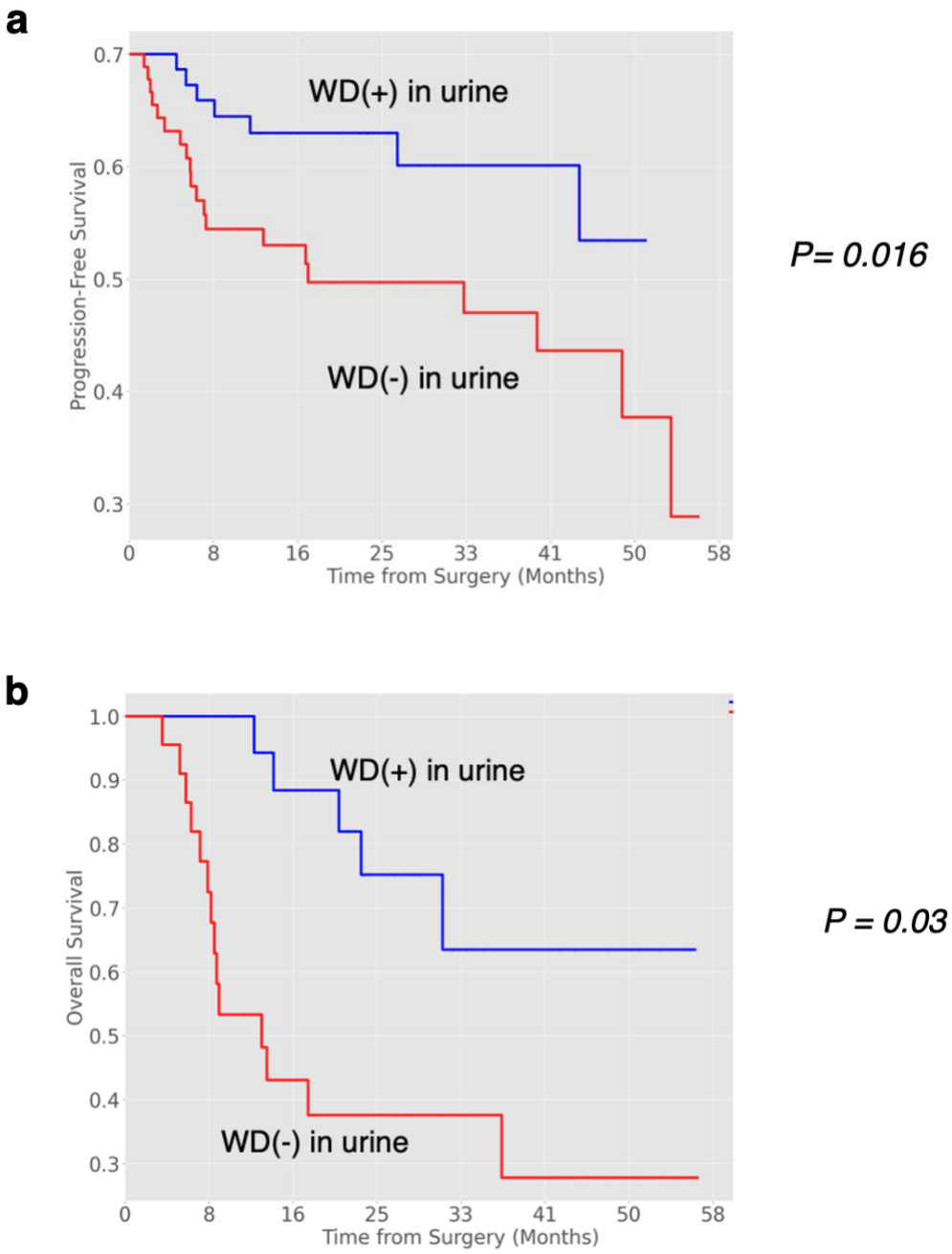

Figure 1.13. Survival analysis using 5hmC weighted diagnostic(WD) score is predictive of patient outcomes.

Kaplan-Meier plots showing **a)** progression-free survival and **b)** overall survival stratified by WD-score status (WD> 0 vs WD <=0). *p* values were calculated by the log-rank test. WD weighted diagnostic score. N=84 patients.

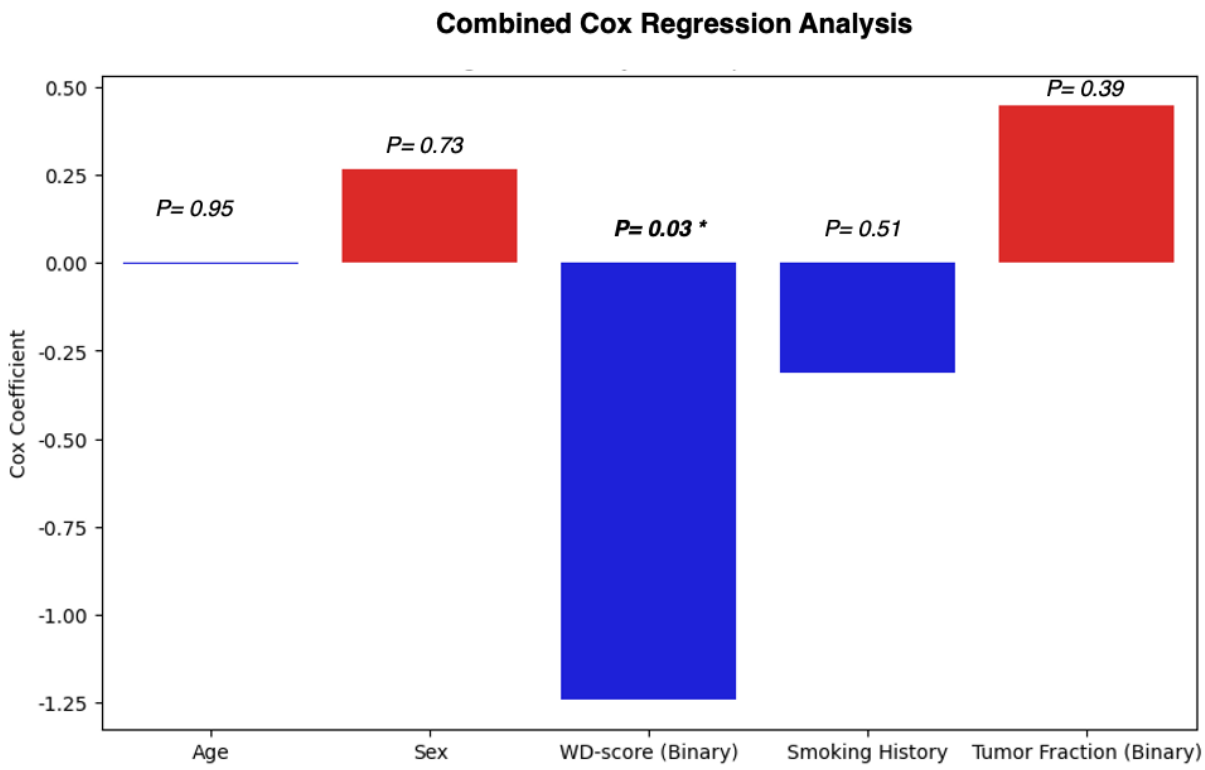

Figure 1.14 5hmC WD-score is more significant than other covariates using multivariate cox regression analysis.

Cox regression analysis was performed on n=70 patients (41 no pCR and 29 pCR). Age, Sex, WD-score (WD+ vs WD-), Tumor fraction (MRD + vs MRD-) based on figure 1.5, and smoking history (binarized) were used as covariates. WD weighted diagnostic score. MRD molecular residual disease.

DISCUSSION

Our study demonstrates the significant potential of using urine-derived 5-hydroxymethylcytosine (5hmC) profiling and a copy number-based approach for the detection of minimal residual disease (MRD) in bladder cancer patients. These novel methods present a breakthrough in urine cfDNA-based bladder cancer diagnostics, offering noninvasive, specific, and sensitive strategies for assessing MRD and predicting patient outcomes. This represents an important advancement over the previous, mutation-based uCAPP-Seq approach, which, while successful, required both blood and urine samples and offered a sensitivity of 81% with an ROC AUC of 0.78[39].

Our low-pass whole genome sequencing (LP-WGS) method, coupled with the ichorCNA computational tool, effectively estimated tumor fraction (TFx) in urine cfDNA. Patients who achieved complete pathologic response (pCR) exhibited significantly lower TFx levels compared to those who did not achieve pCR. We established a TFx threshold of 6.2% for identifying MRD, resulting in an optimal balance between sensitivity (65.8%) and specificity (86.2%). Although the TFx approach demonstrated strong diagnostic potential, it did not exhibit significant prognostic value for predicting progression-free or overall survival.

In contrast, our novel 5hmC profiling method in urine cfDNA yielded insights into gene and pathway modulation related to bladder cancer pathology. Differential expression analysis and elastic net regularization identified a panel of 25 candidate marker genes with potential diagnostic and prognostic value. This approach demonstrated excellent discriminative ability, with a ROC AUC of 0.91, surpassing the diagnostic performance of the uCAPP-Seq approach and the TFx-based classifier (AUC of 0.78). The weighted diagnostic (WD) scores calculated from the 25

marker genes provided significant prognostic utility for progression-free and overall survival, further underscoring the value of 5hmC profiling.

Both approaches rely solely on urine samples, avoiding the need for matched blood samples as required in the uCAPP-Seq approach, thus reducing patient burden and offering a more convenient method of diagnosing disease. This work provides a proof-of-concept for methods that could be applied to noninvasively studying disease progression. The ability to track transcriptional changes and copy number variations in urine cfDNA allows for precise monitoring of disease progression and treatment efficacy. The combination of 5hmC profiling and TFx estimation can guide personalized treatment strategies, providing tailored care based on individual tumor biology.

There are certain limitations to our work. While we used survival analyses as a way to validate our diagnostic approaches, our urine cfDNA analyte was captured pre-surgery. Given that curative-intent radical cystectomy typically diminishes the risk of both disease progression and death, it makes it challenging to assess prognostic implications. A number of MRD-positive patients might have had their elevated risk of disease progression fully addressed by surgery. This could explain why TFx alone was an excellent diagnostic approach for detecting MRD but could not adequately stratify patient outcomes post-surgery. However, this underlies the promising nature of our 5hmC-classifier which was both diagnostic and prognostic. It is intriguing to speculate that the 5hmC model could be shedding light on certain molecular-level changes in patients that were not fully addressed by surgery alone and could require further intervention. Another limitation to our work is that we worked with single timepoint assessment of urine cfDNA which constrains both sensitivity and specificity.

Future research should expand upon these findings with larger cohorts at multiple centers to validate the diagnostic and prognostic capabilities of these methods. Rather than using genome-

wide 5hmC profiling, it would be more cost-effective and efficient to work with targeted panels of marker genes (like the 25-gene marker panel we identified). Serial collections could be used to boost the strength of these predictive approaches.

These noninvasive diagnostic methods have the potential to be integrated into standard care pathways for bladder cancer patients, enabling earlier detection of recurrence and guiding treatment decisions. Given the successful application of these methods in bladder cancer, there is potential for their use in other genitourinary and other types of malignancies, which could improve diagnostics and patient outcomes across various cancer types.

In summary, our study provides strong evidence that urine-based 5hmC profiling and copy number-based estimation of TFx offer robust tools for MRD detection in bladder cancer. By advancing these methods, we contribute to the development of noninvasive, precise, and clinically useful solutions for managing bladder cancer.

CHAPTER 2

INTRODUCTION

Overview of Radiation Therapy In Cancer

Cancer continues to be the deadliest disease globally and is driven by a range of predisposing factors including genetic, environmental, and pathogenic[56]. It arises from the uncontrolled proliferation of actively dividing cells from nearly any organ system in the body, leading to the formation of neoplasms. A variety of treatment modalities exist to combat cancer, including chemotherapy, radiation therapy (RT), surgery, immunotherapy, targeted therapy, hormone therapy, and personalized cell therapies. While there is a broad spectrum of options, chemotherapy, surgery and RT remain the primary pillars of treatment.

Radiation therapy is a cornerstone of cancer treatment, utilized in 50-60% of all cancer patients[56,57]. RT can be administered with curative intent or as a palliative measure to alleviate symptoms. RT is particularly effective for treating cancers such as skin, prostate, lung, and head and neck carcinomas, which can often be cured with radiation therapy alone[57]. Meanwhile, other types of cancer such as breast, rectal and anal, and central nervous system tumors, as well as locally advanced head and neck carcinomas, may require RT in combination with other treatment modalities for optimal results.

RT uses ionizing radiation to cause DNA damage, impacting both cancerous and healthy cells. The concept of dose fractionation was introduced to minimize normal tissue toxicity by delivering smaller doses of radiation daily[57]. Curative fractionated radiotherapy for localized cancers typically involves daily doses of 1.5–2.2 Gy for a total dose of 60–80 Gy, depending on the tumor type and the radiation sensitivity of surrounding normal tissues[57]. Recent advances in imaging, delivery and organ motion tracking, have led to the development of stereotactic body

radiotherapy (SBRT)[57]. SBRT makes it possible to deliver doses ranging from 20-34 Gy in a single treatment with ultra-precise targeting of tumor lesions while minimizing damage to healthy tissue.

Use of radiation in oligometastatic disease

Metastasis is the leading cause of cancer-related mortality and represents one of the major challenges in the fight against cancer[61]. Traditionally, patient disease burdens have been categorized into two distinct categories: localized disease, where curative treatments like surgery, RT and chemotherapy are viable options and metastatic disease, which has been approached with systemic therapies. However, this binary classification fails to address the continuum of disease progression and the potential for more nuanced treatment strategies. Weichselbaum et al. introduced the term "oligometastasis," challenging the conventional classification of patient burden by proposing an intermediate state characterized by a limited number of metastatic lesions confined to specific locations[62]. There is increasing interest in the potentially curative treatment of oligometastatic disease or in patients with more widespread disease by combining SBRT with immune checkpoint blockade (ICB)[63,64,65]. Despite the potentially synergistic immune-activating mechanisms of RT and concurrent ICB[66], combination clinical trials have been unsuccessful, though the underlying mechanisms for this remain unclear.

Through induction of DNA damage and release of immunogenic damage-associated molecular patterns (DAMPs), RT can also stimulate innate and adaptive immunity[57]. On the flip side, RT may also influence the production of certain secreted factors and cytokines that result in immunosuppression through the induction of regulatory T cells (Tregs) and myeloid-derived suppressor cells (MDSCs), which can have an adverse impact on the outcomes of RT treatment[67,68].

The failure of RT to cure patients is has been linked to the persistence of primary tumors or the spread of cancer to distant sites. The Weichselbaum group has proposed a hypothesis to explain the limited efficacy of RT in such cases, particularly in tumors with concurrent metastatic disease or a high tendency to metastasize. They suggest that RT might induce the release of soluble factors in the tumor or its surrounding stroma that could facilitate the growth of distant tumors. This phenomenon has been termed the "badscopal" effect[68,69].

Identification of an EGFR ligand correlating with distant metastases in clinical trials

Piffko et al. profiled the gene expression of patients with advanced solid tumors, who received SBRT to multiple metastatic sites at an institutional clinical trial at the University of Chicago Medical Center (NCT02608385)[70,71]. In 22 matched pre- and post-RT biopsies from irradiated metastases, 60 overlapping genes were found that were both induced by SBRT and positively correlated with growth at distant, unirradiated sites. One of the top hits from this analysis was amphiregulin(AREG)[70]. AREG is EGFR-ligand and can activate signaling pathways involved in cell growth and survival[73]. When the patient cohort was stratified by change in *AREG* expression in response to SBRT, patients whose tumors exhibited increased *AREG* expression had worse survival and progression-free survival suggesting that AREG had a deleterious role in this patient group[70].

AREG is induced by RT and leads to metastatic growth in transplantable tumor models

In order to translate the clinical finding to a preclinical metastasis model, Piffko et al. used the murine Lewis lung carcinoma (LLC) model of spontaneous lung cancer metastasis[70]. LLC

tumors were subcutaneously implanted in the flank of C57BL/6 wildtype (WT) mice and treated with varying doses of local irradiation (5, 10 and 20 Gy)[70]. The resulting tumors showed a high degree of radiation-resistance at these doses. Additionally, Piffko et al. demonstrated that AREG is induced by RT and released into circulation[70]. Intriguingly, they also observed the growth of preexisting metastatic lesions in their model. Piffko et al. found that *Areg* KO tumors produced significantly smaller and fewer lung metastases and the size of LLC^{AR-} lung metastases did not increase following local flank tumor RT[70].

AREG modulates monocytes in patients and murine models

Piffko et al. demonstrated that were no tumor-intrinsic changes in growth or migratory potential upon *Areg* deletion highlighting a potential role for the host-immune response in the AREG-induced metastatic growth phenotype. AREG induces phosphorylation of EGFR at tyrosine residue 992 (Tyr992)[74]. Intriguingly, they found an increase in the frequency of $pEGFR^{+}$ $CD33^{+}$ $CD14^{+}$ monocytes in peripheral blood mononuclear cells (PBMC) of metastatic non-small cell lung cancer (NSCLC) patients treated with SBRT in a second independent clinical trial at the University of Chicago(NCT03223155)[70,75]. Increase in the $CD33^{+}$ $CD14^{+}$ $pEGFR^{+}$ monocytes in patients correlated with significantly worse prognosis[70]. Spectral flow cytometry was also performed on the lungs of LLC tumor-bearing mice. Cell types with highest pEGFR signal were $CD11b^{+}$ $Ly6G^{+}$ granulocytes and $Ly6C^{+}$ $F4/80^{+}$ monocyte-derived macrophages[70]. The $Ly6C^{+}$ $F4/80^{+}$ monocyte-derived macrophages exhibited a unique increase after RT in LLC^{AR+}, but failed to increase in the lungs of LLC^{AR-} tumor-bearing mice post-RT suggesting a possible link between these monocytes/macrophages and RT-induced AREG in the metastatic proliferation observed in LLC^{AR+} models[70].

Open questions to be addressed by single-cell RNA sequencing

Given the findings from Piffko et al. suggesting a potential role of monocyte-derived macrophages in mediating RT-induced AREG driven metastatic proliferation in mice, it is crucial to gain a deeper understanding of the specific mechanisms within the myeloid populations in the lungs. Single-cell RNA sequencing (scRNA-seq) provides a powerful approach to dissect the cellular heterogeneity and intricate transcriptional profiles of these myeloid cells. By examining individual cells at high resolution, this work aims to identify distinct subpopulations, their gene expression patterns, and their potential interactions with other immune populations. Furthermore, this work also aims to study trajectories of myeloid cell differentiation and activation, offering insight into how AREG signaling may influence these immune cells and their roles in the metastatic niche. In concert with the previous findings from Piffko et al., this work aims to provide critical insights into the molecular underpinnings of myeloid cell involvement in metastatic growth, potentially guiding precision medicine strategies for SBRT therapies in the future.

METHODS[a]

Single-cell RNA sequencing(scRNA-seq) library preparation

LLC^{AR+} and LLC^{AR-} flank tumor bearing mice were treated with 20Gy once flank tumors reached 100mm. Mice were sacrificed and lung tissues were collected five days post treatment. Single cells were harvested by digestion and $CD45^{+}$ cells were purified by AriaIII cell sorter (BD Biosciences). Four mice were pooled per biological condition (LLC^{AR+}, LLC^{AR+}+20 Gy, LLC^{AR-}, LLC^{AR-}+20 Gy) and GEX (Gene Expression) and HTO (Hashtag Oligos) libraries were generated for each biological condition and demultiplexed using TotalSeq™ hashtag antibodies (Biolegend). *Performed by Andras Piffko*.

scRNA-seq alignment

For each of the four conditions, both the GEX and HTO libraries were concurrently aligned using *cellranger count* (Cellranger version 7.1.0 from 10X genomics) with the "include-introns" option set to be "true". All four conditions had high alignment proportions with the reads mapped to the mm10-2020A genome ranging between 91-94%. Barcode processing, filtering, UMI counting and aggregation of sequencing runs were also performed using the Cell Ranger analysis pipeline (version 7.1.0 from 10X genomics). Downstream analyses were primarily performed in Python using Pegasus(version 1.8.1)[76], Scanpy (version 1.8.1)[77] and Seurat (version 4.4.0)[78].

[a] Much of this section has been reproduced, with some modification, from Piffko A., Yang K., Panda A., and Weichselbaum R.R. "Radiation-induced amphiregulin drives tumor metastasis." Unpublished manuscript, last modified Jan 3rd 2024. Microsoft Word file.

scRNA-seq doublet detection and filtering

Gene-expression data for cell barcodes from all four biological conditions were pooled together into a single Pegasus Unimodal object. A multi-step approach was used to perform quality control. For each cellular barcode, two metrics were calculated: total number of genes detected and proportion of mitochondrial UMIs (*mito_fraction*). Cell barcodes with <200 genes or *mito_fraction* > 20 were considered to be poor-quality transcriptomes or dying cells and were excluded. Next, we normalized the HTO counts across the filtered cells using the centered log-ratio (CLR) transformation and then demultiplexed using the *HTODemux()* function in Seurat R package and by setting the *positive.quintile* parameter to 99. All cells with > 6000 genes detected or multiple HTO barcodes identified by the HTO classifier were considered multiplets and were filtered out, yielding us a final set of 32,690 $CD45^+$ cells.

To assess technical variability between mice arising from the same biological conditions, we first performed dimensional reduction on the entire transcriptome using principal components analysis (PCA). Using the elbow method, the first 50 PCs were deemed to be significant and a t-SNE projection was subsequently generated using these 50 PCs. A mouse-level t-SNE projection was generated by averaging the t-SNE coordinates of all cells originating from each single mouse and this method was used to visualize global transcriptomic shifts induced by the presence of IR as well as AREG.

scRNA-seq batch correction and cluster identification

Next, we took a multi-step approach to clustering cells for cell type identification in the 32,690 $CD45^+$ cells using the Pegasus package. First, 18,931 robust genes were identified based on expression in at least 5% of all cells in our dataset. Normalized gene expression counts for

robust genes were calculated by scaling counts so that each cell had the same sum of total gene counts (10^5) and then a log-transformation was performed. Then, genes were ranked as "highly variable" based on having moderate mean expression but high dispersion (computed via the loess smoothing method) using Pegasus. Data dimensionality of the dataset was reduced by selecting the top 4000 most highly-variable, robust genes. Next, PCA was calculated on this dimensionally-reduced transcriptome and top 50 PCs were used for downstream analysis. Batch correction was performed using the *run_harmony* function, which implements the Harmony algorithm[79], using top 50 PCs as inputs and *batch* parameter set to the HTO barcode variable which maps a unique barcode to each individual mouse in the experiment. Next, we constructed a k-Nearest-Neighbor (kNN) graph where 30 nearest neighbors were calculated for each cell. Unsupervised clustering was performed using the Leiden algorithm, a modularity optimization algorithm, using a resolution parameter of 0.8 and yielded 21 cell clusters of 11 major lineages. UMAP (Uniform Manifold Approximation and Projection) was performed and used to visualize the clusters in a 2-D space. To identify significant cluster specific gene markers, we performed differential gene expression using the Mann-Whitney U (MWU) test to compare gene expression in the cluster of interest against all other clusters and retained only the genes with MWU q-value <0.05. Cell type labels were assigned by a combination of manual curation by immunology experts and auto-annotation tools in Pegasus.

For mononuclear phagocyte (MNP) focused-analyses, MNP clusters (4095 cells) were selected and previously described steps were repeated (PCA, kNN, leiden with resolution of 0.8) to yield 17 cell clusters. Marker genes were identified as previously described using the MWU test to compare clusters and using q-value of 0.05 to threshold out significant genes. Clusters were visualized using UMAP. Biological condition-specific MNP densities were visualized by

calculating the kernel density estimate of the UMAP embedding using the *kdeplot* function from the Python seaborn library with *thresh* set to 0 and *levels* set to 15.

scRNA-seq Trajectory Analyses

Inflammatory monocyte (IM) populations of interest were manually selected from the MNP population (2600 cells) and underlying differentiation/developmental relationships between these populations were explored by computing the diffusion map (top 100 diffusion components) using the Pegasus package (default parameters were used). Pseudotime was computed in Pegasus to estimate relative progression of cells along the differentiation/developmental processes identified in the diffusion map. Next, input the top 10 diffusion components of the IM populations into the Slingshot R package[80] to infer likely trajectories. The *slingshot* function and *getLineages* with default parameters and IM population 1 as starting population were utilized.

scRNA-seq RNA velocity

The fraction of spliced and unspliced mRNAs for all detected genes was calculated using the Velocyto package (version 0.17.17)[81]. Using these fractions, RNA velocities for the manually-selected 2600 MNP cells, were then computed using the scVelo package (0.3.2)[82]. Specifically, the top 4000 genes were selected and moments were calculated with the number of principal components set to 30 and the number of neighbors set to 15. The dynamical method was used to learn the full transcriptional dynamics of splicing kinetics.

scRNA-seq gene signature calculations

While comparing mean gene expression between IM clusters (either individual gene markers or gene signatures), we applied a more stringent cut-off and filtered out cells with *mito_fraction* > 5 to enable the most accurate comparisons. Gene signature calculations were performed using manually curated signatures from literature as well as gene sets from the Reactome[44] and the KEGG gene sets[45]. Single-cell signature scoring was performed using a well-established gene-scoring methodology[76] method. The gene-scoring method employs a strategy aimed at minimizing the impact of technical variability while amplifying biological signals. It categorizes all genes in the genome into 50 bins and then assesses how many genes from the gene signature E fall into each bin. Next, it computes a weighted sum of normalized counts for each gene in signature E. This sum is then subtracted from a weighted sum derived from 1000 randomly generated "null" signatures, which are gene signatures of the same size as the original set E and share its bin distribution.

scRNA-seq cell-cell interaction

Inference of receptor-ligand interactions was performed using the CellPhoneDB Python library. The algorithm was run on log-normalized expression values for the entire CD45 compartment clusters with so subsampling.

Data Availability

All raw sequencing data are available online (GSE230369).

RESULTS[b]

scRNA-seq data was high quality with some biological condition-specific variation

To explore the role of immune cells in AREG-dependent lung metastasis growth, we conducted single-cell RNA sequencing (scRNA-seq) on lung tissues five days after radiotherapy (RT) of both LLC^{AR+} and LLC^{AR-} flank tumors. Our study specifically focused on the CD45 compartment, allowing us to assess the immune cell dynamics involved in the metastatic process.

We first assessed the quality of the sequencing by evaluating a number of quality-control metrics. Common metrics include library size, mapping ratio to the reference genome, number of genes, number of total transcripts detected per cell and fraction of reads mapping to mitochondrial genome(*mito_fraction*)[77,78]. All four biological conditions demonstrated sufficient library sizes, and their mapping proportions ranged between 91-94% (data not shown). High *mito_fraction* may represent cellular stress, apoptosis or poor RNA quality as mitochondrial genes are more resistant than nuclear genes, and thus high mitochondrial fraction can be considered indicative of poor cell quality[77,78]. When examining the relationship between the number of detected genes and transcripts within each cell, we observed a pattern where cells with low gene counts tended to exhibit high mitochondrial fractions, aligning with our expectations (Fig 2.1a). The number of detected genes showed a consistent distribution in the AREG-knockout conditions, while the LLC^{AR+} contained the greatest subpopulations with low detected genes (Fig 2.1b). The *mito_fraction* distributions were relatively consistent across all biological conditions (Fig 2.1c).

[b] Much of this section has been reproduced, with some modification, from Piffko A., Yang K., Panda A., and Weichselbaum R.R. "Radiation-induced amphiregulin drives tumor metastasis." Unpublished manuscript, last modified Jan 3rd 2024. Microsoft Word file.

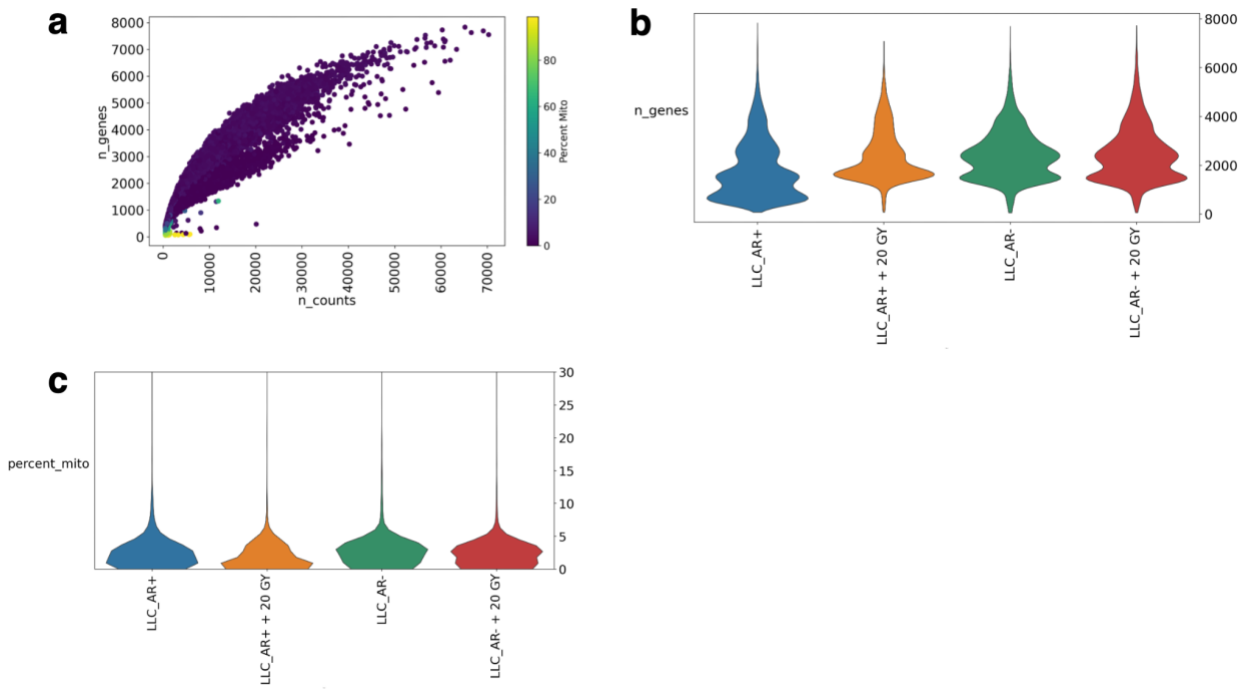

Figure 2.1 Single-cell RNA sequencing quality control metrics.

a, Scatterplot of the number of counts found in each cell (x axis) versus the number of genes (y-axis) that were detected. In addition, the color bar indicates the fraction of mitochondrial reads found in each cell. n = 38,321. **b,** Violin plot representation of the total number of genes detected in each condition. **c,** Violin plot representation of the percentage mitochondrial reads per cell in each condition. Minimal tail distributions above 30% were excluded for visualization purposes.

Filtering and hashtagging were used to identify high-quality cells for downstream analysis

After assessing various quality control metrics, we applied specific filtering thresholds to refine our scRNA-seq data. A key aspect of scRNA-seq data analysis is the identification and exclusion of multiplets and low-quality cells. Doublets occur when two or more cells get inadvertently captured and sequenced together leading to mix transcriptomes and confounding data analysis[77,78]. We removed cells with more than 6,000 detected genes, as they were presumed to be doublets due to a notable discontinuity in the distribution at that point (Fig. 2.1a). Additionally, we set the lower bound of detected genes at 200. Finally, we applied a mitochondrial content threshold, removing cells with a *mito_fraction* exceeding 20% (Figure 2.2a). We also identified the most variable genes in our dataset, which included S100a8, S100a9, Cd74, and Cxcl2 (Figure 2.2b). These genes indicate that myeloid and inflammatory signals were the dominant biological signals in our data.

A second method to exclude doublets utilizes hashtag oligos(HTO), a method that involves tagging each cell with a unique molecular barcode during library preparation so that individual samples can be identified during demultiplexing. We assigned unique barcodes to each individual mouse (n=4) within each biological condition. These barcodes not only enabled accurate tracking of mice but also served as a valuable tool for deconvolving mixed-cell transcriptomes, providing an additional method to detect and filter out doublets. Utilizing these barcodes, we conducted k-medoid clustering on normalized HTO values for each unique hashtag oligo (HTO). Subsequently, we generated a "negative" distribution for each HTO and established a threshold using the 0.99 quantile of this distribution. Any value exceeding this threshold was classified as HTO-positive. While the vast majority of cells were HTO-positive for a single

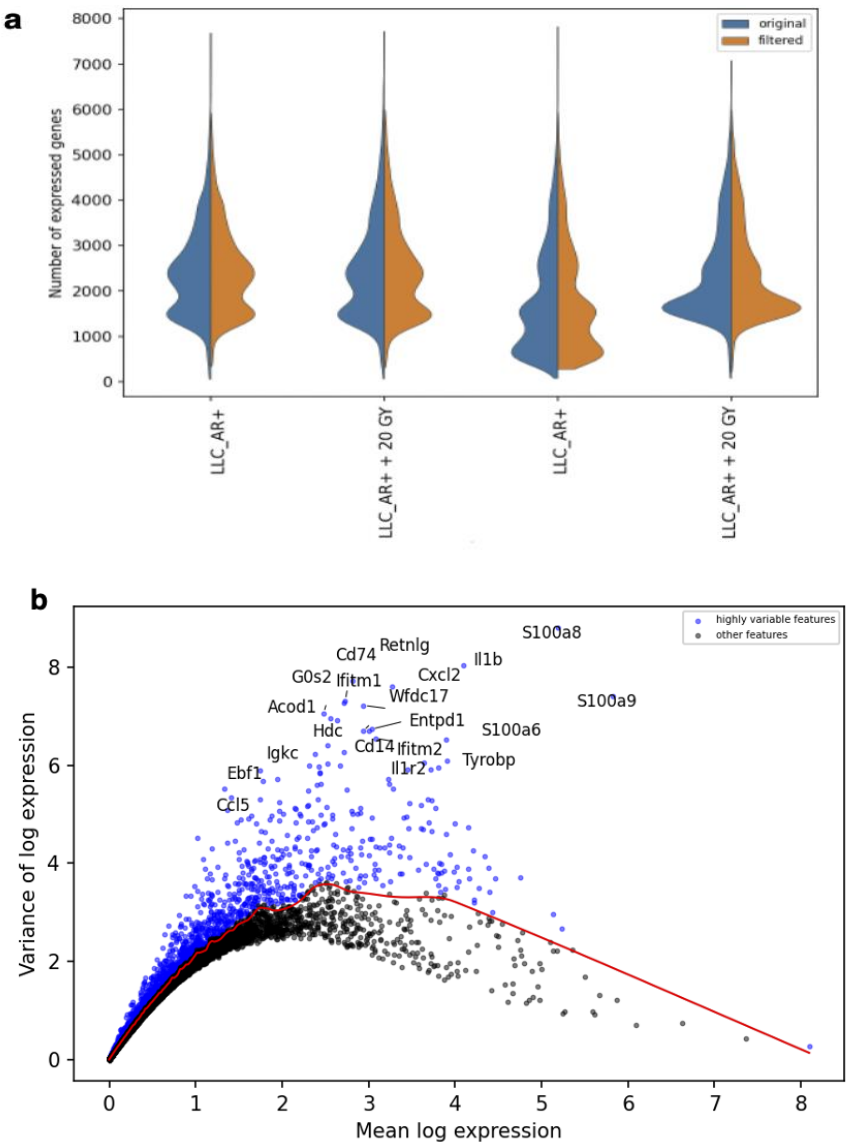

Figure 2.2. Filtering and identification of highly variable genes.

a, Violin plot of the number of genes found in each condition before and after exclusion of low (n_genes>200) and high-expressing genes (n_genes < 6000). **b,** Scatterplot of the most variable genes across the datasets where the x axis is mean log expression and y is variance of log expression. Top 20 genes are labeled.

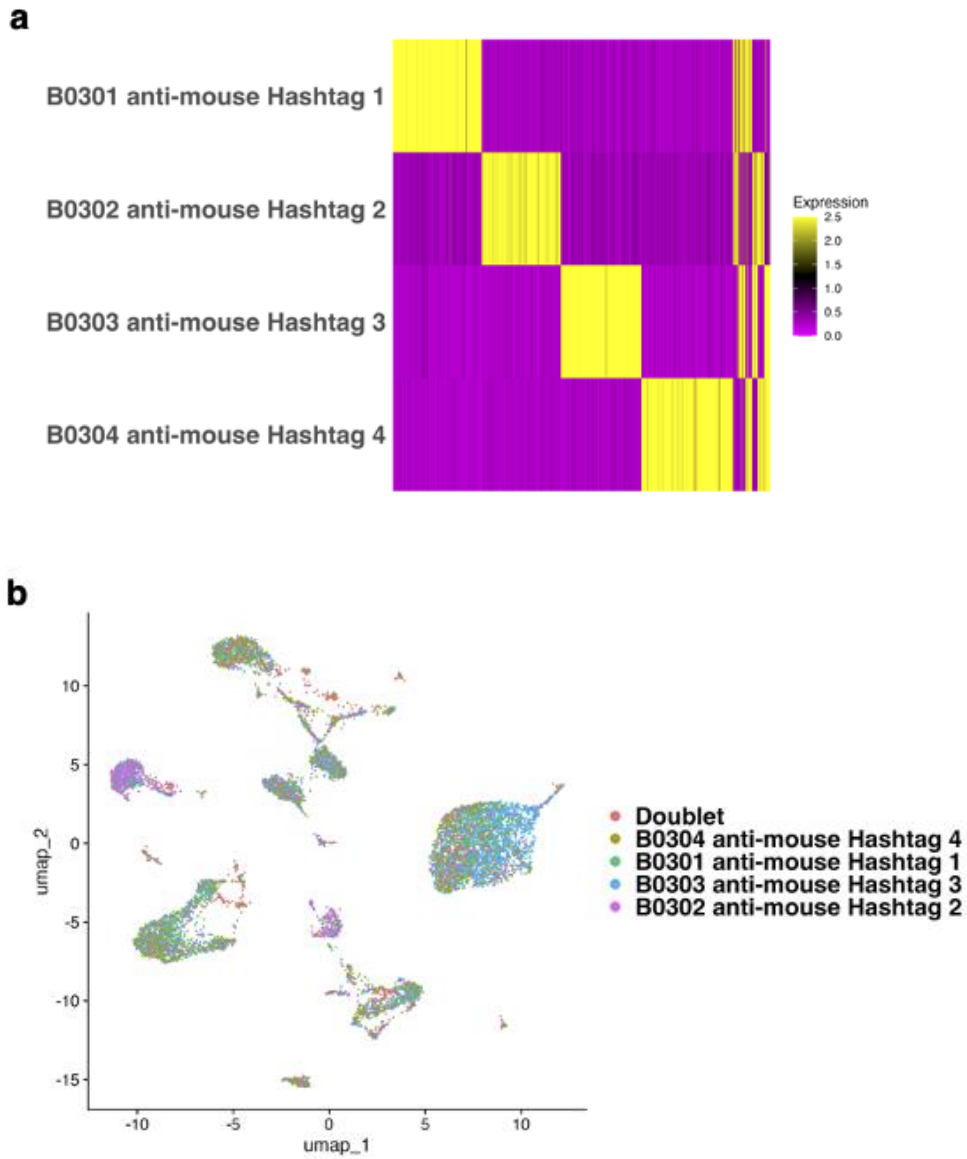

Figure 2.3. Demultiplexing cells using HTOs (hashtag oligos).

a, A row-normalized representative heatmap showing expression intensity from each antibody hashtag found in the LLC^{AR-} + 20 GY condition comprising 10,447 cells. Each hashtag represents a signal from an individual mouse. **b,** Unsupervised uniform manifold approximation and projection (UMAP) representation of all the cells from LLC^{AR-} + 20 GY condition color coded by dominant hashtag detected each cell.

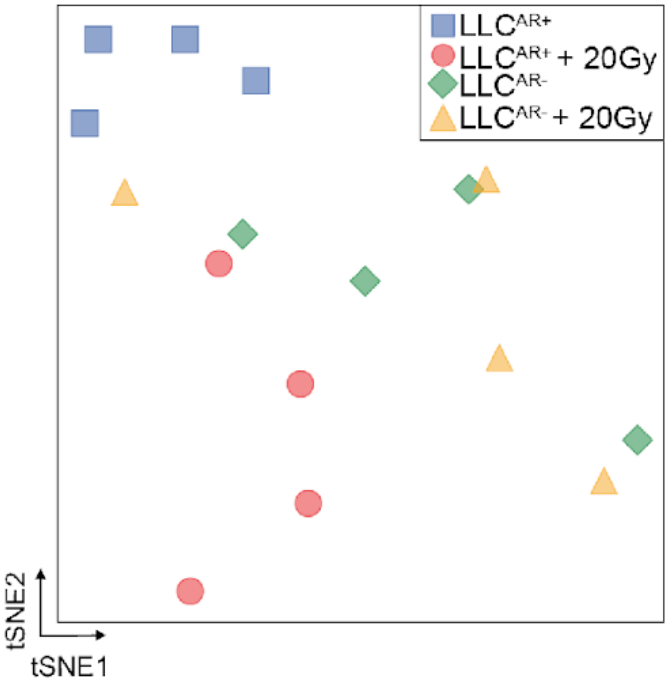

Figure 2.4. Global transcriptional profile of individual mouse replicates.

t-SNE projection of the global transcriptomic profile of 16 individually hashtagged mice from four experimental conditions.

hashtag, we did identify multiplets (Fig 2.3a). UMAP (Uniform Manifold Approximation and Projection) was performed and used to visualize the clusters in a 2-D space and showed each cluster contained mixed populations from each individual mouse (Fig 2.3b). Having excluded doublets, we asked ourselves whether individual mice in each biological group behaved similarly. We observed tight and distinctive clustering of LLC^{AR+} and LLC^{AR+} + 20 Gy samples and a less pronounced separation of AREG-knockout samples (Fig 2.4).

AREG specifically modulates Mononuclear phagocytes (MNPs)

After removing multiplets and unassigned cells using HTO oligos, we performed unsupervised clustering of 32,690 $CD45^+$ cells. We identified 21 clusters arising from 11 major cell lineages including $CD4^+$ and $CD8^+$ T cells, natural killer (NK) cells, naïve and memory B cells, migratory, plasmacytoid, and conventional type I DCs, alveolar macrophages, and inflammatory monocytes that visualized on a UMAP embedding (Fig 2.5a). Each cluster was well-represented across all biological conditions (2.5b). Differential expression analysis identified canonical subset-defining genes such as *Ifitm3*, *Lyz2*, and *S100a4* for inflammatory monocytes, *Gata3*, *Il1r1*, and *Rora* for ILC2, and *Cd3d* and *Cd3e* for T cells (Fig 2.6c).

We next sought to identify the most biologically active clusters in our population by performing ligand-receptor interaction analysis. Inflammatory monocytes, alveolar macrophages, conventional type I DCs, and migratory DCs emerged as the most active cell types based on the total number of interactions (Fig. 2.7a). Using previously published nomenclature[83,84], we label these five cell types mononuclear phagocytes (MNPs). We next attempted to identify clusters that had high levels of EGFR signaling. A group recently published an AREG-driven EGFR

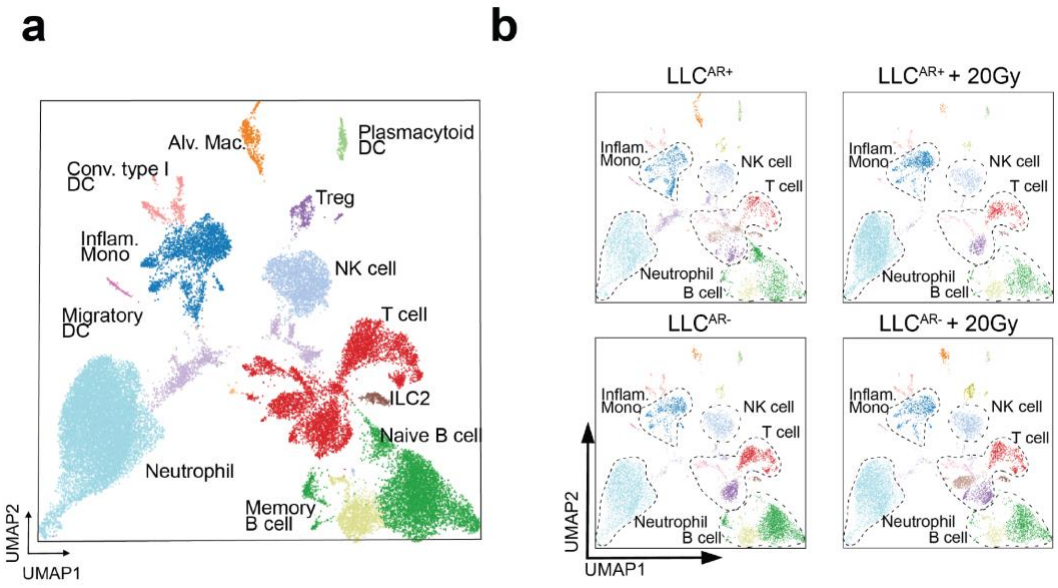

Figure 2.5. UMAP embedding of the filtered, batch corrected CD45 compartment.

a, scRNA-seq and unsupervised uniform manifold approximation and projection (UMAP) clustering of 32,690 live CD45$^+$ cells in the lungs of LLC^{AR+} and LLC^{AR-} tumor-bearing mice ± 20 Gy, *n* = 4 mice per group, annotated as 11 major cell lineages. **b,** UMAP of live CD45$^+$ cells stratified by experimental condition. Inflammatory monocytes, B cells, T cells, NK cells and neutrophils outlined.

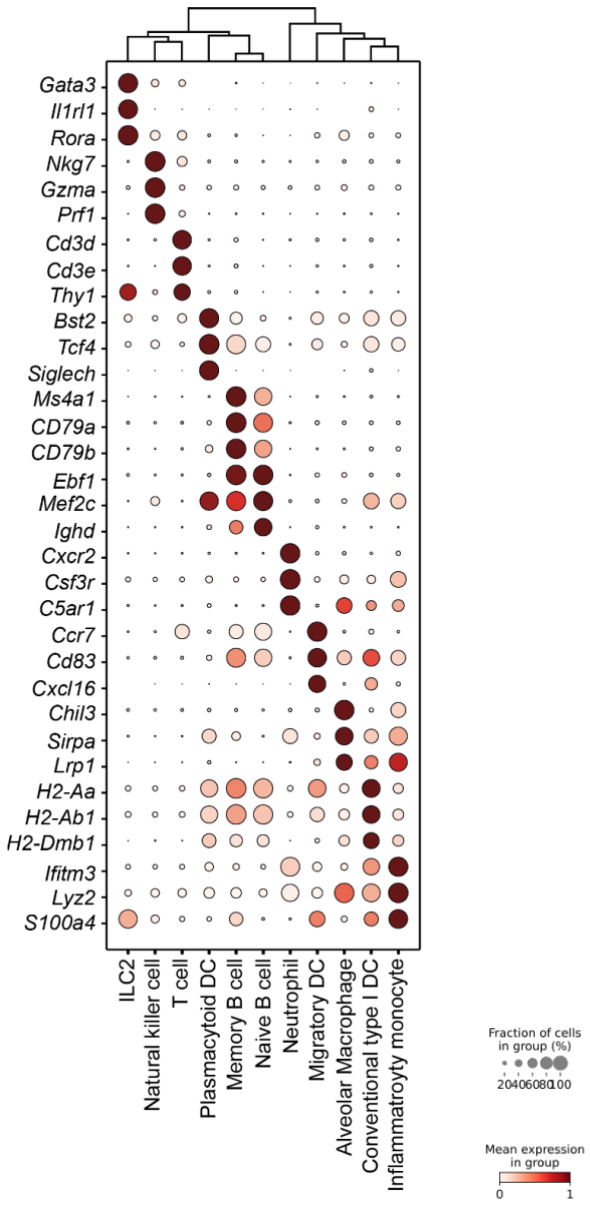

Figure 2.6. Heatmap of Top differential markers.

Bubble plot showing expression of top three marker genes for 11 major cell lineages. DEGs are determined by Wilcoxon rank-sum test and determined in a one-vs-all fashion.

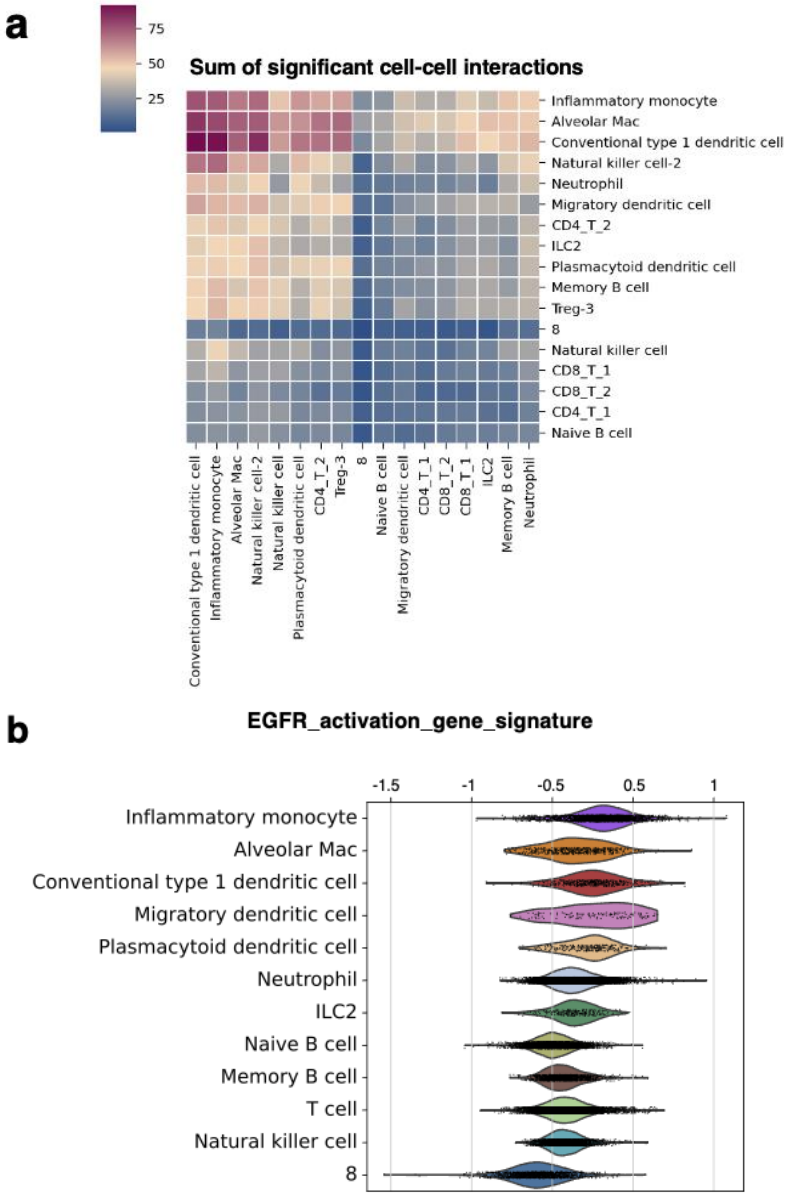

Figure 2.7. EGFR-signaling is increased in monocytes in mouse lungs.

a, Heatmap representing the sum of significant ligand-receptor interactions between all CD45 clusters. **b,** Violin plot representing the gene signature scores of a metastasis promoting EGFR gene signature found in myofibroblasts in single-cell RNA seq of all cell clusters from the CD45 compartment.

signature that was discovered in myofibroblasts that drove metastasis in a model of pancreatic ductal cancer[85]. Employing the same gene signature on all our CD45 clusters, we intriguingly found that MNP populations scored the highest (Figure 2.7b). Taken together, the flow cytometry from human PBMCs[70], mouse lungs[70], as well as the scRNA-seq results demonstrate the following: (i) MNPs show the highest expression levels of Tyr992 pEGFR in PBMC and murine lung immune cells and (ii) an increase of MNPs in metastatic lung tissue correlates with the upregulation of AREG following RT.

AREG leads to immunosuppression in the inflammatory monocyte compartment

We focused our attention to the five CD45 clusters containing MNPs. Subclustering of the MNPs identified 17 populations (Fig. 2.8a) and a distinctive change in the composition of the inflammatory monocyte clusters between the unirradiated and irradiated LLC^{AR+} samples, which was not observed in the AR- conditions (Fig. 2.8b). Samples from LLC^{AR+} were dominated by population 3 (Mono_*S100a9*), a population with a high expression of *S100a9*, *S100a8* and *Satb1*. The most abundant cluster in irradiated LLC^{AR+} samples was population 1, characterized by high expression of genes from the S100 family (*S100a4*, *S100a6*, *S100a10*), immunosuppressive (*Chil3*, *Hp*, *Gas7*, *Anxa1*) and tissue-repair genes (*Fn1*, *F13a1*, *Thbs1*, *Vcan*), as well as the downregulation of genes involved in monocyte to macrophage differentiation (*Pparg*, *Itgax*, *Ly75*, *Hsp90ab1*) (Fig. 2.8b). MNPs from the lungs of AR- tumor-bearing mice showed a more diverse spread between all clusters and a comparatively larger fraction of clusters 4 (Alv_macro_*Mrc1*), 10 (Int_Macro_*C1qa*), and 13 (Class_DC_1_*H2-Eb1*) (Fig. 2.8b). Clusters 10 and 13 expressed pro-inflammatory (*C1qa*, *C1qb*, *C3ar1*), antigen processing/presentation (*H2-Eb1*, *H2-Ab1*, *H2-*

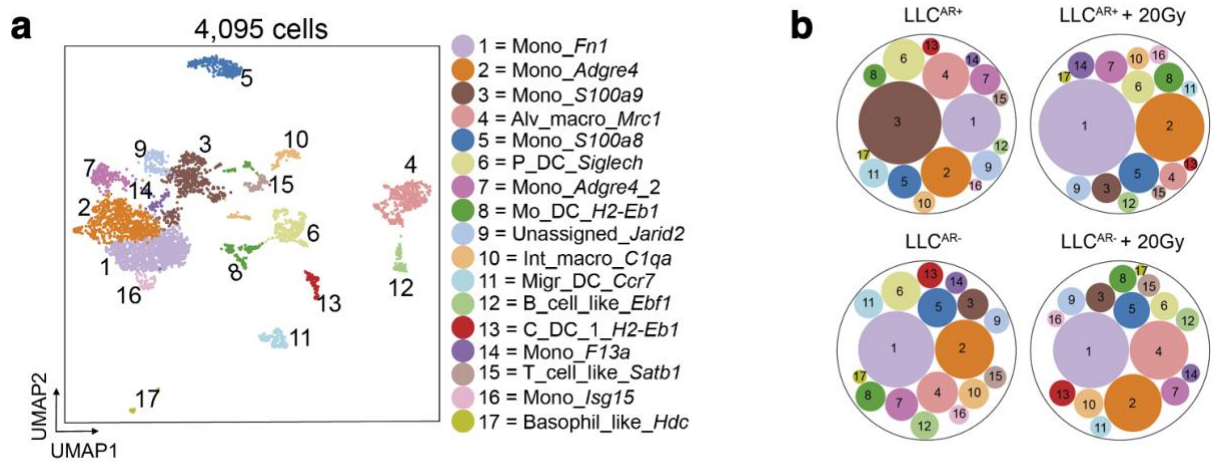

Figure 2.8. EGFR-high mononuclear phagocytes (MNPs) are re-clustered into 17 subclusters.

a) UMAP clustering of 4,095 MNPs (Migratory DC, Inflammatory Monocyre, Conventional Type 1 DC, Alveolar Macrophages, Plasmacytoid DC) in lungs of LLC^{AR+} and LLC^{AR-} tumor-bearing mice ± 20 Gy, identification of 17 subclusters(Mono = monocyte, Alv_macro = alveolar macrophage, P_DC = plasmacytoid DC, Mo_DC = monocyte-derived DC, Int_macro = interstitial macrophage, Migr_DC = migratory DC, C_DC_1 = classical type 1 DC) annotated manually and labeled with the top most differentially upregulated gene in subcluster. Differential genes were determined by Mann-Whitney test with q-value cutoff < 0.05 **b,** Dot plot representing the number of cells of the 17 MNP subclusters in different experimental conditions.

Aa), and phagocytosis-related (*Cd74*, *Cd63*, *Fcgr2b*) genes. Similarly, Cluster 4 (Alv_macro_*Mrc1*) upregulated genes involved in phagocytosis (*Mrc1*, *Mertk*, *Ctsd).*

Spectral flow cytometry results from Piffko et al. had highlighted the increased presence of inflammatory monocytes in the lung, specifically in the LLC^{AR+} conditions[70]. We confirmed this finding in scRNA-seq as well. Local RT of LLC^{AR+} flank tumors significantly increased the frequency of inflammatory monocytes in the lung, which was not recapitulated in the LLC^{AR-} condition (Fig. 2.9a-b). Comparing the inflammatory monocyte compartment from LLC^{AR+} + 20 Gy to all other conditions, known immunosuppressive genes (*Lrg1*, *Retnlg*, *Eno1*) and S100 proteins (*S100a9*, *S100a8*) were upregulated and genes involved in inflammation (*Cd36*, *Dock2*, *Aff3*), antigen processing and presentation (*H2-K1*, *H2-Aa*, *Cd83*), as well as phagolysosomal processes (*Dock10*, *Picalm*, *Satb1*) were downregulated (Fig 2.10a). Further supporting this observation, we found that inflammatory monocytes from LLC^{AR+} + 20 Gy uniformly expressed these genes from a known myeloid-derived suppressor cell (MDSC) immunosuppression[86] signature at higher levels (Fig 2.10b). We then asked whether monocytic suppression correlates with T cell dysfunction. We discovered that T cells displayed a terminal exhaustion[87] transcriptomic phenotype in the LLC^{AR+} + 20 Gy condition (Figure 2.10c).

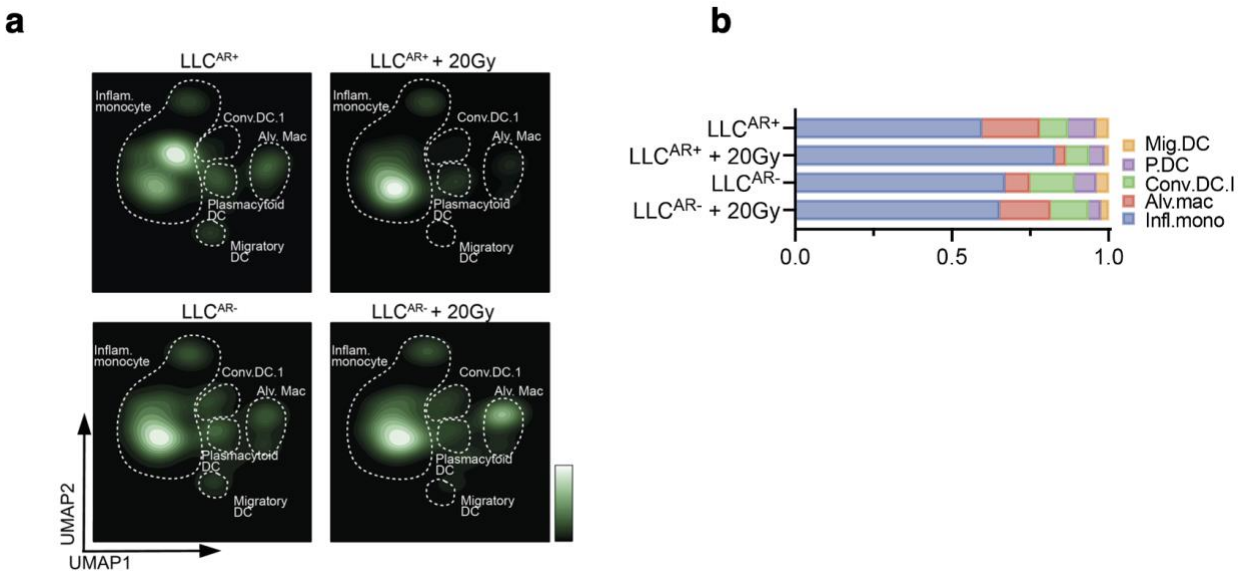

Figure 2.9. There is a specific increase in inflammatory monocytes in the LLC^{AR+} + 20 Gy condition.

a, Density plot of 4,095 MNPs in lungs of LLC^{AR+} and LLC^{AR-} tumor-bearing mice ± 20 Gy. **b,** Quantification of 5 clusters of MNPs (outlined in Fig. 4a) across conditions. P_DC = plasmacytoid DC.

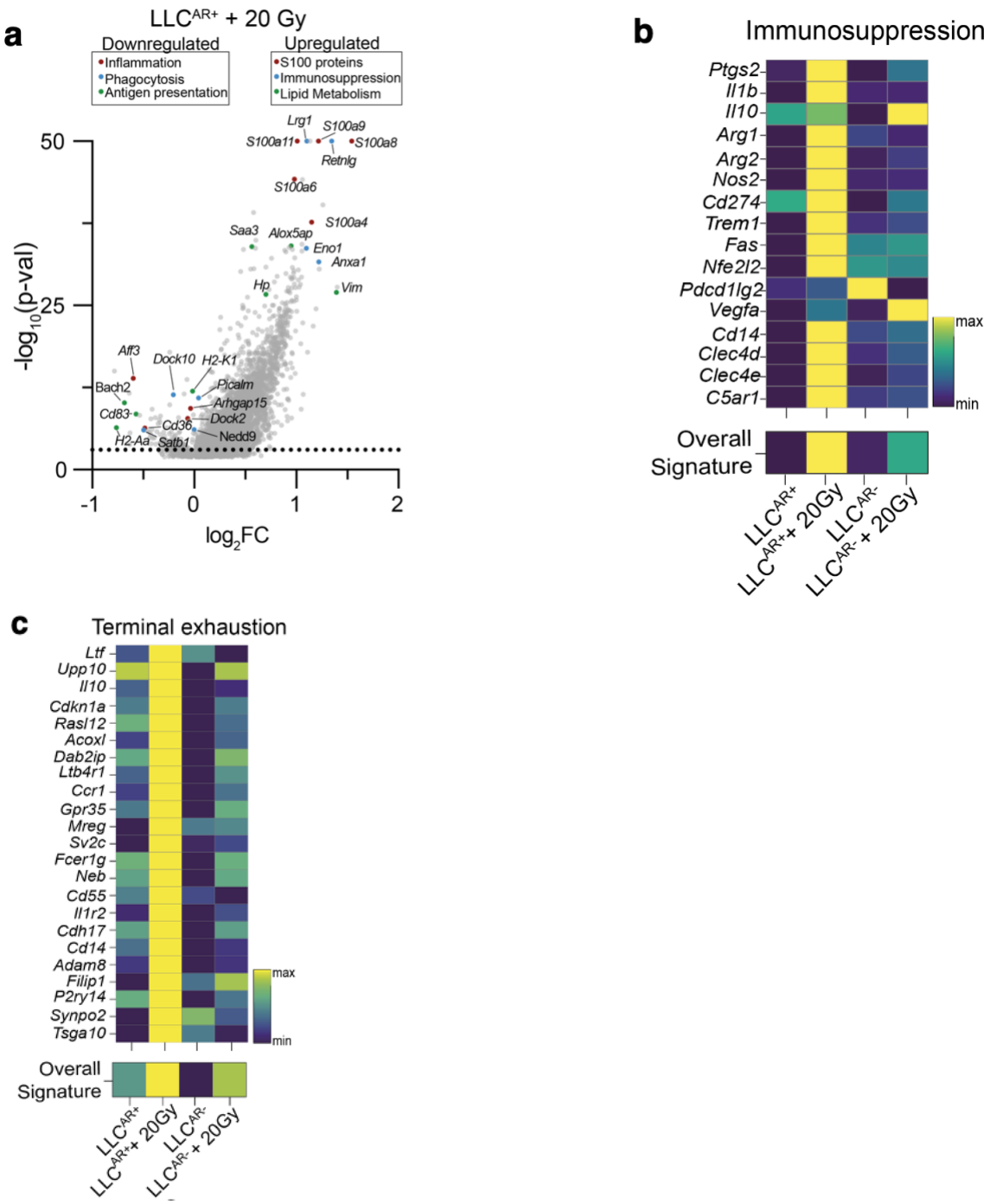

Figure 2.10. The inflammatory monocytes in LLC^{AR+} + 20 Gy condition are immunosuppressive.

a, Volcano plot of up- and downregulated genes in population 1 (Mono_*Fn1*). Genes are colored by function. P-values of 0 are depicted at 1^{-50} (log10). Wilcoxon rank-sum test. **b,** Row-normalized heatmap of individual genes and overall gene signature of immunosuppressive genes in inflammatory monocyte clusters displayed in 2.7a. **c)**, Row-normalized heatmap of CD8 T cells with individual genes from a terminal exhaustion gene signature.

AREG alters differentiation trajectories of inflammatory monocytes in the lung metastasis microenvironment

The phagocytic activity of mononuclear phagocytes (MNPs) is crucial for regulating tumor control within the lung-tumor microenvironment, affecting both tissue-resident and monocyte-derived cells[88]. It has been demonstrated that AREG promotes resistance to apoptosis and diminishes cell death induced by phagocytosis[89], however, the impact of tumor cell-derived AREG on the phagocytic capabilities of MNPs remains uncertain.

To improve our understanding of the effects of tumor-derived AREG on functionally relevant differentiation trajectories of inflammatory monocyte subclusters in the lung metastatic microenvironment, we decided to perform trajectory analysis. One approach relies on diffusion maps and pseudotime analysis. Briefly, diffusion mapping is a dimensionality reduction technique which attempts to use the diffusion process to identify nonlinear relationships within scRNA-seq data enabling the identification of continuous biological processes like cellular development and trajectories[91]. Pseudotime analysis is a complementary analysis which demonstrates temporal progression along the trajectories visualized on diffusion maps[91]. In order to leverage these techniques, we first identified seven subclusters of inflammatory monocytes (n=2600) that were transcriptionally similar for downstream analysis (Fig 2.11a). We inferred their development by computing a diffusion map and ordering subclusters along a pseudotime axis (Fig. 12.11b-c). We next assessed the pseudotime distributions of the seven populations to identify a suitable starting population(2.11d). Differential gene expression analysis on these populations was performed (Figure 2.12). Pop1_*Fn1* was chosen as a starting population due to its low pseudotime distribution and downregulation of known monocyte-differentiation genes (*Hsp90ab1*, *Ly75*, *Itgax*, *Pparg*).

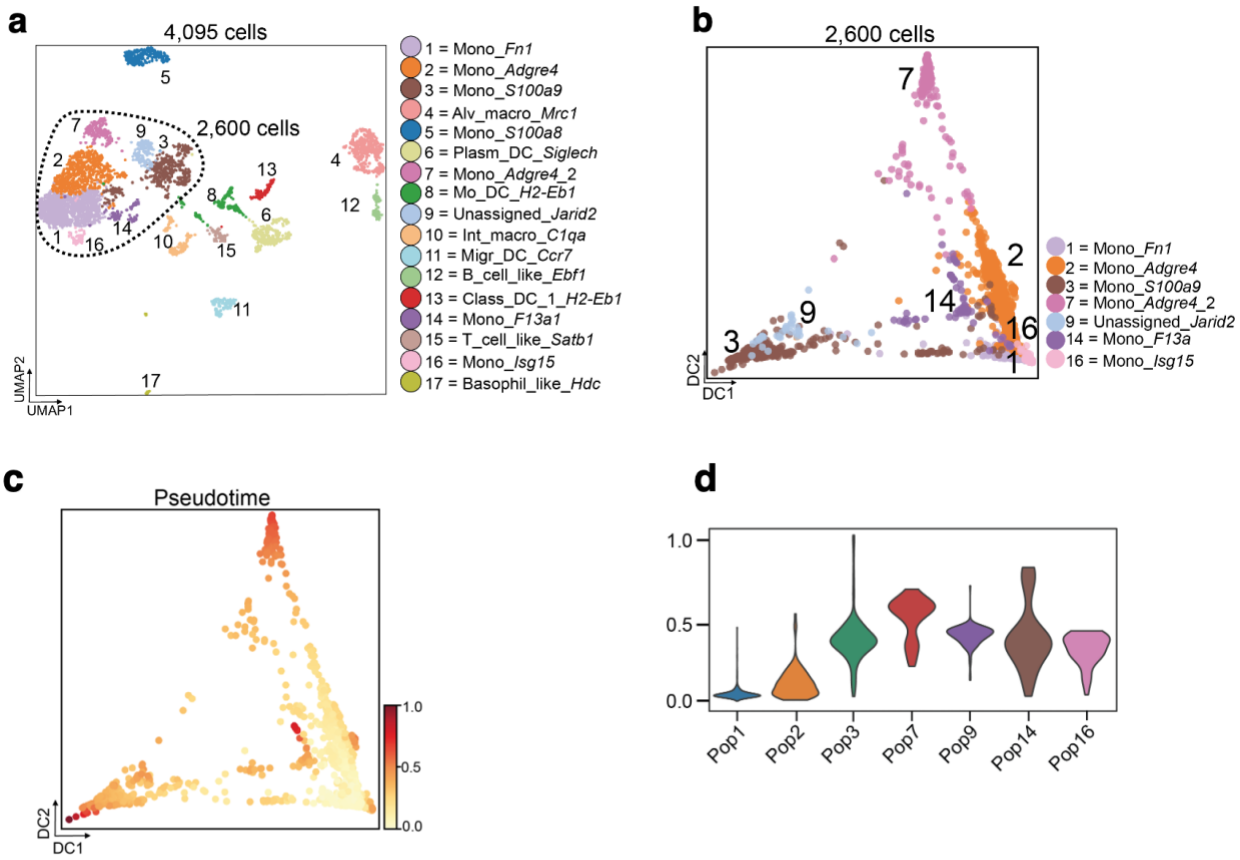

Figure 2.11. Pseudotime is computed for a subset of transcriptionally similar inflammatory monocytes which have different pseudotime distributions.

a, UMAP clustering of 4,095 MNPs identified by scRNA-seq in the lungs of LLC^{AR+} and LLC^{AR-} tumor-bearing mice ± 20 Gy, *n* = 4 mice per group. Clusters used for pseudotime analysis outlined - 2,600 cells. Cluster color and name assignment consistent with Figure 2.10a. **b)** Diffusion map projection of 2,600 outlined inflammatory monocytes computed using Pegasus package. Colors correspond to clusters identified in Figure 2.10a. **c)** Diffusion map colored by the pseudotime variable computed using the Pegasus package. **d)** Violin plot representing the mean pseudotime of cells in each highlighted population corresponding to cell clusters found in 2.13.b.

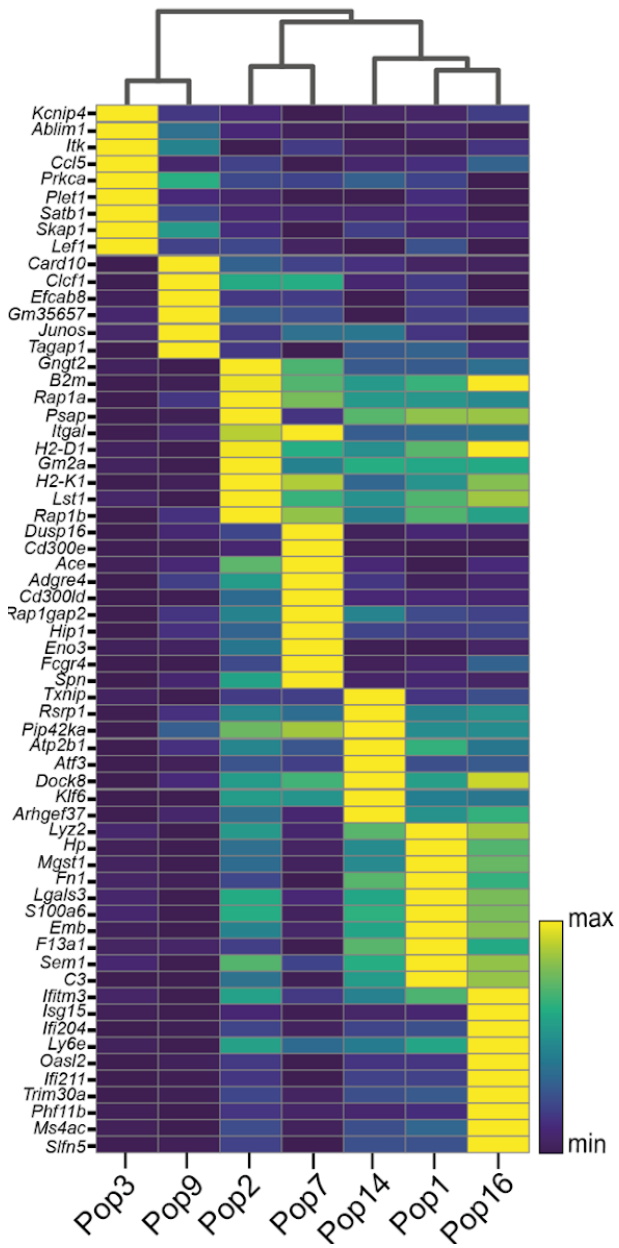

Figure 2.12. Heatmap of differential genes in IM populations selected for trajectory analysis.

Row-normalized heatmap showing expression of top eight marker genes for seven inflammatory monocyte subclusters that were used for trajectory analysis. DEGs are determined by Wilcoxon rank-sum test and determined in a one-vs-all fashion.

Trajectory inference using Slingshot with population 1 as a starting population predicted two main trajectories (Fig. 2.13a): trajectory 1 developed through population 2 to 7. An alternative trajectory, trajectory 2, instead moved through populations 14 and 9, ending in population 3. These major trajectories were confirmed using an orthogonal method – using dynamical models to infer trajectories using RNA velocity (Figure 2.13b). Trajectory 1 involved cells from all four conditions in equal parts, while trajectory 2 was dominated by cells from AR^+ conditions (Figure 2.14b).

The comparison of pseudotime states in the AR^+ conditions showed that a considerable fraction of monocytes from LLC^{AR+} + 20 Gy condition were arrested at an earlier pseudotime state and less likely to progress to later stages (Fig. 2.14a). Corresponding to the differentiation arrest, we found that the largest fraction of cells in Pop1_*Fn1* – the population with the lowest pseudotime – were derived from LLC^{AR+} + 20 Gy (Fig. 2.14b). In contrast, cells from irradiated and unirradiated AR^- conditions were more likely to differentiate, measured by progression along pseudotime and local tumor RT did not affect their pseudotime density (2.14a).

Having established that AREG either causes arrest in population 1 or diversion along Trajectory 2, we next explored these two fates. Pop1, previously identified as a starting population, upregulated immunosuppressive proteins (S100 proteins, *Vcan, Fn1*) as well as proteins involved in tissue repair (Fig 2.15a). At the same time, genes involved in inflammation, antigen presentation and differentiation were downregulated in Pop 1 (Fig 2.15a). Notably, Pop1_*Fn1* also showed a higher signaling through EGFR pathways compared with other populations (Fig 2.15b). Thus, we conclude that the AREG-driven arrested population 1 signals through EGFR and has a pro-tumor transcriptional landscape.

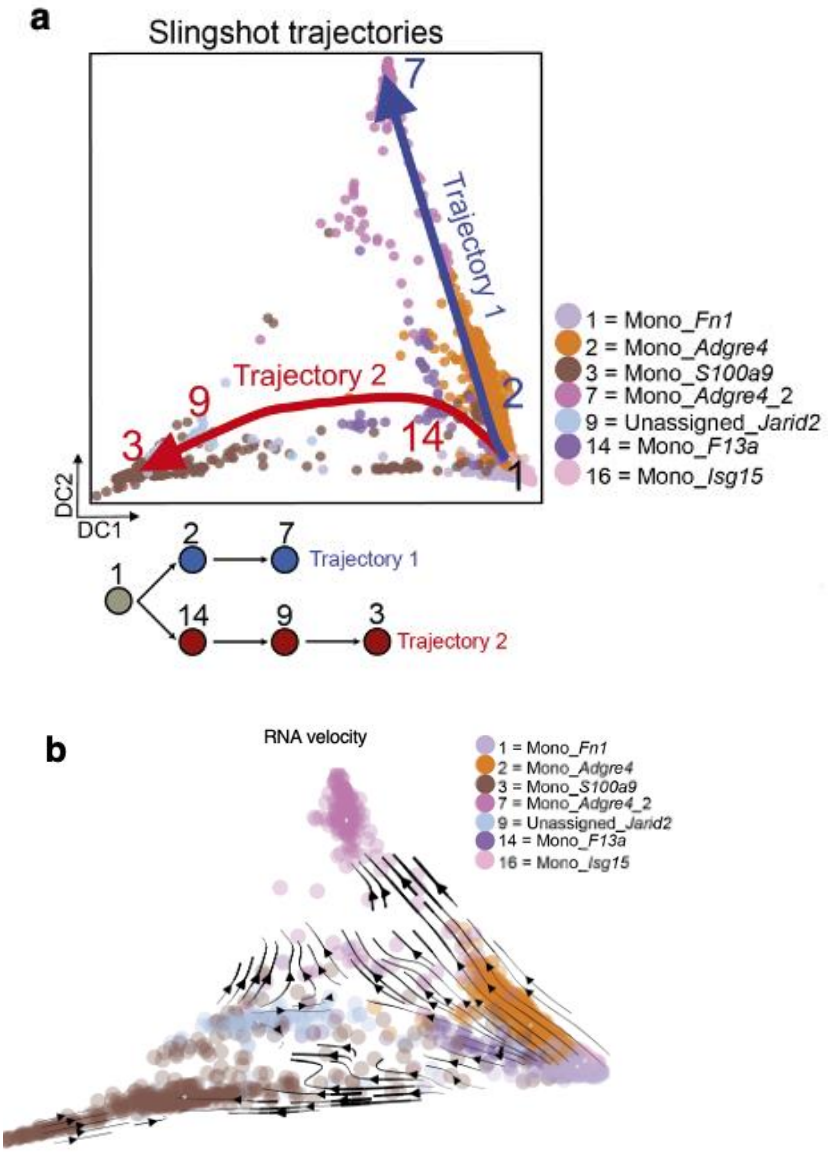

Figure 2.13. Two major monocyte trajectories are identified with AREG confirmed using two orthogonal methods.

a, Trajectory inference of inflammatory monocytes using Slingshot and schematic depiction of two main trajectories. **b,** Proportion of trajectories 1 and 2 in the four experimental conditions. **c,** Violin plot of mean pseudotime in all four experimental conditions. **d,** Proportion of population 1 in the four experimental conditions. **e,** Trajectory inference of the same inflammatory monocytes as in 2.14a using the RNA velocity and the ScVelo package.

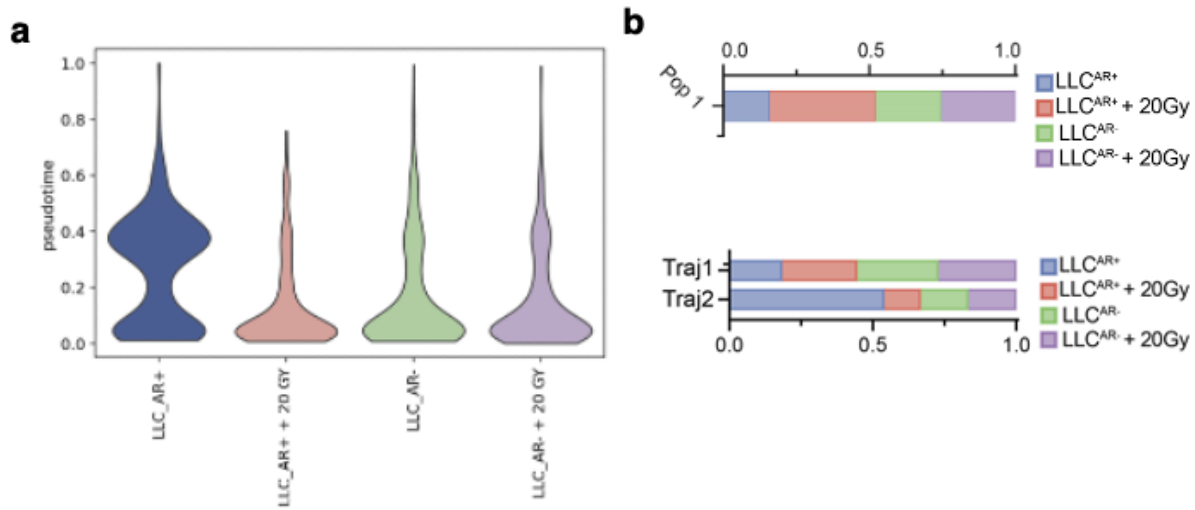

Figure 2.14. AREG causes arrest in population 1 or diversion along Trajectory 2.

a, Violin plot of pseudotime distribution of inflammatory monocytes from each biological condition. **b,** Bar plots representing the biological condition distribution of Pop 1 inflammatory monocytes as well as Trajectory 1 and 2 inflammatory monocytes.

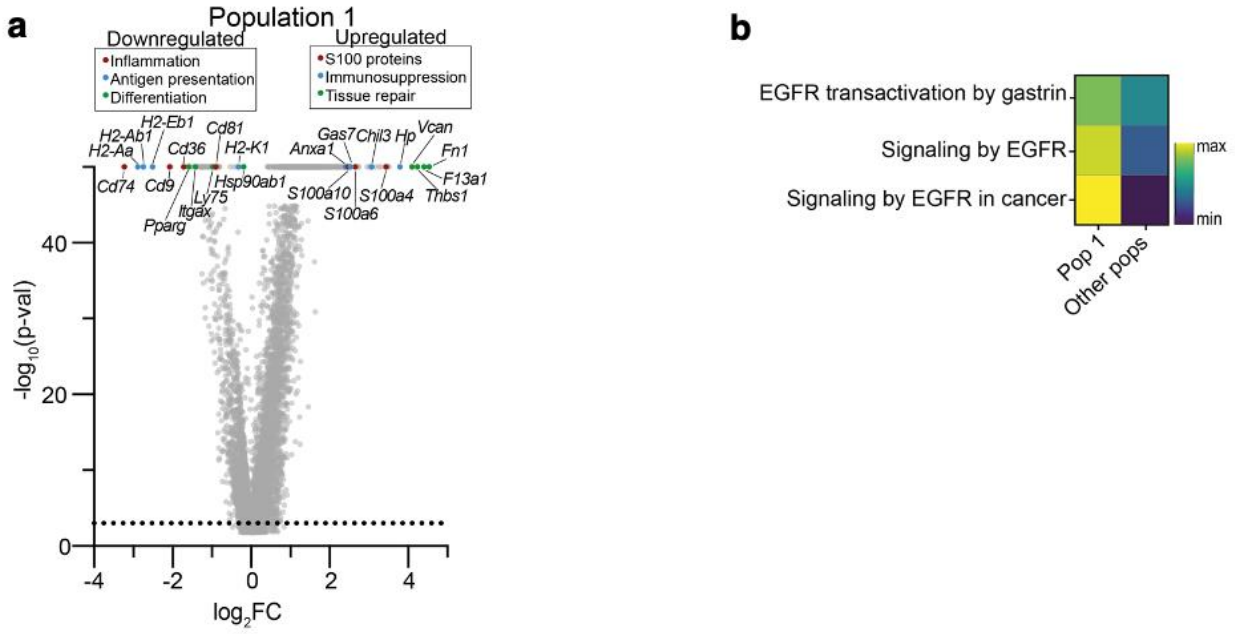

Figure 2.15. AREG-driven arrested population has a pro-tumor transcriptional landscape and upregulated EGFR gene signatures.

a, Volcano plot of up- and downregulated genes in population 1 (Mono_*Fn1*). Genes are colored by function. P-values of 0 are depicted at 1^{-50} (log10). Wilcoxon rank-sum test. **b,** Row-normalized heatmap showing mean gene expression scores of EGFR signaling related Reactome pathways in population 1 vs. all other populations.

Comparison of differentially-expressed gene patterns between the two main trajectories showed that cells following trajectory 1 increasingly express canonical markers of pathways associated with macrophage function. Specifically, trajectory 1 has higher expression of genes involved in phagocytosis, cell killing, as well as antigen presentation and processing (Fig 2.16a-b). While Trajectory 2 showed a global downregulation of gene expression, it significantly upregulates genes linked to immunosuppressive qualities of MDSCs (*S100a8, S100a9, Sell, Il18rap*)(Fig 2.16c). We concluded that the main route of functional monocyte-differentiation occurred along trajectory 1. By contrast, under AR^{+} conditions, cells were more likely to be redirected along an alternative trajectory (trajectory 2), with significantly decreased anti-tumor functions, increased immunosuppression and transcriptional downregulation.

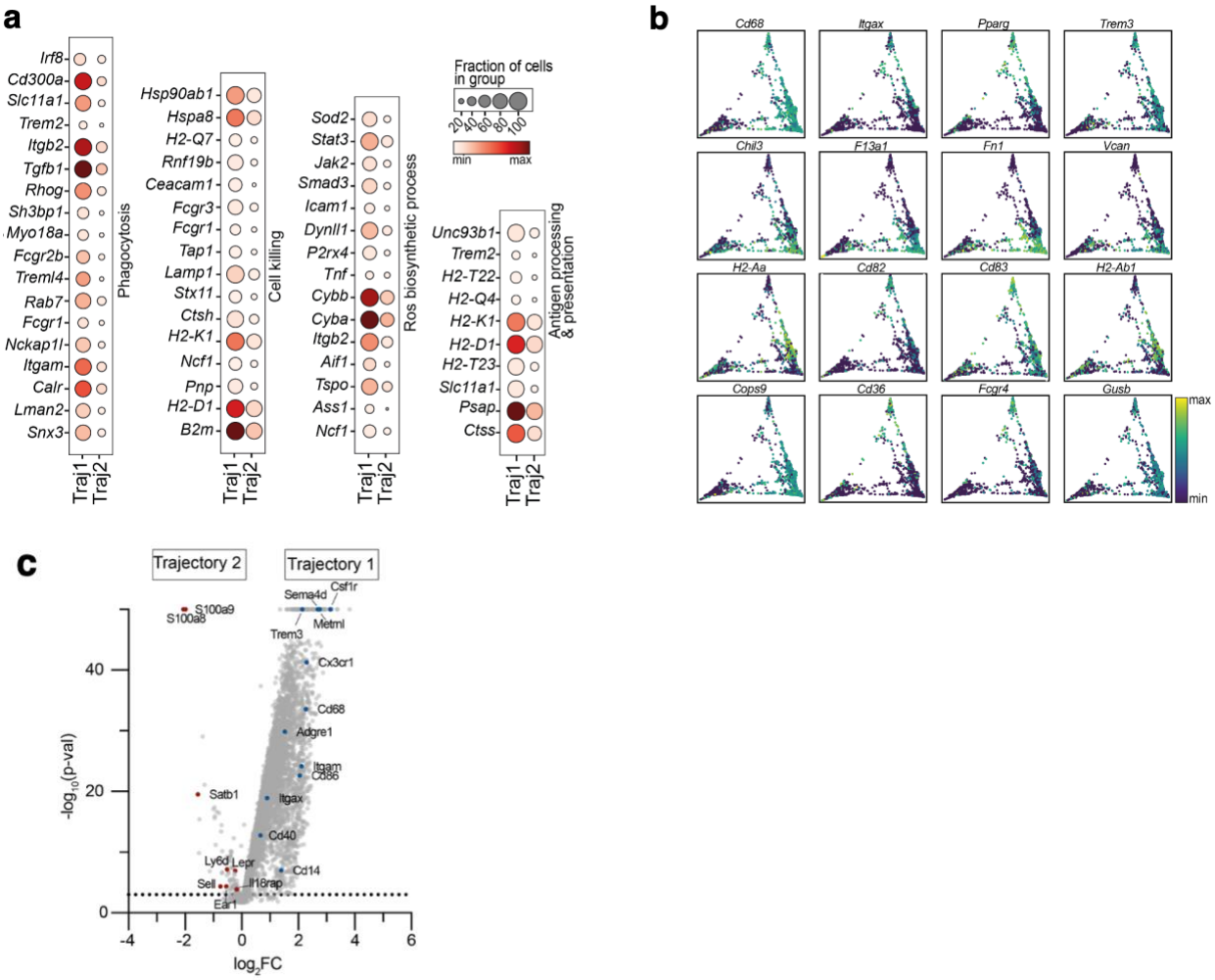

Figure 2.16. AREG-driven Trajectory 2 has lower expression of phagocytosis, cell killing and antigen presentation genes and increased expression of immunosuppressive signatures.

a, Bubble plots depicting proportion of total cells in Traj 1 and 2 expressing selected genes related to phagocytosis, cell killing, reactive oxygen species (ROS) biosynthetic process, or antigen processing and presentation of trajectory (Traj) 1 and 2. Signatures from Wang et al. **b,** Expression of select genes from each pathway in 2.16.a across diffusion map projection. **c,** Volcano plot depicting differentially expressed genes between Trajectory 2 and Trajectory 1.

DISCUSSION

In this study, we employed scRNA-seq to demonstrate that RT-induced AREG reprograms EGFR-expressing myeloid cells at the transcriptional level leading to metastatic growth. Importantly, RT-induced AREG upregulation and an increase of $pEGFR^{+}$ monocytes were previously identified in the PBMCs of cancer patients following SBRT by Piffko et al., and our results validated their observations[70].

Although we cannot exclude the role of immune cells other than MNPs, our results indicate that MNPs were the most transcriptionally active populations and demonstrated a high degree of cell-cell interaction. Piffko et al. experimentally demonstrated that depletion of T cells did not affect the metastasis phenotype while depletion of MNPs did[70]. Combined with those results, and our transcriptomic approaches, we delineated the central role of the tumor-AREG-MNP-EGFR signaling loop (Fig 2.19) that drives distant metastasis proliferation following RT. Our data shows that radiation-induced distal tumor-derived AREG alters the differentiation of infiltrating monocytes in the lung and can potentially enable successful metastatic proliferation by promoting monocyte arrest at an immature suppressive state as well as development along a tumor-tolerogenic trajectory.

There are important caveats to our work. We did not directly examine tumor cells in the scRNA-seq analysis, focusing solely on the immune compartment. While Piffko et al. demonstrated that AREG did not affect primary tumor-intrinsic growth or colonization *in vitro*, we cannot exclude the possibility that the observed effects in our scRNA-seq analysis did not involve metastatic tumor-driven effects. Furthermore, the limitations of our transcriptomic

approach mean that we cannot capture potential post-transcriptional regulatory mechanisms or other modes of regulation that may also play a role in these processes.

In summary, we use scRNA-seq to add additional mechanistic insights to the work by Piffko et.al demonstrating that radiotherapy-induced growth factors can drive distant metastasis growth in SBRT-treated patients and murine models, resulting in shortened survival and adverse outcomes[70]. Taken together, these findings indicate that factors promoting distant growth after RT may be targetable. EGFR signaling has previously been linked to the immunosuppressive[92] and tumor-promoting roles[93] of MNPs, and recently EGFR-targeted therapy has been suggested to mitigate immunosuppression driven by Treg and macrophages in the TME of inflammatory breast cancer[94]. Additionally, AREG has been proposed as a regulator of TAM accumulation in a breast cancer model[95]. All these results propose a paradigm shift in RT use for patients with locally advanced and metastatic tumors, suggesting that growth factors promoting distant metastasis should be monitored during therapy. Personalized approaches to RT may be possible where modulation of these growth factors could dramatically improve patient outcomes.

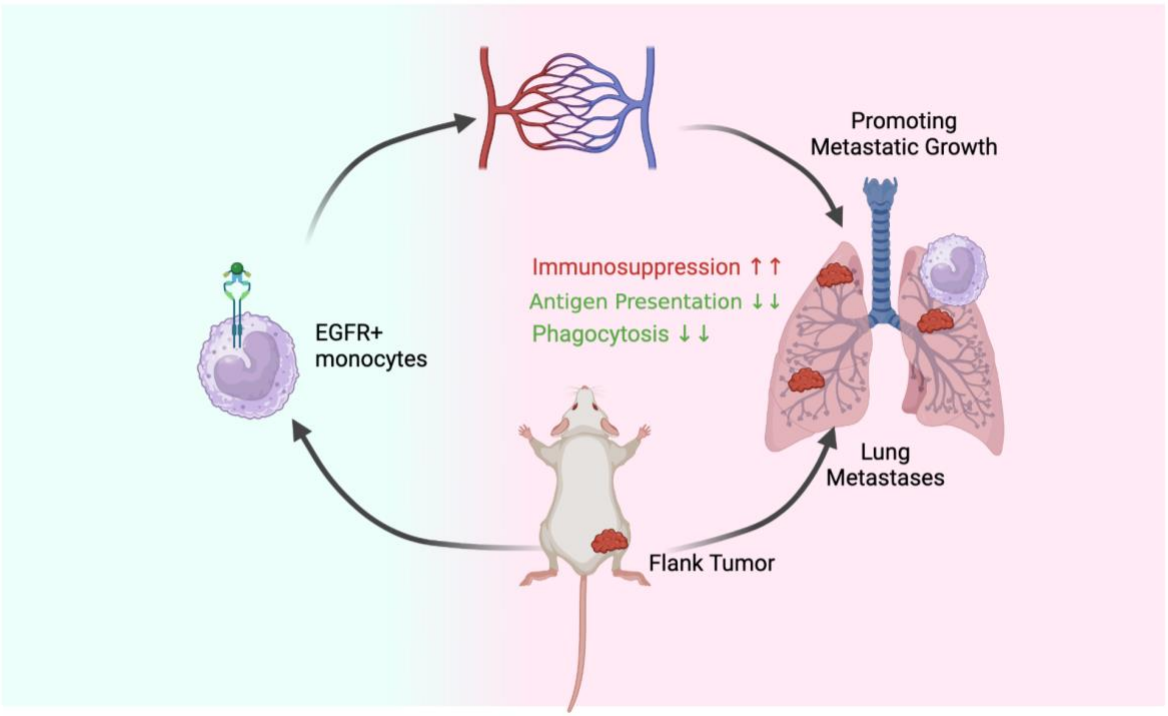

Figure 2.17. Summary Model.

We found that radiation induction of flank tumors lead to an upregulation in circulating monocytes which in turn become immunosuppressed, deficient in antigen presentation and phagocytosis and support metastatic growth of distant tumors.

REFERENCES

1. Mandel P, Métais P. Les acides nucléiques du plasma sanguin chez l' homme. *C R Seances Soc Biol Fil*. **142**, 241–3 (1948).

2. Jahr S, Hentze H, Englisch S, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res*. **61**,1659–65 (2001).

3. Choi JJ, Reich CF, Pisetsky DS. The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells. *Immunology*. **115**, 55–62 (2005).

4. Chaudhuri AA, Binkley MS, Osmundson EC, et al. Predicting radiotherapy responses and treatment outcomes through analysis of circulating tumor DNA. *Semin Radiat Oncol*. **25**, 305–12 (2015)

5. Lui YYN, Chik KW, Chiu RWK, et al. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem*. **48**,421–7 (2002).

6. Snyder MW, Kircher M, Hill AJ, et al. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell*. **164**, 57–68 (2016).

7. Chang CPY, Chia RH, Wu TL, et al. Elevated cell-free serum DNA detected in patients with myocardial infarction. *Clin Chim Acta*. **327**, 95–101 (2003).

8. Lehmann-Werman R, Neiman D, Zemmour H, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci USA*. **113**, E1826–34 (2016).

9. Margraf S, Lögters T, Reipen J, et al. Neutrophil-derived circulating free DNA (CF-DNA/NETs): a potential prognostic marker for posttraumatic development of inflammatory second hit and sepsis. *Shock*. **30**, 352–8 (2008).

10. Stroun M, Anker P, Lyautey J, et al. Isolation and characterization of DNA from the plasma of cancer patients. *Eur J Cancer Clin Oncol*. **23**,707–12 (1987).

11. Corcoran RB, Chabner BA. Application of cell-free DNA analysis to cancer treatment. *N Engl J Med*. **379**, 1754–65 (2018).

12. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat Rev Cancer*. **17**,223–38 (2017).

13. Heitzer E, Haque IS, Roberts CES, et al. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet*. **20(2)**,71–88 (2019).

14. Dudley JC, Schroers-Martin J, Lazzareschi DV, et al. Detection and surveillance of bladder cancer using urine tumor DNA. *Cancer Discov*. **9**,4 (2019).

15. Chaudhuri AA, Chabon JJ, Lovejoy AF, et al. Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov*. **7**,1394–403. (2017)

16. Abbosh C, Birkbak NJ, Wilson GA, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature*. **545**, 446–51 (2017).

17. Newman AM, Bratman SV, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med*. **20**, 548–54 (2014).

18. Khan KH, Cunningham D, Werner B, et al. Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the PROSPECT-C phase II colorectal cancer clinical trial. *Cancer Discov*. **8**,1270–85 (2018).

19. Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol*. **36**,2845–53 (2018).

20. Hu Y, Ulrich BC, Supplee J, et al. False-positive plasma genotyping due to clonal hematopoiesis. *Clin Cancer Res*. **24**,4437–43 (2018).

21. Swanton C, Venn O, Aravanis A, et al. Prevalence of clonal hematopoiesis of indeterminate potential (CHIP) measured by an ultra-sensitive sequencing assay: exploratory analysis of the Circulating Cancer Genome Atlas (CCGA) study. *J Clin Oncol*. **36**,12003 (2018).

22. Cui, XL., Nie, J., Ku, J. et al. A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun*. **11**, 6161 (2020).

23. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).

24. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).

25. Li, W. S. et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* **27**, 1243–1257 (2017).

26. Shao J, Bernicker EH, He C, Li Z. Cell-free DNA 5-hydroxymethylcytosine as a marker for common cancer detection. *Clinical and Translational Dis.* **13**,715 (2022).

27. Sjöström M, Zhao SG, Levy S, et al. The 5-hydroxymethylcytosine landscape of prostate cancer. *Cancer Res.* **82**, 3888-3902 (2022).

28. Chiu BCH, Zhang Z, You Q, et al. Prognostic implications of 5-hydroxymethylcytosines from circulating cell-free DNA in diffuse large B-cell lymphoma. *Blood Adv.* **3**,2790-2799 (2019).

29. Cai J, Chen l, Zhang Z, et al. Genome-wide mapping of 5-hydroxymethylcytosines in circulating cell-free DNA as a non-invasive approach for early detection of hepatocellular carcinoma. *Gut.* **68**, 2195–2205 (2019).

30. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* **70**, 7–30 (2020).

31. Flaig TW, Spiess PE, Agarwal N, Bangs R, Boorjian SA, Buyyounouski MK, et al. Bladder Cancer, Version 3.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* **18**,329–54 (2020).

32. Kirkali Z, Chan T, Manoharan M, Algaba F, Busch C, Cheng L, et al. Bladder cancer: epidemiology, staging and grading, and diagnosis. *Urology*. **66**, 4–34 (2005).

33. Cumberbatch MGK, Jubber I, Black PC, Esperto F, Figueroa JD, Kamat AM, et al. Epidemiology of Bladder Cancer: A Systematic Review and Contemporary Update of Risk Factors in 2018. *Eur Urol.* **74**, 784–95 (2018).

34. Mossanen M, Krasnow RE, Zlatev DV, Tan WS, Preston MA, Trinh QD, et al. Examining the relationship between complications and perioperative mortality following radical cystectomy: a population-based analysis. *BJU Int. 2019*. **124**, 40–6 (2019).

35. Shabsigh A, Korets R, Vora KC, Brooks CM, Cronin AM, Savage C, et al. Defining early morbidity of radical cystectomy for patients with bladder cancer using a standardized reporting methodology. *Eur Urol.* **55**,164–74 (2009).

36. Petrelli F, Coinu A, Cabiddu M, Ghilardi M, Vavassori I, Barni S. Correlation of pathologic complete response with survival after neoadjuvant chemotherapy in bladder cancer treated with cystectomy: a meta-analysis. *Eur Urol.* **65**,350-357 (2014).

37. Chin RI, Chen K, Usmani A, Chua C, Harris PK, Binkley MS, Azad TD, Dudley JC, Chaudhuri AA. Detection of Solid Tumor Molecular Residual Disease (MRD) Using Circulating Tumor DNA (ctDNA). *Mol Diagn Ther*. **23**. 311-331 (2019).

38. Springer SU, Chen C-H, Rodriguez Pena MDC, et al. Non-invasive detection of urothelial cancer through the analysis of driver gene mutations and aneuploidy. *Elife*. **7**:1–27, (2018).

39. Chauhan PS, Chen K, Babbra RK, Feng W, Pejovic N, Nallicheri A, Harris PK, Dienstbach K, Atkocius A, Maguire L, Qaium F, Szymanski JJ, Baumann BC, Ding L, Cao D, Reimers MA, Kim EH, Smith ZL, Arora VK, Chaudhuri AA. Correction: Urine tumor DNA detection of minimal residual disease in muscle-invasive bladder cancer treated with curative-intent radical cystectomy: A cohort study. *PLoS Med*. **18**, e1003876 (2021).

40. Chauhan, P.S., Shiang, A., Alahi, I. *et al.* Urine cell-free DNA multi-omics to detect MRD and predict survival in bladder cancer patients. *npj Precis. Onc.* **7**, 6 (2023).

41. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics.* **14**, 2382-3 (2015).

42. Muzellec B, Teleńczuk M, Cabeli V, Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics.* **39**, btad547 (2023).

43. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics.* **39**, btac757 (2023).

44. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, Haw R, Jassal B, Matthews L, May B, Petryszak R, Ragueneau E, Rothfels K, Sevilla C, Shamovsky V, Stephan R, Tiwari K, Varusai T, Weiser J, Wright A, Wu G, Stein L, Hermjakob H, D'Eustachio P. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research*. **52**: D672-D678 (2024).

45. Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. **28**, 27-30 (2000).

46. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57-74 (2012).

47. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31,** 8-22 (2022).

48. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**,417-425 (2015).

49. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98-W102 (2017).

50. Adalsteinsson VA, Ha G. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun.* **8**,1324 (2017).

51. Fragkoulis C, Ntoumas G, Glykas I, Papadopoulos G, Stathouros G, Kostopoulou A, Choreftaki T, Ntoumas K. Expression of proto-oncogene c-Myc in patients with urinary bladder transitional cell carcinoma. *Curr Urol.* **15**,231-233 (2021).

52. Baer PC, Koch B, Geiger H. Kidney Inflammation, Injury and Regeneration. *Int J Mol Sci.* **21**,1164 (2020).

53. Cai Z, Chen J, Yu Z, Li H, Liu Z, Deng D, Liu J, Chen C, Zhang C, Ou Z, Chen M, Hu J, Zu X. BCAT2 Shapes a Noninflamed Tumor Microenvironment and Induces Resistance to Anti-PD-1/PD-L1 Immunotherapy by Negatively Regulating Proinflammatory Chemokines and Anticancer Immunity. *Adv Sci (Weinh).* **8**,e2207155 (2023).

54. Aine M, Eriksson P, Liedberg F, Sjödahl G, Höglund M. Biological determinants of bladder cancer gene expression subtypes. *Sci Rep.* **5**,10957 (2015).

55. Yu C, Hequn C, Longfei L, Long W, Zhi C, Feng Z, Jinbo C, Chao L, Xiongbing Z. GSTM1 and GSTT1 polymorphisms are associated with increased bladder cancer risk: Evidence from updated meta-analysis. *Oncotarget*. **8**,3246-3258 (2017).

56. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci.* **9**,193-9 (2012).

57. Weichselbaum, R., Liang, H., Deng, L. *et al.* Radiotherapy and immunotherapy: a beneficial liaison?. *Nat Rev Clin Oncol* **14**, 365–379 (2017).

58. Pitroda, S. P., Chmura, S. J. & Weichselbaum, R. R. Integration of radiotherapy and immunotherapy for treatment of oligometastases. *The Lancet Oncology* **20**, e434–e442 (2019).

59. Pitroda, S. P. & Weichselbaum, R. R. Integrated molecular and clinical staging defines the spectrum of metastatic cancer. *Nature Reviews Clinical Oncology* **16**, 581–588 (2019).

60. Katipally, R. R., Pitroda, S. P., Juloori, A., Chmura, S. J. & Weichselbaum, R. R. The oligometastatic spectrum in the era of improved detection and modern systemic therapy. *Nature Reviews Clinical Oncology* **19**, 585–599 (2022).

61. Fares, J., Fares, M.Y., Khachfe, H.H. *et al.* Molecular principles of metastasis: a hallmark of cancer revisited. *Sig Transduct Target Ther* **5**, 28 (2020).

62. Hellman S, Weichselbaum RR. Oligometastases. J Clin Oncol. **1**,8-10 (1995).

63. Pitroda, S. P., Chmura, S. J. & Weichselbaum, R. R. Integration of radiotherapy and immunotherapy for treatment of oligometastases. *The Lancet Oncology* **20**, e434–e442 (2019).

64. Pitroda, S. P. & Weichselbaum, R. R. Integrated molecular and clinical staging defines the spectrum of metastatic cancer. *Nature Reviews Clinical Oncology* **16**, 581–588 (2019).

65. Katipally, R. R., Pitroda, S. P., Juloori, A., Chmura, S. J. & Weichselbaum, R. R. The oligometastatic spectrum in the era of improved detection and modern systemic therapy. *Nature Reviews Clinical Oncology* **19**, 585–599 (2022).

66. Spiotto, M., Fu, Y. X. & Weichselbaum, R. R. The intersection of radiotherapy and immunotherapy: Mechanisms and clinical implications. *Sci. Immunol.* **1**, (2016).

67. Jiménez-Cortegana, C., Galassi, C., Klapp, V., Gabrilovich, D. I. & Galluzzi, L. Myeloid-Derived Suppressor Cells and Radiotherapy. *Cancer Immunology Research* **10**, 545–557 (2022).

68. Wang, L. *et al.* RNA m6A methylation and MDSCs: Roles and therapeutic implications for radiotherapy. *Med* **4**, 863–874 (2023).

69. Pointer, K. B., Pitroda, S. P. & Weichselbaum, R. R. Radiotherapy and immunotherapy: open questions and future strategies. *Trends in Cancer* **8**, 9–20 (2022).

70. Piffko A., Yang K., Panda A., and Weichselbaum R.R. "Radiation-induced amphiregulin drives tumor metastasis." Unpublished manuscript, last modified Jan 3rd 2024. Microsoft Word file.

71. Luke, J. J. *et al.* Safety and clinical activity of pembrolizumab and multisite stereotactic body radiotherapy in patients with advanced solid tumors. *J. Clin. Oncol.* **36**, 1611–1618 (2018).

72. Luke, J. J. *et al.* Improved survival associated with local tumor response following multisite radiotherapy and pembrolizumab: Secondary analysis of a phase i trial. *Clin. Cancer Res.* **26**, 6437–6444 (2020).

73. Berasain, C. & Avila, M. A. Amphiregulin. *Semin. Cell Dev. Biol.* **28**, 31–41 (2014).

74. Miyawaki, M. *et al.* Enhanced phosphorylation of the epidermal growth factor receptor at the site of tyrosine 992 in esophageal carcinomas. *APMIS* **116**, 1097–1106 (2008).

75. Bestvina, C. M. *et al.* A Phase 1 Trial of Concurrent or Sequential Ipilimumab, Nivolumab, and Stereotactic Body Radiotherapy in Patients With Stage IV NSCLC Study. *J. Thorac. Oncol.* **17**, 130–140 (2022).

76. Li B, Gould J, Yang Y, Sarkizova S, Tabaka M, Ashenberg O, Rosen Y, Slyper M, Kowalczyk MS, Villani AC, Tickle T, Hacohen N, Rozenblatt-Rosen O, Regev A. Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. Nat Methods. **17**, 793-798 (2020).

77. Wolf, F., Angerer, P. & Theis, F. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).

78. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. *Cell.* **184**,3573-3587.e29 (2021).

79. Korsunsky, I., Millard, N., Fan, J. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).

80. Street, K., Risso, D., Fletcher, R. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

81. La Manno, G., Soldatov, R., Zeisel, A. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

82. Bergen, V., Lange, M., Peidli, S. *et al.* Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408–1414 (2020)

83. Swirski, F. K., Hilgendorf, I. & Robbins, C. S. From proliferation to proliferation: Monocyte lineage comes full circle. *Seminars in Immunopathology* **36**, 137–148 (2014).

84. Aegerter, H., Lambrecht, B. N. & Jakubzick, C. V. Biology of lung macrophages in health and disease. *Immunity* **55**, 1564–1580 (2022).

85. Mucciolo G, Araos Henríquez J, Jihad M, Pinto Teles S, Manansala JS, Li W, Ashworth S, Lloyd EG, Cheng PSW, Luo W, Anand A, Sawle A, Piskorz A, Biffi G. EGFR-activated myofibroblasts promote metastasis of pancreatic cancer. *Cancer Cell.* **42**, 101-118.e11 (2024).

86. Hegde, S., Leader, A. M. & Merad, M. MDSC: Markers, development, states, and unaddressed complexity. *Immunity* **54**, 875–884 (2021).

87. Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).

88. Noy, R. & Pollard, J. W. Tumor-Associated Macrophages: From Mechanisms to Therapy. *Immunity* **41**, 49–61 (2014).

89. Zaiss, D. M. W., Gause, W. C., Osborne, L. C. & Artis, D. Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. *Immunity* **42**, 216–226 (2015).

90. Huang C, Wang X, Wang Y, Feng Y, Wang X, Chen S, Yan P, Liao J, Zhang Q, Mao C, Li Y, Wang L, Wang X, Yi W, Cai W, Chen S, Hong N, He W, Chen J, Jin W. Sirpα on tumor-associated myeloid cells restrains antitumor immunity in colorectal cancer independent of its interaction with CD47. *Nat Cancer.* **5**, 500-516 (2024).

91. Haghverdi L, Ludwig LS. Single-cell multi-omics and lineage tracing to dissect cell fate decision-making. *Stem Cell Reports.* **18**,13-25 (2023).

92. Zhang, W. *et al.* Polarization of macrophages in the tumor microenvironment is influenced by EGFR signaling within colon cancer cells. *Oncotarget* **7**, 75366–75378 (2016).

93. Lanaya, H. *et al.* EGFR has a tumour-promoting role in liver macrophages during hepatocellular carcinoma formation. *Nat. Cell Biol.* **16**, 972–981 (2014).

94. Wang, X. *et al.* EGFR is a master switch between immunosuppressive and immunoactive tumor microenvironment in inflammatory breast cancer. *Sci. Adv.* **8**, (2022).

95. Nickerson, N. K., Mill, C. P., Wu, H. J., Riese, D. J. & Foley, J. Autocrine-Derived Epidermal Growth Factor Receptor Ligands Contribute to Recruitment of Tumor-Associated Macrophage and Growth of Basal Breast Cancer Cells In Vivo. *Oncol. Res.* **20**, 303 (2013).

96. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc.* **15**, 1484-1506 (2020).