Myeloid–T-Cell Communication States and BCG Response in Non-Muscle-Invasive Bladder Cancer

Fadia Mohammed M Ashmawi
f.m.ashmawi@gmail.com

Mentor: **Dr Hugo Tavares**
hm533@cam.ac.uk
Department of Genetics, University of Cambridge

KAUST Academy — Bioinformatics Summer Programme
21 August 2026

Code: github.com/fadia-ashmawi/BCG-communication-states
Project page: f-ash.github.io/bcg-nmibc-website_final

Abstract

Intravesical Bacillus Calmette–Guérin (BCG) is the standard immunotherapy for intermediate- and high-risk non-muscle-invasive bladder cancer (NMIBC), yet roughly 30–50% of high-risk patients fail to achieve a durable response and no validated pre-treatment biomarker exists. Because the mechanism of BCG is fundamentally one of antigen presentation and T-cell priming, we asked whether the communication between myeloid cells and T cells — rather than their abundance — distinguishes responders from non-responders. We reanalysed GSE322495, a CD45-enriched BD Rhapsody single-cell RNA-sequencing dataset of 21 libraries from 19 NMIBC patients, through a fully reproducible pipeline: separation of 115 AbSeq antibody-derived features that the deposit had mixed into the RNA assay, adaptive median-absolute-deviation quality control grouped by library and compartment, dual-assay normalisation, Harmony integration, graph-based clustering into 20 annotated identities (9 myeloid, 4 T-cell), pseudobulk DESeq2 differential expression, compositional differential-abundance testing, and ligand–receptor communication inference with CellChat v2 restricted to tissue and to myeloid senders and T-cell receivers. No cell-type proportion survived Benjamini–Hochberg correction across twelve comparisons, under either percentage-scale or centred-log-ratio analysis. The composition of predicted myeloid→T-cell signalling nevertheless differed sharply between arms: the MHC-II:MHC-I ratio was 0.33 in BCG-naive, 2.97 in BCG-responsive and 0.68 in BCG-unresponsive tissue, a pattern robust to a sensitivity analysis removing CD4-receptor edges on CD8 T cells (2.65 and 0.64). Signal magnitude did not track sender abundance. These results position myeloid–T-cell communication state as a hypothesis-generating biological feature of BCG response that is invisible to conventional composition analysis, and requiring independent-cohort and experimental validation.

Contents

1 Introduction 3
1.1 Clinical background 3
1.2 Motivation: communication rather than composition 3
1.3 Contributions 4
1.4 Report structure 4
2 Related Work 5
2.1 Single-cell characterisation of the BCG-treated bladder 5
2.2 Biomarkers of BCG response 5
2.3 Inference of cell–cell communication 5
2.4 Statistical practice in single-cell analysis 6
3 Methodology 7
3.1 Dataset 7
3.2 Preprocessing: separation of antibody-derived features 7
3.3 Quality control 8
3.4 Normalisation and a memory-constrained implementation 8
3.5 Integration, dimensionality reduction and clustering 9
3.6 Cell-type annotation, and the separation of identity from state 10
3.7 Differential expression 11
3.8 Differential abundance 11
3.9 Cell–cell communication inference 11
3.10 Sensitivity analysis 12
3.11 Validation of published claims 12
3.12 Software and reproducibility 12
4 Results 13
4.1 Atlas and annotation 13
4.2 Cellular composition does not separate the response arms 13
4.3 Predicted communication networks differ in structure 13
4.4 Pathway composition and the MHC-II:MHC-I balance 14
4.5 The pattern survives a sensitivity analysis 15
4.6 Signal composition is not explained by sender abundance 15
4.7 Macrophages dominate CD8 T-cell input in every arm 16
4.8 Systematic validation of published claims 16
4.9 Dissemination 17
5 Discussion 18
6 Limitations and Future Work 20
6.1 Limitations 20

6.2 Future work 20
7 Conclusion 22
A Appendix: Reproducibility and Implementation Notes 23
References 25

1 Introduction

1.1 Clinical background

Bladder cancer is among the most commonly diagnosed malignancies worldwide, and roughly three quarters of new cases present as non-muscle-invasive bladder cancer (NMIBC): papillary tumours confined to the mucosa (stage Ta), tumours invading the lamina propria but not the muscularis propria (stage T1), and flat high-grade carcinoma in situ (CIS). NMIBC is rarely immediately lethal but is defined clinically by its propensity to recur and, in the high-risk subgroup, to progress to muscle-invasive disease. Risk stratification in routine practice follows the EORTC risk tables, which combine tumour number, size, prior recurrence rate, stage, grade and the presence of CIS into estimates of recurrence and progression probability [1].

For intermediate- and high-risk disease, the standard of care after transurethral resection of the bladder tumour is adjuvant intravesical immunotherapy with Bacillus Calmette–Guérin (BCG), a live attenuated strain of Mycobacterium bovis that has been in continuous clinical use for this indication since the 1970s. Its mechanism is immunological rather than cytotoxic. Instilled bacilli attach to the urothelium through fibronectin, are internalised by urothelial cells and by professional antigen-presenting cells, and engage pattern-recognition receptors including TLR2, TLR4 and NOD2. The resulting NF-κB-driven release of IL-1β, IL-6, IL-8, TNF-α and IL-12 recruits neutrophils, monocytes, macrophages, natural killer cells and T lymphocytes into the bladder wall and polarises the response towards a T-helper-1 phenotype, in which interferon-γ licenses cytotoxic CD8 T-cell activity against tumour cells. BCG additionally induces trained immunity: epigenetic and metabolic reprogramming of myeloid cells and their progenitors that raises the magnitude of the innate response on re-challenge [2].

Despite this, approximately 30–50% of high-risk patients fail to achieve a durable response. BCG-unresponsive disease — defined by consensus as persistent or recurrent high-grade disease despite an adequate course of BCG [3] — carries a substantial risk of progression and frequently results in radical cystectomy, a life-altering operation. There is at present no validated biomarker that predicts response before treatment begins, so patients who will not benefit are nonetheless exposed to months of instillations and to the delay in definitive treatment that failure entails.

1.2 Motivation: communication rather than composition

Because the therapeutic mechanism of BCG is fundamentally a problem of antigen presentation and T-cell priming, the interface between myeloid cells and T cells is a mechanistically motivated place to look for the difference between responders and non-responders. Most single-cell analyses of the bladder tumour immune microenvironment have characterised that interface in terms of who is present: which cell types are enriched or depleted in each response arm, and which genes they express differentially. Composition, however, is a weak summary of a signalling system. Two tissues can contain the same populations in the same proportions and yet differ completely in the ligand–receptor programmes those populations run.

This report asks whether the composition of predicted myeloid-to-T-cell signalling carries information that cell abundance does not, and whether that information tracks BCG response. The question is deliberately narrow, and its scope was fixed in writing before analysis began. General tumour-microenvironment characterisation, all-to-all cell communication, broad immune profiling, biomarker discovery, survival prediction and spatial transcriptomics were all placed explicitly out of scope. In an

exploratory single-cell dataset it is always possible to find something interesting; a question that expands to accommodate whatever the data offers cannot be tested.

1.3 Contributions

The contributions of this work are as follows.

1. **A data-integrity correction.** The public deposit mixes 115 BD Rhapsody AbSeq antibody-derived protein features into the RNA assay, where in two libraries they account for approximately 76% of counts per cell. We identify this, separate the features into a dedicated ADT assay before quality control, and re-run the affected stages (Section 3.2).
2. **A defensible negative result on composition.** Under both percentage-scale Welch testing and a centred-log-ratio sensitivity analysis, no cell-type proportion survives Benjamini–Hochberg correction across twelve comparisons (Section 4.2). This negative result is what makes the communication analysis informative rather than redundant.
3. **A communication-state signature of BCG response.** Restricting inference to tissue and to myeloid senders and T-cell receivers, we show that the relative contribution of signalling pathways differs sharply between response arms, with an MHC-II:MHC-I ratio of 0.33 (naive), 2.97 (responsive) and 0.68 (unresponsive), and that this difference is not explained by sender-cell abundance (Sections 4.4–4.6).
4. **A sensitivity analysis of the central claim.** Because CD4 is annotated as an MHC class II receptor and is detectable on CD8 T cells, the class II estimate is vulnerable to a labelling artefact. Removing CD4-receptor edges on CD8 T cells preserves the ordering and the qualitative pattern (Section 4.5).
5. **A systematic, published audit of concordance.** Rather than reporting only the claims of the source study that we could confirm, we test fifteen of its specific claims and report every outcome, including the eleven that our restricted network cannot address (Section 4.8).
6. **An openly documented, reproducible implementation.** A single configuration file, one seed, code-based validation gates at the end of every script, 575 automated checks, three explicit human decision gates, and a coded audit changelog (Appendix A), together with a public project page describing every pipeline stage and its parameters.

1.4 Report structure

Section 2 situates the work with respect to single-cell studies of BCG response, biomarker efforts in NMIBC, and the methodological literature on communication inference and single-cell statistics. Section 3 describes the dataset and every stage of the pipeline, with the parameter values and the reasoning behind them. Section 4 reports the findings. Section 5 interprets them biologically and methodologically, Section 6 states the limitations and the planned extensions, and Section 7 concludes.

2 Related Work

2.1 Single-cell characterisation of the BCG-treated bladder

Single-cell RNA sequencing has been applied to the NMIBC immune microenvironment to ask which populations differ between patients who respond to BCG and those who do not. The dataset reanalysed here was deposited with the study of Brown et al. [4], which profiled CD45-enriched bladder tissue and matched peripheral blood from BCG-naive, BCG-responsive and BCG-unresponsive patients and reported differences in myeloid and T-cell states between arms, including elevated MHC class II signalling in responsive tissue and elevated TIGIT in unresponsive tissue. That study also proposed a 57-gene signature associated with response.

Our work is a reanalysis rather than a replication. We start from the same author-processed count matrices but differ in quality control, integration method and cell-labelling scheme, and we restrict the communication network to a single directed interface. This means we should not expect complete concordance, and it makes the systematic reporting of non-concordance (Section 4.8) part of the contribution rather than an embarrassment to be omitted.

2.2 Biomarkers of BCG response

Clinical risk stratification in NMIBC remains dominated by the EORTC tables [1], which use only clinicopathological variables and were derived before modern immunotherapy practice. Molecular efforts have largely been transcriptomic. Meghani et al. [5] assembled a bulk RNA-sequencing cohort of 103 patients with recurrence-free survival data, which is the natural external validation set for any single-cell-derived signature; the present study does not yet use it, and doing so is the principal item of planned future work (Section 6.2). The consensus definition of BCG-unresponsive disease used throughout this report follows the recommendations of the International Bladder Cancer Group [3], and the immunological framing of BCG as an inducer of trained immunity follows Netea et al. [2].

2.3 Inference of cell–cell communication

A family of methods now infers intercellular signalling from single-cell expression by scoring the co-expression of cognate ligand–receptor pairs across cell groups. CellChat, used here, differs from simpler permutation-based approaches in three ways that matter for this analysis: it curates signalling into pathway-level families rather than treating each ligand–receptor pair independently; it models multi-subunit receptor complexes and soluble or membrane-bound cofactors, including agonists and antagonists; and it formulates communication probability through a law-of-mass-action-inspired Hill function rather than a mean-expression product [6]. This pathway-level aggregation is precisely what makes the MHC-I versus MHC-II comparison in this report expressible.

The important caveat is shared by every method in this family. What is computed is a probability derived from co-expression, not an observation of interaction. It carries no evidence of spatial proximity, of protein-level expression, or of causality, and it is sensitive to upstream choices — quality control, clustering resolution and cell labelling — that determine which cell groups exist to be connected in the first place. We therefore use the term predicted communication throughout, and we treat the choice of which cells may legitimately be connected (Section 3.9) as a biological decision rather than a computational one.

2.4 Statistical practice in single-cell analysis

Two methodological literatures bear directly on the design of this study. The first concerns differential expression. Testing across tens of thousands of individual cells drawn from a handful of donors treats cells as independent replicates when the true unit of replication is the donor; Squair et al. [7] showed that this pseudoreplication inflates false discovery rates severely, and that aggregating cells into per-sample pseudobulk profiles before applying a count-based model such as DESeq2 [8] restores calibration. We follow that recommendation (Section 3.7).

The second concerns compositional data. Cell-type proportions within a sample are constrained to sum to one, so an increase in one population mechanically forces a decrease in the others and standard tests on proportions can report differences that are arithmetic rather than biological. The centred-log-ratio transformation is the standard remedy, and we report differential abundance both on the percentage scale — to remain comparable with the source study — and after centred-log-ratio transformation as a sensitivity analysis (Section 3.8).

3 Methodology

3.1 Dataset

All analyses use GSE322495, the public deposit accompanying Brown et al. [4] and released under CC BY 4.0 [9]. The deposit contains 21 CD45-enriched single-cell RNA-sequencing libraries from 19 patients with NMIBC, generated on the BD Rhapsody platform and sequenced on an Illumina NovaSeq X Plus. Most libraries contain both bladder tissue and matched peripheral blood mononuclear cells, distinguished by sample multiplexing tags. Libraries are distributed across three clinical arms: BCG-naive (7 libraries), BCG-responsive (6) and BCG-unresponsive (8). Three patients (P1, P2 and P4) contribute paired pre- and post-treatment libraries and are therefore not independent replicates. Table 1 summarises the dataset.

Because the deposit provides author-processed Seurat objects rather than raw reads, no alignment, barcode correction or cell calling was performed; the analysis begins at the count matrix. This is a limitation as much as a convenience, since upstream decisions made by the original authors are inherited and cannot be varied.

Table 1: Composition of the GSE322495 dataset used in this study.

Attribute	Value
Accession	GEO GSE322495 (CC BY 4.0)
Source study	Brown et al., J Clin Invest 2026;136(12):e200442
Platform	BD Rhapsody; Illumina NovaSeq X Plus (GPL34284)
Enrichment	CD45-positive
Libraries / patients	21 libraries from 19 patients
Compartments	Bladder tissue and matched PBMC (sample-tag multiplexed)
BCG-naive	7 libraries
BCG-responsive	6 libraries
BCG-unresponsive	8 libraries
Paired patients	P1, P2, P4 (pre-/post-treatment; not independent)
Cells retained after QC	89,753
Starting material	Author-processed Seurat objects (no raw FASTQ)

3.2 Preprocessing: separation of antibody-derived features

The first substantive finding of the project was a data-integrity problem rather than a biological one. The deposited objects mix 115 BD Rhapsody AbSeq (pAbO) antibody-derived protein features into the RNA assay. In two libraries these protein counts constitute approximately 76% of the total counts per cell.

Left in place, they would dominate every count-based quality-control metric and every normalisation in those libraries: cells would be filtered on library size that is mostly antibody signal, size factors would be set by staining intensity, and highly variable gene selection would be driven by protein features. Cells in the affected libraries would then be clustered on the basis of antibody staining rather than transcription. A dedicated script therefore separates the AbSeq features into their own ADT assay before quality control, and the three downstream stages were re-run on the corrected objects.

3.3 Quality control

Quality control uses adaptive thresholds derived from the median absolute deviation (MAD) rather than fixed cut-offs. For a metric x within a group of cells, the outlier bounds are given by Equation 1, with n_low = 2 for the number of detected genes and for total counts, and n_high = 3 for mitochondrial percentage.

$$\text{lower} = \text{med}(x) - n_{\text{low}} \cdot \text{MAD}(x), \qquad \text{upper} = \text{med}(x) + n_{\text{high}} \cdot \text{MAD}(x) \tag{1}$$

The consequential decision is the grouping. Thresholds are computed within each library × Sample_Tag stratum rather than globally, and the justification is empirical rather than stylistic: mitochondrial content in this cohort separates completely by compartment, running from 0.55% to 3.5% in blood and from 9.8% to 17.9% in tissue with no overlap, while remaining unconfounded with response arm. A single global threshold would therefore have preferentially deleted tissue cells — precisely the cells the research question concerns — for a technical rather than a biological reason.

Because 8 of the 21 libraries produced a negative lower threshold and 7 produced a mitochondrial ceiling below 7%, hard floors were added so that an unusually tight library could not impose an implausibly strict filter on itself: a minimum of 350 detected genes and 500 unique molecular identifiers, and a floor of 15% on the mitochondrial ceiling. These values, and the reasoning above, are recorded in `config/qc_thresholds.R` rather than buried in a script.

One library was excluded outright: the tissue half of GSM9553271 (patient P1, unresponsive), with 91.8% median mitochondrial content and approximately 238 median detected genes, which represents a failed tissue dissociation rather than a biological state. The blood half of the same cartridge passed quality control and was retained. After filtering, 89,753 cells were carried forward.

3.4 Normalisation and a memory-constrained implementation

Two normalisations are carried in parallel by design. SCTransform, which fits a regularised negative binomial model per gene and returns Pearson residuals [10], is used for integration, dimensionality reduction and clustering, where variance stabilisation is what matters. Log-normalised counts are retained for differential expression, because the count-based models used for testing are calibrated on counts and not on residuals. Using one representation for both purposes is a common and consequential error.

Implementation was shaped by a hardware constraint worth recording, because it is the kind of constraint that silently changes analyses. The merged log-normalised object occupies 5.2 GB and the analysis machine has 16 GB of memory, so running SCTransform on all cells as a single object aborted repeatedly. The step was restructured to run SCTransform per library and to assemble only a shared-feature residual matrix for principal component analysis, which reduced peak memory to approximately 2 GB. This is a container-level deviation from the reference workflow: the statistical model is unchanged, only the granularity at which it is fitted.

3.5 Integration, dimensionality reduction and clustering

Integration used Harmony [11] with library as the sole batch variable. Compartment was deliberately excluded from the correction, so that the biological difference between blood and tissue is preserved rather than removed; correcting it away would have eliminated the very axis that motivates restricting communication inference to tissue.

Thirty principal components were retained. The elbow of the variance curve lies at components 10–12 and 30 components capture 91.6% of cumulative variance, but the additional components were kept deliberately to preserve rare populations — plasmacytoid dendritic cells, mast cells — which are the first to be lost when dimensionality is trimmed to the elbow. PC1 alone carries 34.9% of the variance and corresponds to the blood–tissue compartment axis, with library structure nested inside it rather than dominating it, which is itself evidence that the integration was correctly specified.

A shared-nearest-neighbour graph was built with k = 20 and partitioned by Louvain modularity optimisation at resolution 0.5, chosen after inspecting six candidate resolutions, yielding 24 clusters. UMAP embeddings used `n_neighbors` = 30 and `min_dist` = 0.3 for visualisation only; no analysis depends on the embedding. All parameters are collected in Table 2.

Table 2: Pipeline parameters. All values are set once in config/config.R.

Stage	Parameter	Value
Quality control	MAD multipliers (low / high)	2 / 3
	Grouping	library × Sample_Tag
	Hard floors	350 genes; 500 UMIs
	Mitochondrial ceiling floor	15%
Normalisation	Clustering assay	SCTransform (per library)
	Testing assay	Log-normalised RNA
Reduction	Principal components	30
	Integration	Harmony (batch = library)
	UMAP	n_neighbors 30; min_dist 0.3
Clustering	SNN k.param	20
	Louvain resolution	0.5
	Clusters / identities	24 / 20
Markers	Test	Wilcoxon, positive only
	min.pct / logfc.threshold	0.25 / 0.25
Differential expression	Model	Pseudobulk DESeq2
	Significance	padj < 0.05; |log2FC| > 0.5
Communication	Package	CellChat v2.2.0.9001
	Compartment	Tissue only
	Network	Myeloid → T cell
	min.cells / trim	10 / 0.1
	population.size / raw.use	FALSE / TRUE
Reproducibility	Seed	20260819

3.6 Cell-type annotation, and the separation of identity from state

Clusters were annotated from marker genes identified by one-versus-rest Wilcoxon testing restricted to positive markers, with `min.pct` = 0.25 and `logfc.threshold` = 0.25. The 24 clusters mapped to 20 cell identities, of which nine are myeloid — classical and non-classical monocytes, macrophages, inflammatory myeloid, interferon-associated myeloid, plasmacytoid dendritic cells and dendritic-cell-like subsets — and four are T-cell identities: CD8 T cells, CD4 T cells, regulatory T cells and a generic T-cell population.

Neutrophils were deliberately excluded from the communication network. Their abundance and ligand repertoire would otherwise overwhelm the monocyte and macrophage signal that the research question concerns, and the resulting network would describe granulocyte biology under a myeloid label.

A methodological distinction was enforced throughout: identity and state are kept separate. Five myeloid transcriptional programmes — M1-like, M2-like, lipid-associated tumour-associated macrophage (LA-TAM), interferon TAM, and antigen presentation — were scored as continuous per-cell programmes but were never used to define clusters. Macrophage polarisation is a continuum, and allowing a state score to determine a discrete label would manufacture the very categories the analysis then claims to discover.

3.7 Differential expression

Differential expression was carried out on pseudobulk profiles with DESeq2 [8], aggregating cells to one profile per patient per cell type so that the patient, not the cell, is the unit of replication [7]. Paired patients were blocked on patient identity rather than treated as independent samples. Significance was called at padj < 0.05 (Benjamini–Hochberg [12]) and |log2 FC| > 0.5.

3.8 Differential abundance

Cell-type abundance was tested in two ways. The first is a Welch t-test on percentage-scale proportions, which reproduces the approach of the source study and therefore permits direct comparison. The second is a sensitivity analysis on centred-log-ratio (CLR) transformed proportions, Equation 2, which removes the sum-to-one constraint that otherwise induces spurious negative associations between populations. Multiple testing across the twelve comparisons performed was controlled by the Benjamini–Hochberg procedure.

$$\mathrm{clr}(\mathbf{p})_i = \ln \frac{p_i}{\left(\prod_{j=1}^{D} p_j\right)^{1/D}} \tag{2}$$

3.9 Cell–cell communication inference

Communication was inferred with CellChat v2.2.0.9001 [6], following the complete documented workflow rather than a shortened approximation: object construction, ligand–receptor database assignment, data subsetting, identification of over-expressed genes and interactions, projection onto the human protein–protein interaction network, computation of communication probabilities, filtering of interactions supported by too few cells, aggregation to pathway level, and network aggregation.

CellChat models the probability of communication between sender group i and receiver group j for a ligand–receptor pair as a Hill function of the ensemble ligand and receptor expression, modulated by agonists and antagonists and, optionally, by the sizes of the two cell populations (Equation 3), where L and R are trimmed-mean ligand and receptor expression, AG and AN are agonist and antagonist terms, and n are cell counts.

$$P_{i,j} = \frac{L_i R_j}{K_h + L_i R_j} \cdot \left(1 + \frac{\mathrm{AG}_i}{K_h + \mathrm{AG}_i}\right) \cdot \left(1 + \frac{\mathrm{AN}_i}{K_h + \mathrm{AN}_i}\right)^{-1} \cdot \underbrace{\frac{n_i\, n_j}{\bar{n}^2}}_{\text{disabled here}} \tag{3}$$

Setting `population.size = FALSE` removes the final factor, so that a predicted interaction reflects expression rather than how many cells happen to be present — which matters here, because sender abundance is itself one of the quantities under comparison (Section 4.6).

Two restrictions were imposed for biological rather than computational reasons.

- **Tissue only.** Cells in blood and cells in the bladder wall are not in physical contact, so an inferred interaction between them would be an artefact of co-expression. Blood cells are retained in the atlas but excluded from communication inference.
- **Myeloid senders, T-cell receivers.** The network is restricted to the directed interface named in the research question rather than surveying all possible pairs, which both matches the hypothesis and avoids a multiplicity problem created by the analyst.

Remaining parameters were `min.cells` = 10 per cell type per group, `trim` = 0.1 and `raw.use` = TRUE.

Pathway-level results are reported as a relative contribution. For response arm g and pathway p, with S the aggregated predicted communication strength, the relative contribution and the MHC-II:MHC-I balance are defined in Equation 4.

$$c_{g,p} = \frac{S_{g,p}}{\sum_q S_{g,q}}, \qquad \rho_g = \frac{S_{g,\text{MHC-II}}}{S_{g,\text{MHC-I}}} \tag{4}$$

3.10 Sensitivity analysis

CD4 is annotated as an MHC class II receptor in the CellChat database and is detectable at the transcript level on CD8 T cells. The class II estimate is therefore vulnerable to a labelling artefact rather than a biological one. The full pathway analysis was repeated with all CD4-receptor edges on CD8 T cells removed, and both values of the ratio are reported.

3.11 Validation of published claims

Fifteen specific claims from the source study [4] were extracted and each was assigned, before inspection of the corresponding result, to one of four outcomes: supported, direction differs, not supported, or not testable. A claim was recorded as not testable when the ligand–receptor pair falls outside the restricted myeloid→T-cell network, when cell labels could not be mapped unambiguously between the two schemes, or when fewer than `min.cells` cells support the relevant group. Reporting this category is deliberate: a claim that this analysis cannot address is not evidence for or against the source study, and omitting it would misrepresent the strength of the concordance.

3.12 Software and reproducibility

Analyses were carried out in R with Seurat and SeuratObject [13], sctransform, harmony, DESeq2, CellChat v2.2.0.9001, Matrix, dplyr, tidyr, ggplot2 and `here`. A single seed (20260819) is set once in `config/config.R`. The engineering controls that support reproducibility — validation gates, human decision gates, and the audit changelog — are described in Appendix A.

4 Results

4.1 Atlas and annotation

After AbSeq separation, adaptive quality control and the exclusion of one failed tissue library, 89,753 cells were retained. Harmony integration on library produced an embedding in which the dominant axis (PC1, 34.9% of variance) is the blood–tissue compartment rather than any single library, and Louvain clustering at resolution 0.5 yielded 24 clusters that marker-based annotation mapped to 20 identities: nine myeloid, four T-cell, and seven remaining lymphoid and non-myeloid populations. Figure 1 shows the annotated embedding.

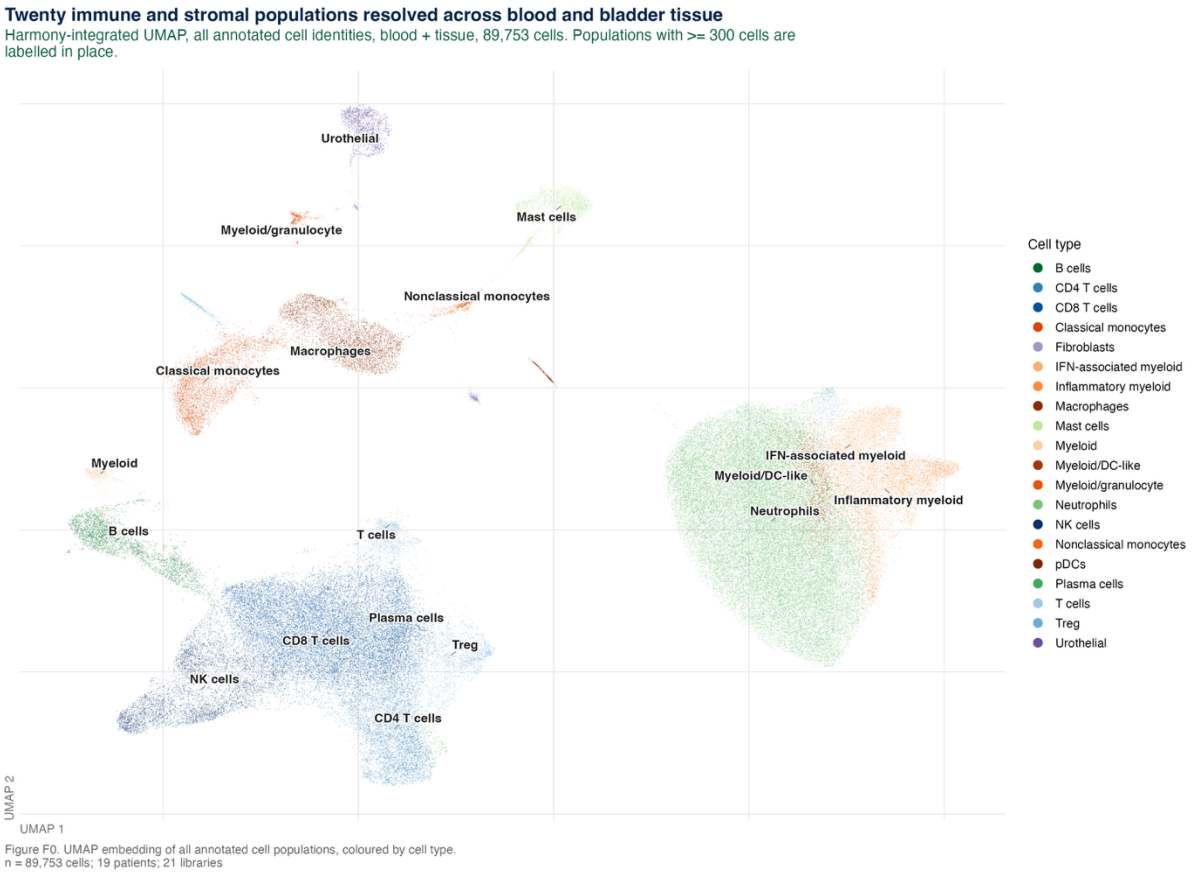


Figure 1: UMAP embedding of 89,753 cells after Harmony integration on library, coloured by annotated cell identity. Myeloid and T-cell compartments are the populations carried forward into communication inference; blood cells are retained in the atlas but excluded from the communication analysis.

4.2 Cellular composition does not separate the response arms

Differential abundance was tested across twelve comparisons: on the percentage scale, to match the source study, and again after centred-log-ratio transformation as a sensitivity analysis. The result is the same under both approaches, and it is negative: **no cell-type proportion survives Benjamini–Hochberg correction**. The myeloid-to-T-cell ratio in particular shows substantially overlapping distributions between arms.

Pseudobulk differential expression with patient as the replicate unit likewise yielded no large, consistently reproducible set of arm-discriminating genes within the myeloid and T-cell compartments at padj < 0.05 and $|\log2\ FC| > 0.5$ once paired patients were blocked.

This negative result is reported first because it is what makes everything that follows informative. Had cellular composition separated the response arms, a communication analysis would largely be restating the same information in a more elaborate form.

4.3 Predicted communication networks differ in structure

Restricting inference to tissue and to myeloid senders and T-cell receivers, CellChat produced visibly different network structures in the three arms (Figure 2). Edge weights, and the identity of the dominant sender–receiver pairs, vary across arms in a way that is not predicted by node size — that is, by cell abundance.

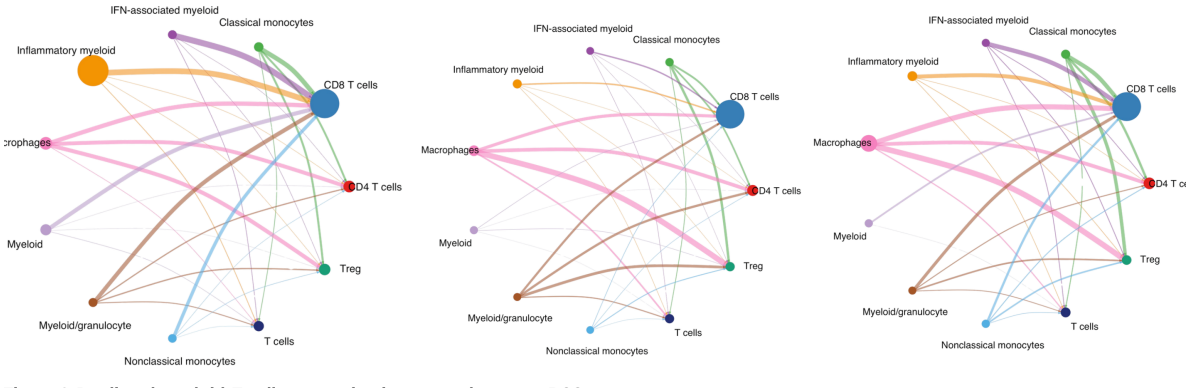


Figure 2: Predicted myeloid–T-cell communication networks across BCG response groups. CellChat v2 was used to infer ligand–receptor communication between myeloid and T-cell populations in bladder tissue. Networks are shown for BCG-naive, BCG-responsive and BCG-unresponsive groups. Node size represents cell abundance; edge width represents predicted communication strength.

4.4 Pathway composition and the MHC-II:MHC-I balance

The informative result is compositional rather than structural. The relative contribution of signalling pathways (Equation 4) shifts sharply between arms (Figure 3, Table 3).

In BCG-naive tissue, MHC class I accounts for 56.3% of predicted myeloid-to-T-cell signalling and MHC class II for 18.6%, a class-II-to-class-I ratio of 0.33. In BCG-responsive tissue the relationship inverts: MHC class I falls to 16.3% while MHC class II rises to 48.5%, giving a ratio of 2.97 — the only arm in which class II signalling exceeds class I. BCG-unresponsive tissue sits between the two at 35.9% and 24.3% (ratio 0.68) and is further distinguished by galectin signalling rising to 15.1%, a level it does not reach in either other arm.

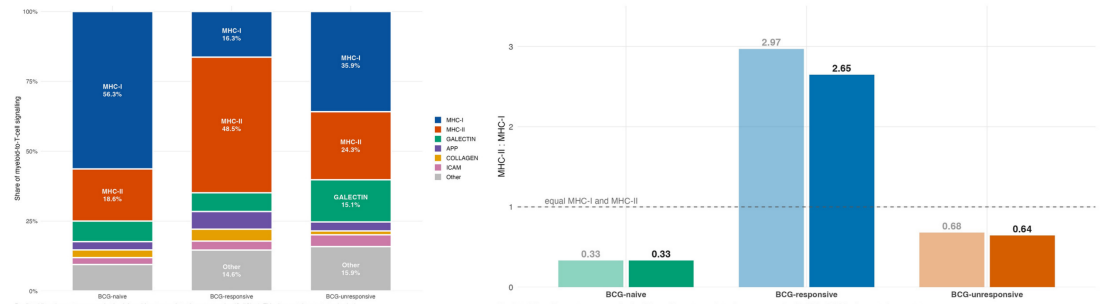


Figure 3: Composition of predicted myeloid-to-T-cell signalling by response arm (left) and the MHC-II:MHC-I ratio with and without CD4-receptor edges on CD8 T cells as a sensitivity check (right).

Responsive tissue is the only arm in which class II signalling exceeds class I; galectin signalling is a leading pathway only in unresponsive tissue.

Table 3: Composition of predicted myeloid→T-cell signalling by response arm. Percentages are relative contributions to total predicted signalling within each arm; the ratio is the MHC-II:MHC-I balance defined in Equation 4.

Measure	BCG-naive	BCG-responsive	BCG-unresponsive
MHC-I (% of predicted signalling)	56.3	16.3	35.9
MHC-II (% of predicted signalling)	18.6	48.5	24.3
Galectin (% of predicted signalling)	—	—	15.1
MHC-II : MHC-I ratio	0.33	2.97	0.68
MHC-II : MHC-I ratio, sensitivity analysis	0.33	2.65	0.64
Myeloid sender cells	3,561	2,098	3,664
Total predicted signalling strength	9.353	10.127	11.565

4.5 The pattern survives a sensitivity analysis

Because CD4 is annotated as an MHC class II receptor in the CellChat database and is detectable on CD8 T cells, the class II estimate could in principle be a labelling artefact. Repeating the analysis with all CD4-receptor edges on CD8 T cells removed moved the responsive ratio from 2.97 to 2.65 and the unresponsive ratio from 0.68 to 0.64, leaving the naive arm unchanged at 0.33. The ordering — responsive far above unresponsive, and unresponsive above naive — and the qualitative pattern are preserved, so the finding is not an artefact of that single annotation.

4.6 Signal composition is not explained by sender abundance

If the pathway differences simply tracked how many myeloid cells were present, the result would be a restatement of composition. They do not (Figure 4, Table 3). BCG-naive and BCG-unresponsive tissue contain almost the same number of myeloid sender cells — 3,561 against 3,664 — yet generate total predicted signalling strengths of 9.353 and 11.565 respectively. BCG-responsive tissue contains the fewest myeloid senders of the three arms, 2,098, yet produces an intermediate total signal of 10.127 and by a wide margin the highest MHC class II share. Signal composition and signal magnitude therefore vary independently of sender-cell abundance, which is exactly the regime in which communication analysis adds information.

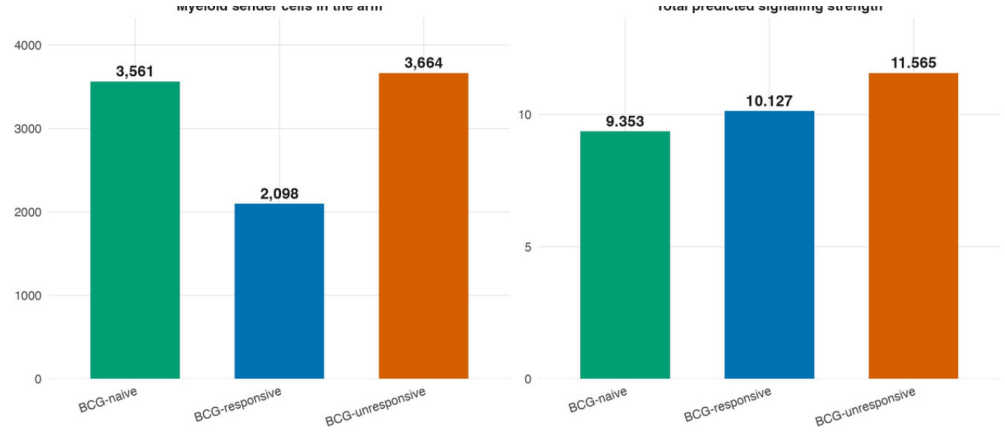


Figure 4: Myeloid sender-cell abundance (left) against total predicted myeloid-to-T-cell signalling strength (right) in each response arm. The responsive arm has the fewest senders but not the weakest signal.

4.7 Macrophages dominate CD8 T-cell input in every arm

Decomposing the predicted input into CD8 T cells by sender identity, macrophages are the strongest non-autocrine predicted source in all three response groups — not only in non-responders, as the corresponding figure of the source study suggests. This is a point of partial disagreement with [4] and is recorded as such in Section 4.8.

4.8 Systematic validation of published claims

Fifteen specific claims from the source study were tested against these results under the protocol of Section 3.11. The outcomes are given in Table 4: two claims are supported, one differs in direction, one is not supported, and eleven are not testable within a network deliberately restricted to myeloid senders and T-cell receivers.

Table 4: Outcome of the systematic validation of fifteen claims from Brown et al. [4].

Outcome	n	Detail
Supported	2	MHC-II signalling elevated in BCG-responsive tissue; TIGIT higher in BCG-unresponsive tissue
Direction differs	1	CXCL16 predicted higher in unresponsive tissue here, whereas the source study reports the opposite direction
Not supported	1	TGF-β shows no directional difference between response arms in this analysis
Not testable	11	Ligand–receptor pairs absent from the restricted myeloid→T-cell network, unmapped between cell-label schemes, or supported by too few cells — including PD-L1, PD-L2, CXCL9/10/11/12, IL2 and TNFSF10

Three methodological differences from the source analysis bear on this table: our quality-control strategy, our integration method (Harmony rather than canonical correlation analysis) and our cell-labelling scheme all differ from those used by the original authors, and any one of them can change which interactions survive filtering.

4.9 Dissemination

The work was presented as an A0 conference poster (Figure 5) and is documented in full — research question, dataset, all nine pipeline stages with their parameters, findings, claim validation and limitations — on the public project page linked on the cover of this report [14].

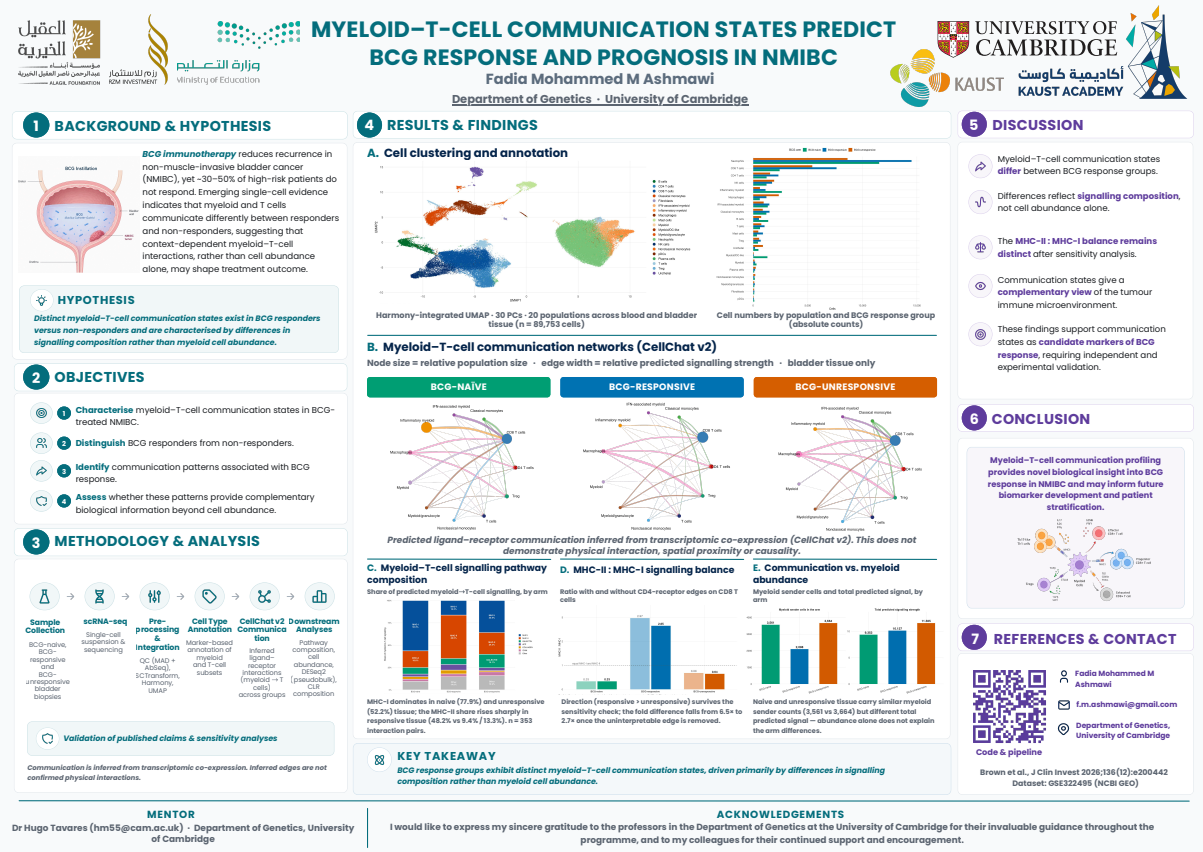


Figure 5: Conference poster summarising the project, presented at Christ's College, University of Cambridge, on 21 August 2026.

5 Discussion

5.1 Biological interpretation

The central observation is that an MHC class II-dominant myeloid-to-T-cell axis is specific to responsive tissue, and this is coherent with what BCG is understood to require. Class II presentation on professional antigen-presenting cells engages CD4 T cells, and it is the CD4 helper compartment that supplies interferon-γ and the licensing signals underpinning the T-helper-1 polarisation and cytotoxic CD8 activity on which BCG efficacy depends [2]. A tissue whose myeloid compartment is predicted to talk to T cells predominantly through class I is, in this reading, running a presentation programme that is poorly matched to the mechanism the therapy relies on — even where the same cell types are present in the same numbers.

The naive arm is instructive as a control. It is not a failed response; it is tissue that has not yet been asked to mount one, and its strongly class I-dominant profile (ratio 0.33) is what an untreated bladder immune compartment might be expected to look like. Unresponsive tissue sits between naive and responsive (ratio 0.68), which is consistent with a partial or aborted shift towards the helper axis rather than with a total absence of engagement.

The rise of galectin signalling to 15.1% of predicted signalling only in unresponsive tissue is worth noting for the opposite reason. Galectin-9 engagement of TIM-3 is a well-characterised inhibitory axis associated with T-cell exhaustion, and its appearance as a leading pathway precisely in the arm that fails is the kind of coincidence that generates hypotheses. We present this as interpretation and not as a finding: nothing in this analysis demonstrates that these interactions occur, still less that they cause the difference in outcome. Establishing that would require spatial confirmation of proximity, protein-level validation of the ligand–receptor pair, and a functional perturbation.

5.2 Why communication adds information beyond composition

The pairing of a negative composition result with a positive communication result is the methodological core of this study. Cell-type proportions in these samples do not separate the response arms once multiple testing is corrected for (Section 4.2), and predicted signalling strength does not track sender abundance (Section 4.6): the responsive arm has the fewest myeloid senders yet neither the weakest total signal nor anything close to the lowest class II share. What differs between arms is therefore which ligand–receptor programmes the same populations are predicted to run, not how many of those populations there are.

This is an argument for treating communication state as a distinct descriptive layer of the tumour immune microenvironment rather than as a downstream visualisation of composition. Disabling CellChat's population-size term (Equation 3) was necessary for the argument to be made at all: with the term enabled, arm differences in predicted strength would partly encode arm differences in cell number, and the two layers could not be separated.

5.3 Methodological reflections

Three aspects of the analysis generalise beyond this dataset.

Data integrity precedes data analysis. The AbSeq features mixed into the RNA assay (Section 3.2) would not have produced an error, a warning or an implausible figure. They would have produced a quality-control step that silently filtered on antibody staining intensity in two libraries and a clustering

that partly reflected staining rather than transcription. This class of defect is invisible to a pipeline that begins at normalisation, and it argues for inspecting the contents of a deposit — assay by assay, feature by feature — before trusting it.

Grouping is a scientific decision, not a default. Computing quality-control thresholds globally rather than within library and compartment would have been the conventional choice and would have deleted tissue cells preferentially, because mitochondrial content separates completely by compartment in this cohort with no overlap. The threshold values matter far less than the stratum within which they are computed.

Inference should be named as inference. Every quantity reported here is a predicted communication probability derived from co-expression under the model of Equation 3. Reporting these as observed interactions would overstate the evidence by a wide margin, and the vocabulary used in a report shapes how its results are subsequently cited.

5.4 Relation to the source study

Of fifteen claims from Brown et al. [4] tested here, two are supported, one differs in direction, one is not supported and eleven are not testable within our restricted network (Table 4). The concordance on MHC class II elevation in responsive tissue is the substantive agreement, and it is reached by a different route: a different quality-control strategy, a different integration method, a different cell-labelling scheme, and a network restricted to a single directed interface. Agreement obtained under different analytical choices is stronger evidence than agreement obtained by repeating the same choices.

The disagreements are equally worth stating. Our finding that macrophages are the strongest non-autocrine predicted source into CD8 T cells in all arms, rather than characteristically in non-responders, and the reversed direction for CXCL16, both plausibly follow from those same methodological differences. The eleven untestable claims are not evidence against the source study; they are a consequence of a deliberately narrow network, and they are reported so that the two supported claims are not read as a higher rate of concordance than the analysis can justify.

6 Limitations and Future Work

6.1 Limitations

The value of this study depends on its limitations being stated plainly rather than qualified away.

- **Cohort size.** Nineteen patients, with as few as one to seven per arm for some cell types. This is hypothesis-generating work, not a confirmatory study, and no effect size reported here should be treated as an estimate for a population.
- **Inference, not observation.** CellChat probabilities derive from co-expression of ligand and receptor genes under Equation 3. They carry no evidence of spatial proximity, protein-level expression, or causality, and no co-culture, blocking or functional validation was performed.
- **No outcome data.** GSE322495 is cross-sectional and carries no recurrence-free survival information. Nothing here can be described as prognostic in the clinical sense; response arm is a label assigned at sampling, not an observed time to event.
- **Non-independence.** Three patients (P1, P2, P4) contribute paired pre- and post-treatment libraries. This is handled by blocking in the differential-expression model but cannot be handled within a per-arm communication analysis, which pools libraries by arm.
- **Severe cell imbalance.** The tissue CD8:CD4 ratio is approximately 6:1 and non-classical monocytes are substantially under-represented, so several comparisons rest on few cells even after the min.cells filter.
- **Inherited upstream processing.** The deposit provides processed objects rather than raw reads, so alignment, barcode correction and cell calling are inherited and could not be varied.
- **Method sensitivity.** The quality-control strategy, the integration method (Harmony rather than canonical correlation analysis) and the cell-labelling scheme all differ from the source study, and any of them can change which interactions survive filtering. The CD4-edge sensitivity analysis addresses one specific vulnerability; it does not establish robustness to these broader choices.
- **Pooling by arm.** Communication is currently inferred per response arm rather than per patient, so no measure of between-patient variability accompanies the ratio, and no formal statistical test of the arm difference is reported.

6.2 Future work

The limitations above define the work programme, which is organised in three components. Component A — the descriptive myeloid–T-cell communication analysis reported here — is complete. Components B and C are outstanding.

Per-patient communication features. Re-running CellChat per patient rather than per arm would convert the MHC-II:MHC-I ratio from a pooled descriptive statistic into a per-sample feature with a distribution, allowing a proper statistical test of the arm difference, an estimate of between-patient variability, and use of the ratio as a covariate. This is the single highest-value next step and requires no new data.

Completion of the published gene signature. Only 6 of the 57 genes in the signature reported by the source study are publicly named in the main text; the remainder are recoverable only by manual transcription from supplemental figures. Completing this is a prerequisite for any comparison between a communication-state feature and the existing transcriptional signature.

External validation with outcome data (Component B). The natural validation set is the 103-patient bulk RNA-sequencing cohort of Meghani et al. [5], which carries recurrence-free survival. A per-patient communication feature cannot be computed directly from bulk data, so this requires deconvolution or a surrogate — for example an MHC-II:MHC-I expression balance restricted to a myeloid signature — and validation against both the EORTC risk score [1] and recurrence-free survival.

Comparison against checkpoint markers (Component C). Whether a communication-state feature adds information over individual checkpoint-marker expression (TIGIT, TIM-3, PD-L1) in the same validation cohort is the question that determines whether this line of work is clinically interesting or merely internally coherent.

Experimental and spatial validation. Confirming that the predicted MHC class II and galectin interactions occur in tissue requires spatial transcriptomics or multiplexed imaging to establish proximity, and functional assays — co-culture with blocking antibodies against galectin-9/TIM-3, or class II blockade — to establish consequence. Only this class of experiment can convert a predicted interaction into a mechanism.

7 Conclusion

This project asked a narrow, mechanistically motivated question: does the communication between myeloid cells and T cells differ between BCG-responsive and BCG-unresponsive non-muscle-invasive bladder cancer, and does it carry information that cell abundance does not? Reanalysing 89,753 cells from 19 patients in GSE322495 through a fully reproducible pipeline, the answer to both parts is a qualified yes.

The qualification comes first because it is what makes the result meaningful. Cellular composition does not separate the response arms: no cell-type proportion survives Benjamini–Hochberg correction across twelve comparisons, on the percentage scale or after centred-log-ratio transformation. Against that background, the composition of predicted myeloid-to-T-cell signalling differs sharply. The MHC-II:MHC-I balance is 0.33 in BCG-naive tissue, 2.97 in BCG-responsive tissue and 0.68 in BCG-unresponsive tissue — responsive tissue being the only arm in which class II signalling exceeds class I — and this ordering survives a sensitivity analysis removing CD4-receptor edges on CD8 T cells. The difference is not an abundance effect: the responsive arm contains the fewest myeloid senders of the three yet neither the weakest total predicted signal nor anything close to the lowest class II share. Galectin signalling rises to a leading pathway only in the arm that fails.

That pattern is biologically coherent with a therapy whose efficacy depends on CD4-mediated help and T-helper-1 polarisation, and it is offered as a hypothesis rather than a conclusion. The cohort is small, the design is cross-sectional and carries no survival data, and every interaction reported is a probability inferred from co-expression rather than an observed event. What can be claimed is narrower and, we think, more useful: that myeloid–T-cell communication state is a descriptive layer of the tumour immune microenvironment that is not recoverable from cell composition, that it varies systematically with BCG response in this cohort, and that it is therefore worth measuring per patient and testing against outcome data in an independent cohort.

Two secondary contributions are worth restating. The first is a data-integrity correction: 115 AbSeq antibody-derived features mixed into the RNA assay of a public deposit, accounting for approximately 76% of counts per cell in two libraries, which would have silently corrupted quality control and clustering had they not been separated. The second is the systematic audit of fifteen claims from the source study, in which the eleven claims this analysis could not test are reported alongside the two it supports — because a claim that cannot be addressed is not evidence, and reporting only the concordant subset would overstate the agreement.

A Appendix: Reproducibility and Implementation Notes

A substantial share of the project's effort went into making the analysis reproducible rather than merely correct. This appendix records the engineering controls, because they are part of the claim: a result that cannot be regenerated is a report about a computation rather than a finding.

A.1 Configuration and determinism

Every tunable value in the pipeline is declared once, in `config/config.R`, and read from there by each script; no threshold, resolution or parameter is written inline. The random seed (20260819) is set once in the same file. Paths are resolved relative to the project root through the `here` package rather than hard-coded, so the pipeline runs unchanged from any working directory. Quality-control thresholds and their written justification live in a separate file, `config/qc_thresholds.R`, so that the reasoning travels with the numbers.

Figure-generating code is separated from analytical code. Redrawing a figure therefore cannot alter a result, and a change to a plot cannot silently propagate into a statistic.

A.2 Validation gates

Each script terminates in a code-based validation gate that asserts the properties its output must satisfy — object dimensions, assay presence, metadata integrity, the existence of expected files — so a failure halts the pipeline at its origin instead of propagating quietly into a downstream figure. Across configuration loading, path resolution, script parsing and figure generation, 575 automated checks run in total.

A.3 Human decision gates

Three decisions in this pipeline are scientific judgements and cannot be automated: the quality-control thresholds, the clustering resolution, and the mapping from cluster to cell identity. Each is guarded by an explicit status flag — `QC_STATUS`, `CLUSTER_STATUS` and `LABEL_STATUS` — that a human must set and that an assertion enforces. The pipeline refuses to run past an unmade decision rather than falling back on a default, so a silent default can never masquerade as a considered choice. Once set, the decision is applied deterministically thereafter.

A.4 Audit changelog

Every change to the analysis is recorded in `reports/audit/CHANGELOG.md` under a fixed schema (Table 5). The log deliberately includes corrections that would have been easier to leave out: reconciling three conflicting myeloid/T-cell label sets to a single authoritative owner; repairing a malformed bracket assignment that prevented the CellChat script from executing; and correcting a mismatch in which the configuration file recorded `population.size = TRUE` while the script had actually run with `FALSE`. That last entry matters directly to the interpretation in Section 5, since the population-size term is exactly what would have conflated signal strength with cell abundance.

Table 5: Schema of the coded audit changelog. Every entry records all six fields.

Field	Content
OLD	The previous value, behaviour or label set
NEW	The replacement
WHY	The reason the change was necessary
SOURCE	The evidence or authority for the new value
EXPECTED EFFECT	What the change was predicted to alter downstream
VALIDATION	The check confirming the predicted effect occurred

A.5 Course-compliance audit

Because the pipeline was built on the single-cell curriculum taught during the programme [15], it was audited against that curriculum rather than assumed to follow it. Of the ten available practicals, eight are implemented; one is not applicable, since the deposit contains no raw sequencing data and no alignment or cell calling was performed; and one — conversion to a `SingleCellExperiment` container — was intentionally not used. There are no method-altering deviations. Two deviations are container-level only and are driven by the memory constraint described in Section 3.4: SCTransform is fitted per library rather than on a single merged object, and the residual matrix is assembled over shared features. The statistical model is unchanged in both cases.

A.6 Reproducing the analysis

The pipeline runs as an ordered sequence of nine stages — data acquisition, quality control, normalisation, dimensionality reduction and clustering, annotation, differential expression, differential abundance, myeloid state scoring, and communication inference — each of which reads its inputs from the outputs of the previous stage and writes to a versioned output directory. Stale outputs are archived rather than overwritten, so a rerun can be compared against its predecessor instead of silently replacing it. The repository and the project page linked on the cover of this report document each stage, its parameters and its outputs in full.

References

[1] R. J. Sylvester, A. P. M. van der Meijden, W. Oosterlinck, et al. “Predicting Recurrence and Progression in Individual Patients with Stage Ta T1 Bladder Cancer Using EORTC Risk Tables”. In: *European Urology* 49.3 (2006), pp. 466–477. doi: 10.1016/j.eururo.2005.12.031.

[2] M. G. Netea, L. A. B. Joosten, E. Latz, et al. “Trained Immunity: A Program of Innate Immune Memory in Health and Disease”. In: *Science* 352.6284 (2016), aaf1098. doi: 10.1126/science.aaf1098.

[3] A. M. Kamat, R. J. Sylvester, A. Böhle, et al. “Definitions, End Points, and Clinical Trial Designs for Non–Muscle-Invasive Bladder Cancer: Recommendations from the International Bladder Cancer Group”. In: *Journal of Clinical Oncology* 34.16 (2016), pp. 1935–1944. doi: 10.1200/JCO.2015.64.4070.

[4] R. J. Brown, M. T. Khan, A. J. Houston, et al. “Single-Cell Analysis of the Immune Microenvironment in BCG-Treated Non-Muscle-Invasive Bladder Cancer”. In: *Journal of Clinical Investigation* 136.12 (2026), e200442. doi: 10.1172/JCI200442.

[5] K. Meghani, L. F. Cooley, B. Choy, et al. “First-in-Human Intravesical Delivery of Immune Checkpoint Inhibition and Transcriptomic Correlates of Response in Non–Muscle-Invasive Bladder Cancer”. In: *European Urology* 89.4 (2024), pp. 346–352.

[6] S. Jin, C. F. Guerrero-Juarez, L. Zhang, et al. “Inference and Analysis of Cell–Cell Communication Using CellChat”. In: *Nature Communications* 12 (2021), p. 1088. doi: 10.1038/s41467-021-21246-8.

[7] J. W. Squair, M. Gautier, C. Kathe, et al. “Confronting False Discoveries in Single-Cell Differential Expression”. In: *Nature Communications* 12 (2021), p. 5692. doi: 10.1038/s41467-021-25960-2.

[8] M. I. Love, W. Huber, and S. Anders. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2”. In: *Genome Biology* 15 (2014), p. 550. doi: 10.1186/s13059-014-0550-8.

[9] Gene Expression Omnibus. *GSE322495: Single-Cell RNA Sequencing of BCG-Treated Non-Muscle-Invasive Bladder Cancer*. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE322495. NCBI Gene Expression Omnibus, CC BY 4.0. Accessed August 2026. 2026.

[10] C. Hafemeister and R. Satija. “Normalization and Variance Stabilization of Single-Cell RNA-Seq Data Using Regularized Negative Binomial Regression”. In: *Genome Biology* 20 (2019), p. 296. doi: 10.1186/s13059-019-1874-1.

[11] I. Korsunsky, N. Millard, J. Fan, et al. “Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony”. In: *Nature Methods* 16 (2019), pp. 1289–1296. doi: 10.1038/s41592-019-0619-0.

[12] Y. Benjamini and Y. Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society: Series B* 57.1 (1995), pp. 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[13] Y. Hao, T. Stuart, M. H. Kowalski, et al. “Dictionary Learning for Integrative, Multimodal and Scalable Single-Cell Analysis”. In: *Nature Biotechnology* 42 (2024), pp. 293–304. doi: 10.1038/s41587-023-01767-y.

[14] F. M. M. Ashmawi. *Myeloid–T-Cell Communication States and BCG Response in Bladder Cancer — Project Website*. https://f-ash.github.io/bcg-nmibc-website_final/. Accessed August 2026. 2026.

[15] Centre for Research Informatics Training, University of Cambridge. *Bioinformatics and Data Science Training Materials*. https://cambiotraining.github.io/. Accessed August 2026. 2026.