

Learning gene co-regulation across cells with Multicellular Transformer

Maciej Wiatrak^{1,2,3}, Alma Andersson¹, Kam Hon Hoi¹, Tommaso Biancalani¹,
Aïcha BenTaieb^{†1}

¹Genentech Inc., ²University of Cambridge, ³Cambridge Stem Cell Institute

Capturing the intricate molecular interplay between cells within their native tissue environment remains a fundamental challenge in understanding cellular communication and the coordination of essential biological processes. Current computational models often treat cells as isolated entities, neglecting the context of multicellular interactions. We present Multicellular Transformer, a single-cell foundation model that learns gene co-regulation by incorporating both intra- and inter-cellular interactions within a tissue. Unlike traditional approaches that model gene expression within individual cells, our model generates highly contextual gene embeddings by considering thousands of other cells in a multicellular environment. Multicellular Transformer reveals gene-gene interactions critical for cell-cell communication, including receptor-ligand pairs, thus paving the way for the discovery of novel putative receptor-ligand interactions. In lupus, using the contextual gene embeddings produced by Multicellular Transformer we identified disease-specific genes undetectable by differential expression analysis. By taking multiple cells as input, Multicellular Transformer estimates gene perturbation effects at the population level, allowing for more accurate safety assessments and the identification of novel, highly specific therapeutic targets. Overall, Multicellular Transformer is a large pre-trained model that facilitates systematic analysis of co-regulated genes within a multicellular environment.

1. Introduction

Cells do not operate in isolation; their functions are profoundly influenced by interactions with neighboring cells and the surrounding tissue environment ([Armingol et al. 2020](#); [Daneshpour and Youk 2019](#); [Nakahama 2010](#)). Gene expression enables cells to send and receive signals, coordinating immune responses and maintaining tissue health ([Nakahama 2010](#)). Advances in single-cell RNA sequencing (scRNA-seq) have revealed intricate gene relationships within individual cells ([Tang et al. 2009](#); [Kolodziejczyk et al. 2015](#); [Aibar et al. 2017](#)), while spatial transcriptomics has illuminated how cellular organization within tissues drives collective behavior ([Cang et al. 2023](#); [Ji et al. 2020](#); [Qi et al. 2022](#)). Despite these technological advancements, neither approach – individually or combined – directly captures how gene-level regulation is distributed

[†]Corresponding author. Aïcha BenTaieb (bentaieb.aicha@gene.com)

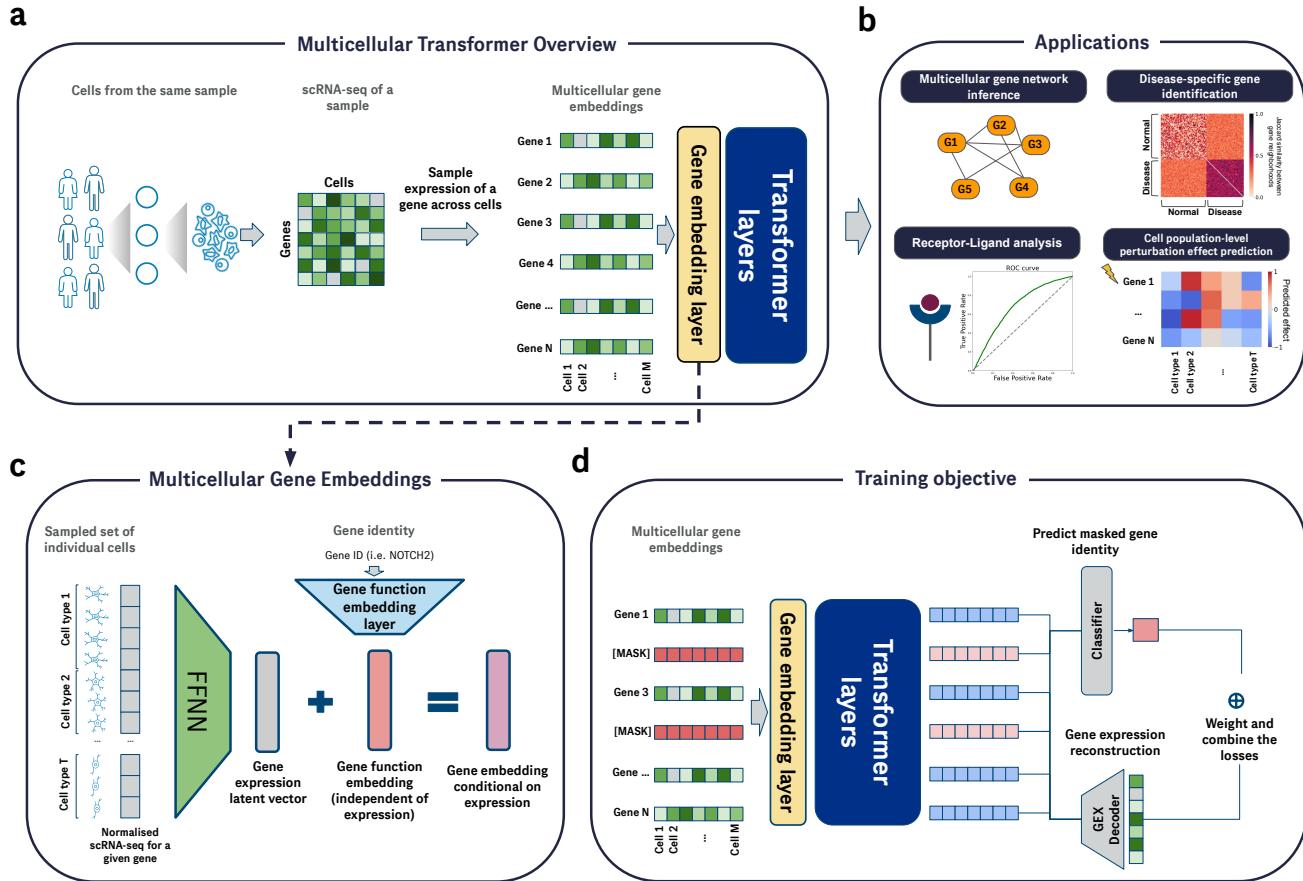


FIGURE 1. Multicellular Transformer learns gene co-regulation in a tissue environment. (a) Multicellular Transformer takes as input scRNA-seq of a set of cells originating from the same sample. The normalized expression across distinct cells in the sample for each gene are used to compute embeddings for each gene, which are then used as input to the transformer model. Transformer computes contextual gene embeddings across cells in a sample which can be then used in downstream tasks. (b) Multicellular Transformer can be applied to a number of downstream tasks, including gene network inference, disease-specific gene identification, receptor-ligand analysis, zero-shot perturbation effect prediction at cell population level and more. (c) Multicellular gene embeddings. The expression across cells for each gene is fed into a feedforward neural network (FFNN), which is shared across all genes to prevent overfitting. To allow for capturing the interactions between different cell types, the input units of a FFNN are fixed to certain cell types during pretraining. The final gene embedding consists of a gene expression latent vector which varies with expression and a gene function embedding, which is independent of expression. The gene expression latent vectors captures the activity of the gene in the sample, while the gene function embedding represents the sample-independent function of the gene and is learned from scratch. The gene embedding layer and gene expression FFNN are summed to form a final gene embedding vector used as input to the transformer model. (d) Multicellular Transformer is trained using masked learning objective, randomly masking 15% of gene function embedding and/or expression as well as gene expression reconstruction loss (Methods).

across multiple cells. This gap is critical: in complex biological processes—such as tissue repair and disease progression—each cell’s behavior is influenced by its environment. Understanding such distributed regulation is essential for translating basic discoveries into applications in drug development, tissue engineering, and beyond.

Existing computational tools address only parts of this challenge. Methods such as SCENIC (Sande et al. 2020) and CellOracle (Kamimoto et al. 2023) reconstruct intracellular gene regulatory networks, while DIALOGUE (Jerby-Arnon and Regev 2022) attempts to capture intercellular programs but requires large homogeneous datasets. More recent foundation models (Theodoris et al. 2023; Cui et al. 2024; Hao et al. 2024; Heimberg et al. 2023; Kalfon et al. 2025) have broadened the scope of single-cell analyses but remain confined to within-cell processes. Similarly, approaches like CellPLM (Wen et al. 2023), Nicheformer (Schaar et al. 2024), and NicheCompass (Birk et al. 2024) focus on cell–cell interactions at the level of cell types, leaving gene-level communication largely unexplored. These limitations underscore the need for a method that can capture gene regulatory interactions across cells in their native multicellular context across diverse environments and conditions.

To address this, we introduce the Multicellular Transformer (MCT), a novel single-cell foundation model that learns dynamic, context-aware representations by modeling gene expression across ensembles of cells rather than within individual cells. Unlike previous models, MCT leverages both local cell context and global tissue information, offering a unified framework for uncovering distributed gene regulatory programs. Pretrained on a vast dataset of over 14 million cells from 3,500 donors spanning diverse tissues and diseases, MCT processes entire samples—comprising thousands of cells—enabling a holistic analysis of multicellular interactions.

We demonstrate MCT’s broad utility through rigorous validation on three key tasks: (1) *de novo receptor-ligand pair identification* where MCT not only outperforms state-of-the-art methods but also reveals disease-specific receptor-ligand interactions; (2) *cell–cell communication gene set recovery* where MCT uncovers signaling pathways, including those associated with tumor progression, by recovering gene sets involved in intercellular communication; and (3) *spatial gene–gene co-localization reconstruction* where MCT successfully infers spatial co-localization patterns from non-spatial scRNA-seq data, highlighting its versatility in capturing multicellular organization. These results demonstrate MCT’s ability to generate meaningful gene representations, excel in zero-shot settings, and adapt effectively with fine-tuning.

Building on these capabilities, we further showcase two novel applications that underscore MCT’s broad applicability: (1) population-level perturbation analysis, enabling insights into tissue-wide effects of genetic interventions; and (2) biomarker identification, which leverages a neighborhood-based approach in the gene-embedding space to identify genes influenced by specific conditions, such as tissue type or disease state.

In summary, MCT represents a transformative advance in modeling multicellular gene regulation. By providing unprecedented insights into gene-level intercellular interactions and supporting a range of applications from clinical diagnostics to therapeutic development, our approach opens new avenues for understanding and manipulating complex biological systems.

2. Results

2.1. Learning gene co-regulation across cells with Multicellular Transformer

Cells within the same tissue environment communicate and coordinate their activities by expressing genes that produce signaling molecules and receptors. This intercellular signaling is highly context-dependent, and understanding its function often requires examining the population of cells in the environment rather than studying individual cells in isolation. We propose that by learning gene representations that consider a gene's expression across multiple cells, we can better decipher gene co-regulation in multicellular environments. To this end, we developed the Multicellular Transformer (MCT), a model that takes as input thousands of cells from a sample along with their gene expression data, and models contextual gene-gene interactions by considering gene expression across different cells rather than focusing on a single cell.

Multicellular Transformer takes as input a set of single-cell profiles, which may originate from scRNA-seq or spatial transcriptomics data, and come from the same sample (i.e., same donor and technical batch). These single-cell profiles are represented as a cell-by-gene matrix $X \in \mathbb{R}^{M \times N}$, where M and N correspond to the number of cells and genes, respectively. Multicellular Transformer treats each gene as a token and uses its expression across different cell types as input to the model (Fig. 1a). This approach is based on the understanding that a cell's gene expression is partially influenced by intercellular signaling from other cells in its surrounding environment. Each gene is represented by an embedding learned by mapping its expression across the cells in the sample into a latent space. We then combine the gene expression latent vector with a gene function embedding, which is independent of expression and represents the inherent function of the gene, consistent across all samples. The combined gene embedding is then fed into a transformer, which computes contextual gene embeddings based on the embeddings of other genes present in the sample.

The rationale for combining gene expression and gene function embeddings is that the expression of a gene captures its context-dependent activity, while the gene function embedding represents the gene's inherent function that remains the same across samples. Without the gene expression embedding, we could not represent the activity of the gene in a sample, which is crucial for modeling the dynamic behavior of genes in specific cellular contexts. Conversely, without the gene function embedding, the model would lack essential information about the role of the gene, as genes with the same level of expression in a sample would have exactly the same representation.

The gene expression latent vector is computed using a two-layer feed-forward neural network (FFNN), which is shared across all genes to prevent overfitting. Importantly, the input units of this neural network are fixed to specific cell types during pretraining, based on the cell type composition across the entire dataset (Methods). For example, input units 1–20 are assigned to *CD8 T cells*, while units 21–32 are assigned to *Macrophages*. This design allows us to account for cell-type-specific gene expression and helps prevent the model from simply learning average gene expression patterns regardless of cell type. The gene function embedding, independent of context, projects the gene name into a multidimensional vector representing the gene's inherent function, independent of its sample-specific expression. Both the gene function embeddings and the gene expression FFNN are learned from scratch during pretraining, leading to gene representations where co-regulated genes are close in the embedding space. The gene expression and gene function embeddings are combined by summing them together.

For pretraining, we compiled a corpus of over 14 million cells from more than 3,500 donors across various

tissues and diseases, harmonizing cell types between datasets (Supp. Fig. S1). The transformer encoder consists of 12 layers, and the total number of unique genes is nearly 60,000. Notably, the context size in MCT is significantly larger than in single-cell foundation models such as scGPT (Cui et al. 2024) or GeneFormer (Theodoris et al. 2023), because we consider all genes expressed in a sample, rather than only those expressed in a single cell. We demonstrate that this increased context leads to significant improvements in downstream task performance (Fig. 2b).

The Multicellular Transformer learns gene co-regulation across cells by predicting masked gene identities based on the genes' expressions and their context, encouraging the model to learn meaningful gene-gene relationships (Fig. 1d). To further promote the learning of meaningful latent representations of gene expression, we combined the masked learning objective with a reconstruction loss, which is optimized to reconstruct the expression of a gene from its contextual gene embedding. The gene expression reconstruction serves as a secondary loss to the masked gene function prediction and is computed by projecting the contextual gene embeddings (the output of the transformer's last layer) to a vector equal in length to the number of cells used as input, representing the reconstructed raw expression. During pretraining, we randomly mask 15% of the genes (Methods).

In the next sections we describe how we validated that MCT captures intercellular interactions by evaluating it across three tasks on unsupervised receptor-ligand prediction, learning gene sets involved in cell-cell communication and recovering spatial co-localization patterns between genes. Finally, we demonstrate MCT's novel capabilities by showing how one can use it to identify disease- and tissue-specific genes as well as perform *in silico* perturbation effect prediction at cell-population, rather than cell-line level.

2.2. Multicellular Transformer uncovers gene interactions involved in cell-cell communication

Identifying receptor-ligand interactions. We firstly evaluated whether Multicellular Transformer (MCT) learned the relationship between known receptor-ligand genes—key components of cellular signaling networks. Identifying receptor-ligand relationships from scRNA-seq or spatial profiling data is challenging due to their often low expression levels and high context dependency. Given MCT's inherent ability to capture multicellular context and model non-linear interactions, we tested whether it could identify receptor-ligand interactions in an unsupervised manner. We measured the relationships between receptor and ligand pairs by computing pairwise cosine similarity between pretrained gene embeddings. These results were evaluated against a consensus set of known receptor-ligand pairs from Liana (Dimitrov et al. 2024) (Methods). MCT outperformed alternative methods significantly, achieving an area under the receiver operating characteristic curve (AUROC) of 0.68, compared to 0.59 by the second-best method, CellPLM (Wen et al. 2023) (Fig. 2a). Intercellular co-expression achieved only 0.52 AUROC, highlighting the limitations of relying solely on scRNA-seq co-expression for receptor-ligand identification. Moreover, MCT demonstrated robustness by achieving highest AUROC across receptors specific to various cell types (Supp. Fig. S2a-b).

We hypothesized that the high performance of MCT is due to model considering the entire multicellular environment, rather than expression of two genes of interest. We evaluated the impact of the environment by varying the model context size, which refers to the number of genes included in the input (Methods). We observed that AUROC increased significantly with larger context lengths, emphasizing the importance of considering the multicellular context for receptor-ligand identification (Fig. 2b).

Identifying additional genes across distinct cell types involved in receptor-ligand interactions is crucial,

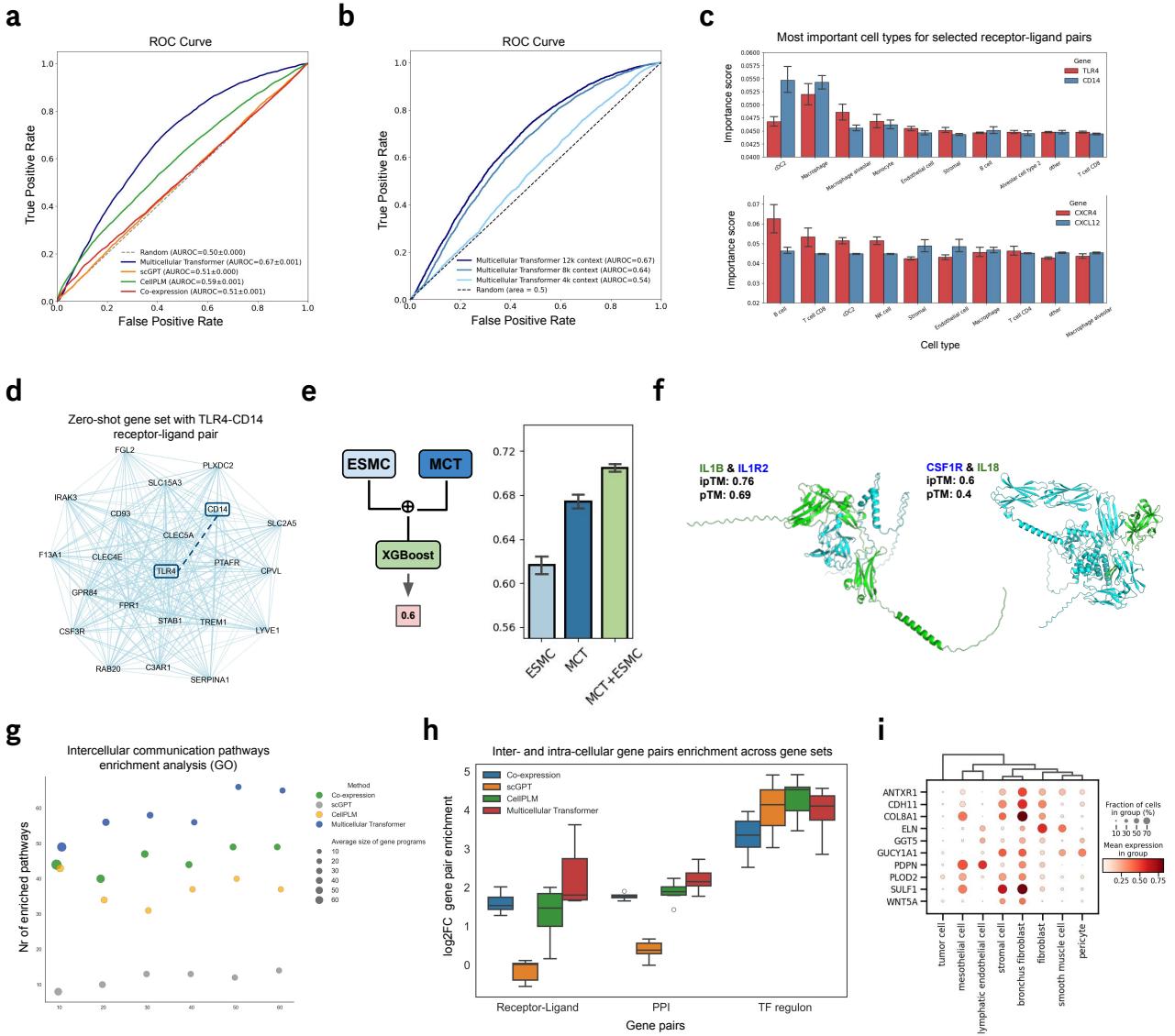


FIGURE 2. Multicellular Transformers identifies genes driving multicellular interactions. (a) Performance on the receptor-ligand prediction task using pretrained gene embeddings. Co-expression was computed using normalized expression aggregated at sample-level using training corpus. The true receptor-ligand pairs were taken from Liana consensus receptor-ligand set (Dimitrov et al. 2024), while the negatives were randomly sampled (Methods). (b) Comparison of Multicellular Transformer model pretrained with various context size (i.e. number of genes). Same dataset was used as in (a). (c) Most important cell types for TLR-CD14 and CXCR4-CXCL12 receptor-ligand pairs using integrated gradients (Sundararajan, Taly, and Yan 2017) on lung cancer atlas (Salcher et al. 2022). Cell type importance scores have been computed by applying softmax function to aggregated absolute attribution scores across samples. The error bars represent 95% confidence interval. cDC stands for conventional dendritic cells. (d) Example of a zero-shot gene set with a receptor (CD14) - ligand (TLR4) pair identified with Multicellular Transformer. (e) Combining transcriptomic and protein information from Multicellular Transformer (MCT) and ESMC protein language model leads to improved performance on receptor-ligand prediction task. Scores for each receptor-ligand pair were computed by calculating cosine similarity between embeddings coming from each model. The scores were then used as input to the XGBoost model which was trained and evaluated using 5-fold cross validation on the Liana consensus set (Methods). (f) Structures of protein complexes predicted with AlphaFold3 (Abramson et al. 2024) and predicted to interact with Multicellular Transformer. The pTM score measure how well different parts of protein fold together, while ipTM assesses the confidence of protein-protein interfaces, higher is better. (g) Enrichment of zero-shot gene sets extracted using various methods in cell-cell communication related pathways across Leiden resolutions. The genes selected for the analysis were highly variable genes from the tumor microenvironment atlas (Guimarães et al. 2024). The co-expression was computed using the aggregated gene expression at the sample-level using the entire pretraining corpus. (h) Enrichment of receptor-ligand, PPI and TF regulon (TF-target gene) gene pairs across Leiden resolutions using various methods. The receptor-ligand pair enrichment measures how well does the model capture the intercellular interactions, TF regulon the intracellular interactions and PPI both, with the majority of its gene pairs representing intracellular interactions. The enrichment was calculated by counting the number of gene pairs present in the gene set and normalizing for the expected number of gene pairs given the size of the gene set. The box represents the interquartile range (IQR), horizontal like marks the median, and whiskers extend to farthest data points within 1.5 times the IQR. (i) Normalized expression of a selected gene set involved in Wnt signaling in lung cancer across a set of cell types. For brevity we have limited cell types to the ones highly expressed in the gene set.

as these genes often encode components that regulate or facilitate signaling pathways. Understanding these genes can reveal novel therapeutic targets for diseases caused by dysregulated cell communication. Using gene sets extracted from MCT’s pretrained embeddings, we identified groups containing known receptor-ligand pairs and analyzed the other genes present. MCT discovered several gene sets with receptor-ligand pairs, such as *TLR4-CD14* (Fig. 2d), essential for initiating immune responses against bacterial infections (Ciesielska, Matyjek, and Kwiatkowska 2021), and *WNT5A-FZD1*, part of the crucial Wnt signaling pathway (Ring et al. 2014) (Supp. Fig. S3a). To determine which cell types are particularly important for these interactions, we employed integrated gradients (Sundararajan, Taly, and Yan 2017), attributing the model’s predictions to its input features (Fig. 2c). For the *TLR4-CD14* pair, MCT identified dendritic cells (cDC2) and monocytes as the cell types with the highest attribution scores. These cells are pivotal in initiating immune responses against bacterial pathogens, where *TLR4-CD14* plays a critical role as mentioned above. For the *CXCR4-CXCL12* pair, MCT highlighted B and T cells, as well as stromal and endothelial cells—cell types involved in inflammation and angiogenesis, respectively (Liekens, Schols, and Hatse 2010). Both processes are known to involve this receptor-ligand pair. Additionally, MCT identified regulatory T cells as most significant for the *IL2RA-IL2R* interaction (Cheng, Yu, and Malek 2011) and tumor cells as top-scoring for the *MET-HGF* pair (Cecchi, Rabe, and Bottaro 2010) (Supp. Fig. S3b), further validating our findings.

Integrating multicellular and protein-level information. MCT predicts receptor-ligand interactions purely from scRNA-seq data, however, the receptor-ligand interaction is ultimately mediated by the physical binding of the proteins. To incorporate protein-level information, we combined a state-of-the-art protein language model, ESMC (ESM Team 2024) with MCT’s gene embeddings in a boosting framework (Fig. 2e). By jointly modeling expression context and protein-level interactions, we achieved better precision in identifying receptor-ligand pairs.

Many human cell surface receptors are orphan, meaning their ligands are unknown, and running protein-protein binding simulations is very time consuming. Therefore, we investigated whether MCT could help identify gene pairs that form protein complexes. Focusing on cytokines as key mediators in immune responses (Belardelli 1995), we utilized data from a lupus study (Perez et al. 2022). We computed cosine similarity between contextual gene embeddings and extracted top-scoring pairs. After filtering for pairs of secreted and membrane proteins, we employed AlphaFold3 (Abramson et al. 2024) to compute binding scores (ipTM) for the top 70 candidates. Remarkably, out of 70 pairs, 6 showed binding scores greater than 0.5 (mean = 0.26 ± 0.13), indicating moderate binding confidence, despite the model only having access to expression data. MCT recovered known interactions such as *IL1B* and *IL1R2* (ipTM: 0.76), which are established binding partners (Colotta et al. 1993) (Fig. 2f). Additionally, top-scoring pairs included *CSF1R* and *IL18* (ipTM: 0.6), known to interact (Ma et al. 2023), among others (Supp. Fig. S4a). These results suggest that MCT can aid in predicting potential receptor-ligand interactions. However, to validate novel pairs, further experimental studies, including structural characterization, are essential.

Capturing intercellular communication pathways. Having validated that MCT learns the relationships between receptors and ligands, we sought to further study the co-regulatory interactions learned by the model during pretraining. Using the pretrained gene embeddings from MCT, we extracted co-regulated genes based on their embedding similarities and clustered the genes into distinct, mutually exclusive gene sets (Methods). These gene sets consist of genes with similar latent representations, implying functional relationships in shared biological processes.

MCT recovered several gene sets that play key roles in signaling pathways (Supp. Fig. S5a), including cytokine-related pathways such as type I interferon and chemokine-mediated signaling, as well as T-cell receptor signaling. To demonstrate that MCT effectively models the multicellular context, we performed enrichment analysis (Fang, Liu, and Peltz 2023) on the gene sets discovered by MCT and compared them to those identified by other methods, focusing specifically on pathways related to cell-cell communication (CCC). MCT showed significantly higher enrichment in CCC-related pathways across Leiden resolutions (Fig. 2g & Supp. Fig. S6a), while maintaining smaller gene set sizes compared to the second-best method, co-expression analysis (Supp. Fig. S6c). This demonstrates that MCT can capture true interactions with high specificity without requiring fine-tuning. We also investigated the overlap between gene sets identified by co-expression, CellPLM, and MCT. All three methods shared 12 Gene Ontology (GO) pathways (Ebert et al. 2023), while MCT uniquely discovered an additional 23 pathways (Supp. Fig. S6b).

Interactions between different cell types are crucial in health and disease and key for cell-cell communication, so identifying gene sets that modulate these interactions is particularly important. We examined the expression of gene sets identified by MCT across cell types in the lung cancer atlas (Salcher et al. 2022) (Supp. Fig. S6d). The gene sets were expressed across diverse cell types, validating that our approach captures relationships across distinct cell states.

MCT is trained to predict masked genes based on the expression of other genes in the sample, allowing the model to learn both intra- and inter-cellular interactions. To assess the extent to which MCT captures these interactions, we evaluated whether the discovered gene sets are enriched in gene pairs related to intercellular and intracellular processes.

We extracted three sets of gene pairs: (1) receptor-ligand pairs (intercellular interactions) using LIANA (Dimitrov et al. 2024). (2) TF-target gene pairs (intracellular interactions) using CollecTRI (Müller-Dott et al. 2023). (3) protein-protein interaction (PPI) Pairs (both inter- and intra-cellular interactions) using OmniPath (Türei, Korcsmáros, and Saez-Rodriguez 2016). MCT showed the highest enrichment in receptor-ligand pairs across Leiden resolutions (Fig. 2h). Additionally, MCT performed best in capturing PPI pairs and performed similarly to scGPT in capturing TF-target gene pairs, although it underperformed compared to CellPLM in the latter. These results indicate that MCT effectively captures gene interactions both across cells and within individual cells. The strong performance of CellPLM in intracellular interactions suggests that its pretraining objective emphasizes learning intracellular mechanisms compared to other methods, highlighting the advantage of MCT's approach in learning gene-gene interactions across cells.

Revealing interactions promoting tumor progression in lung cancer. Having validated that MCT recovers known biological interactions, we explored gene sets identified by MCT to gain novel insights into multicellular interactions. We focused on a gene set extracted from the lung cancer atlas (Salcher et al. 2022). Expression analysis across cell types showed enrichment in tumor cells, fibroblasts, lymphatic endothelial cells, and others (Fig. 2i). Notably, several genes in this set are associated with epithelial-mesenchymal transition (EMT), a critical process in cancer metastasis. Genes like *CDH11*, *PDPN*, *WNT5A*, and *PLOD2* influence cell adhesion, motility, and extracellular matrix (ECM) remodeling, promoting EMT and tumor progression (Kiener et al. 2006; Wicki and Christofori 2007; Dissanayake et al. 2007). Specifically, the expression patterns of these genes suggest that the gene set contributes to: (1) extracellular matrix (ECM) remodeling: genes like *PLOD2*, *COL8A1*, and *ELN* may collaborate to strengthen the ECM, facilitating cancer cell invasion (Hong et al. 2024). (2) Cell adhesion and motility: *CDH11* and *WNT5A* influence cell adhesion and motility,

promoting tumor cell migration (Kiener et al. 2006; Dissanayake et al. 2007). (3) Modulation of signaling pathways: *SULF1* modifies heparan sulfate chains, regulating the availability of signaling molecules like *WNT5A*, impacting cell-cell communication (Lai et al. 2008). (4) Angiogenesis and lymphangiogenesis: *PDPN* and *ANTXR1* are involved in forming new blood and lymphatic vessels that supply the tumor (Wicki and Christofori 2007; Carson-Walter et al. 2001).

The presence of these genes in cell types such as mesothelial cells, fibroblasts, pericytes, and endothelial cells indicates a coordinated effort among different cells to remodel the ECM and enhance signaling processes that support tumor growth, EMT, and metastasis. This gene set illustrates how lung cancer leverages interactions between genes across various cell types to promote tumor progression through EMT, demonstrating the potential of MCT in revealing complex multicellular interactions in disease contexts.

2.3. Multicellular Transformer discovers disease-specific genes with context-aware gene embeddings

The Multicellular Transformer (MCT) uncovers subtle, condition-specific gene signals by embedding each gene in a tissue- and disease-aware space. Rather than relying on expression within a single cell, MCT captures gene-gene relationships that change with the local microenvironment—paving way for the discovery of biomarkers that are context-dependant and which standard differential expression often misses. To systematically measure these context-specific shifts, we computed each gene’s top k nearest gene neighbors in different samples across conditions and then quantified their overlap via the Jaccard similarity (Methods; Supp. Fig. S7). A low Jaccard index across conditions combined with high within a condition, such as a disease indicates that the gene’s network “rewires” significantly across conditions, flagging it as highly condition-specific. For instance, interferon-stimulated genes (ISGs) like *IFI44* or *OAS1* cluster with interferon pathway members in lupus or severe COVID-19, yet occupy more generic neighborhoods in healthy samples. Even if these ISGs display modest expression changes, MCT detects their contextual transition by observing the swap in embedding neighbors (Fig. 3a). We computed neighborhood shift across genes in SLE (Perez et al. 2022), COVID-19 (Stephenson et al. 2021) and multi-tissue dataset (Suo et al. 2022), showing consistent neighborhood specificity for a range of disease- and tissue-specific genes (Supp. Fig. S8).

Beyond individual genes, we also clustered sample-specific embeddings from MCT’s last hidden layer to form context-aware gene sets—effectively capturing larger-scale pathway signatures that differ by tissue or disease. For instance, clusters derived from glioblastoma samples were enriched for pathways involving *NFKB* (Fig. 3b), *p53* and stress responses, reflecting processes central to tumor progression. Meanwhile, healthy peripheral blood mononuclear cell (PBMC) samples showed enrichment for normal immune function pathways (e.g., B-cell differentiation, T-cell receptor signaling) (Supp. Fig. S9). This context-awareness not only validates MCT’s capacity to discover disease-oriented gene sets, but also underscores its usefulness in revealing novel gene subsets that may otherwise remain hidden.

Identifying disease-Specific Genes in SLE and COVID-19. Applying this framework to systemic lupus erythematosus (SLE) and COVID-19 highlights how MCT pinpoints condition-specific genes that play pivotal roles despite modest changes in expression. In lupus (Perez et al. 2022), *IFI44* and *MX1* (Fig. 3c) emerged as robustly SLE-specific genes—consistent with their known associations to interferon signaling (Rodríguez-Carrio et al. 2019; AlFadhli et al. 2016)—while *OAS1* and *CD74* were identified as COVID-19-specific (Stephenson

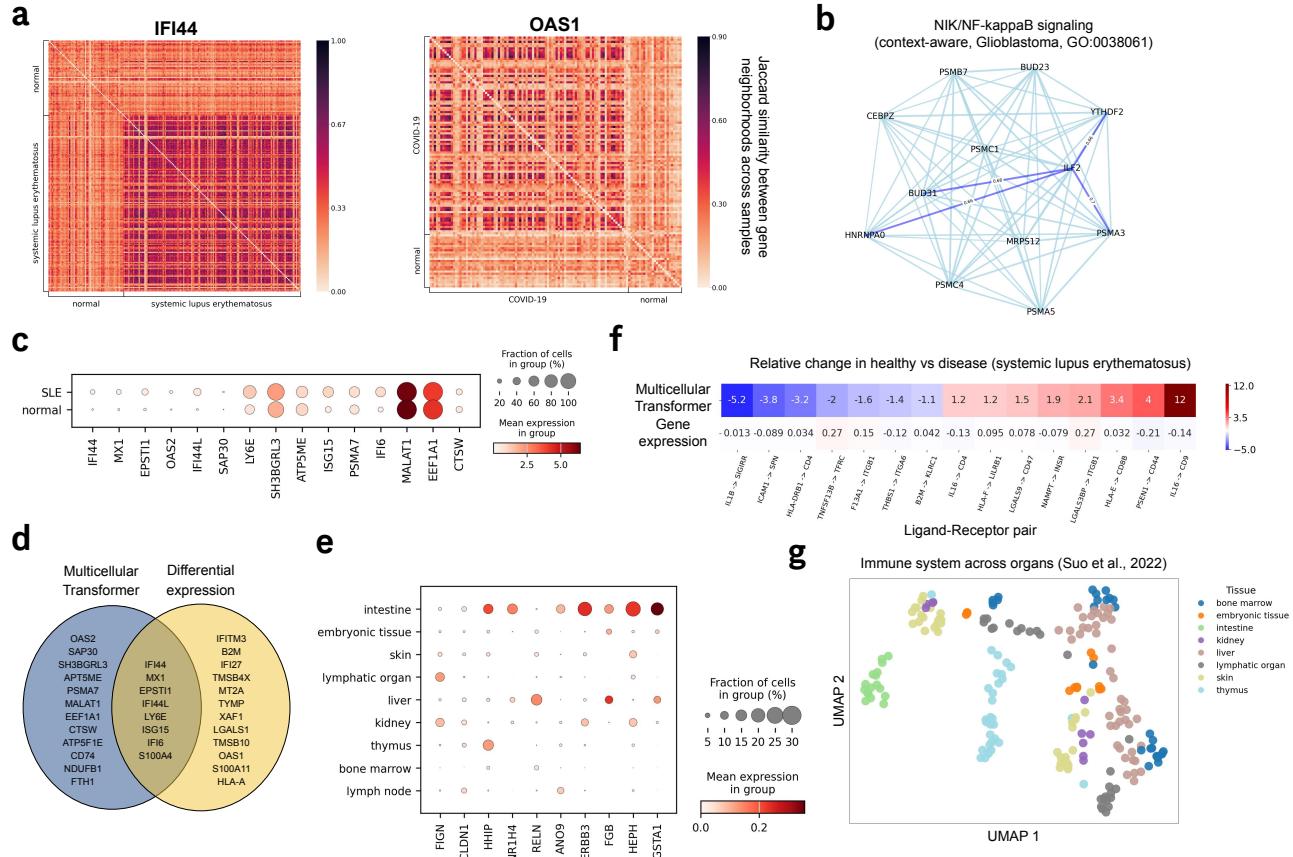


FIGURE 3. Context-aware gene embeddings from Multicellular Transformer identify disease- and tissue-specific genes. (a) Genes with high condition-specificity across SLE (Perez et al. 2022) & COVID-19 (Stephenson et al. 2021). The condition-specificity was calculated by dividing the mean overlap across distinct samples between gene neighborhoods in the same condition by the mean overlap between gene neighborhoods across conditions. (b) Example of a context-aware gene set enriched in *NF-κB* pathway identified at Leiden resolution of 50 with Multicellular Transformer in a glioblastoma sample (Ruiz-Moreno et al. 2022). (c) The expression of disease-specific genes in SLE (Perez et al. 2022) computed using context-aware embeddings from Multicellular Transformer. Many genes have been shown to be associated with the disease but do not exhibit a significant change in expression between healthy and disease. (d) The overlap between top 20 genes discovered by the Multicellular Transformer and differential expression (Wilcoxon). Overlap = 40%, $N = 20$. (e) The expression of highly tissue-specific genes according to the gene neighborhood across samples computed with MCT contextual gene embeddings on the immune across organs dataset (Suo et al. 2022) across different tissues. (f) Relative change in cosine similarity (top) vs expression (bottom) in healthy vs disease patients with SLE across a set of known receptor-ligand pairs with highest change according to cosine similarity between MCT contextual gene embeddings. Receptor-ligand pairs with highest change according to expression can be found in Supp. Fig. S10b. (g) UMAPs of sample-level embeddings across tissues on the immune across organs dataset (Suo et al. 2022). The sample embeddings were extracted by averaging out all of the contextual gene embeddings in the sample using MCT.

et al. 2021) genes (Supp. Fig. S10a) (Banday et al. 2022; Westmeier et al. 2023).

Comparing MCT's top 20 condition-specific genes to those found by DEX in an SLE dataset (Perez et al. 2022) shows a notable (40%) overlap, confirming that MCT recovers many established biomarkers (Fig. 3d). Crucially, it also uncovers additional candidates (e.g., *OAS2*, *FTH1*, *SH3BGRL3*) by leveraging gene-gene relationships to amplify weak signals—an approach that reveals gene network shifts even when overall expression changes are minimal. For instance, *SH3BGRL3* associates with T cell activation genes in lupus patients but not in healthy controls, indicating an immune function overlooked by standard analyses. Similarly, *OAS2* exhibits little expression difference yet is recognized by its shifted neighborhood of IFN-inducible genes. By identifying these subtle but consequential changes, MCT expands the repertoire of potential biomarkers and therapeutic targets beyond those detectable via purely expression-based pipelines.

MCT's context-awareness extends naturally to tissue specificity. Genes such as *FIGN*, *CLDN1*, and *GSTA1* from the immune system across organs dataset (Suo et al. 2022) exhibit strong tissue-restricted expression patterns—validated by MCT's ability to keep these genes clustered together in tissues where they are functionally relevant (Fig. 3e). Indeed, many genes that drive inflammation or cell differentiation can have drastically different roles in lung, blood, or brain tissue (Lavin et al. 2014). By incorporating whole-sample context, MCT captures these often subtle, tissue-bound gene functions without requiring considerable expression signatures.

Revealing receptor-ligand interactions in systemic lupus erythematosus. Many disease processes emerge from coordinated actions among multiple genes. For example receptor-ligand pairs whose binding events can shift an entire signaling landscape. To capture these multi-gene interactions, we leveraged MCT to evaluate changes in receptor-ligand similarity between healthy and disease states in SLE.

Traditional scRNA-seq methods may miss lowly expressed or context-dependent interactions, yet MCT uncovered receptor-ligand pairs that shift significantly between healthy and SLE states—even when their expression remains stable (Fig. 3f & Supp. Fig. S10b). For instance, pairs such as *IL16-CD9* and *IL1B-SIGIRR* showed notable changes consistent with lupus pathology (Lee et al. 1998; Jeffries et al. 2011; Song et al. 2014; Wang et al. 2013). MCT also recovered established lupus interactions (e.g., *TNFSF13B-TNFRSF13C*) and grouped *CD9* with genes like *OAS1*, *TNFSF10*, and *WARS1* (Supp. Fig. S10c), linking them to immune regulation, inflammation, and apoptosis (Dinarello 2000; Postal and Appenzeller 2011; Gao, Tan, and Luo 2020; Kurabayashi et al. 2008; Scandura et al. 2004; Weedon 2007; Nguyen et al. 2020). Collectively, these results reveal how inter-cellular communication can drive chronic inflammation and autoimmune responses in SLE.

Finally, we observe that sample-level embeddings obtained by aggregating gene embeddings per sample cluster in a biologically meaningful way. Samples from distinct tissues or disease states separate clearly in the embedding space, reflecting true biological variation rather than technical noise and outperforming expression-based methods (e.g., PCA) on normalized mutual information (NMI) in 3 out of 4 comparisons (Fig. 3g Supp. Fig. S11). This context-sensitive clustering proves robust to donor identity and assay type, indicating that MCT not only pinpoints condition- or tissue-specific genes but also yields globally interpretable sample-level structures.

2.4. Multicellular Transformer allows for *in silico* zero-shot perturbation prediction at cell population level

MCT has shown the capability to capture the contextual interactions between genes in a multicellular environment, therefore, we hypothesized that it can also predict the effect of genetic perturbation on entire cell populations in a zero-shot manner. Understanding how therapeutic interventions impact not only target cells but also other cell types in the tumor microenvironment is crucial for evaluating drug safety. Approximately 30% of drugs fail in clinical trials due to safety issues (Waring et al. 2015), highlighting the need for methods that consider the broader cellular context.

Existing perturbation prediction methods primarily focus on effects at the cell line level (Roohani, Huang, and Leskovec 2024; Lotfollahi et al. 2023; Cui et al. 2024), often neglecting the microenvironment of target cells. In contrast, MCT incorporates the entire multicellular environment and allows for *in silico* genetic perturbations by leveraging gene embeddings. Due to the lack of data on genetic perturbations at the cell population level, we conducted all predictions in a zero-shot manner without fine-tuning the model.

Uncovering gene-gene relationships with *in silico* genetic perturbations. First, we assessed whether MCT could capture gene-gene relationships upon genetic perturbation. We curated a set of gene pairs known to play key roles in cell-cell communication and performed perturbation effect predictions across samples using the model fine-tuned on the lung cancer atlas (Salcher et al. 2022). We simulated gene knockouts *in silico* by removing the target gene from the sample and comparing the results to the baseline model where the gene is present (Methods).

MCT accurately learned dependencies between genes, outperforming a control baseline across gene pairs (Fig. 4a) and achieving high AUROC scores (Supp. Fig. S12a). This demonstrates MCT's ability to predict the impact of gene perturbations on other genes in a multicellular setting, opening avenues for evaluating target specificity and safety.

Predicting the effect of genetic perturbation on cell types. Moving beyond individual gene-gene interactions, we explored MCT's ability to predict the effects of genetic perturbations on different cell types in a population. Current perturbation approaches focus on predicting effects on target cell types, such as tumor cells, often using cell line models. While useful for understanding intracellular mechanisms, these methods do not inherently consider other cells in the environment. MCT's input includes cells from various cell types, enabling us to predict the impact of gene perturbations on an entire cell population.

To quantify the impact of perturbing a gene at the population level, we used the reconstructed gene expression and compared the predicted normalized expression aggregated at the cell type level, with and without the gene knockout (Methods). We evaluated our approach using data from the DepMap consortium (Tsherniak et al. 2017), which provides viability readouts from CRISPR gene knockouts across various cell lines, indicating the effect on tumor cell populations.

Using the model fine-tuned on the lung cancer atlas, we performed *in silico* gene knockouts across lung cancer samples (Methods). We retrieved CRISPR gene knockout viability readouts from relevant lung cancer cell lines in DepMap. Comparing the genes with the highest predicted effects (i.e., those reducing the tumor population, Fig. 4b), MCT showed a considerable overlap (30%) with DepMap data (Supp. Fig. S12b). To ensure the effect was not solely due to removing the expression of the knocked-out gene, we assessed the overlap between MCT's top-effect genes and the genes with highest expression in tumors, finding a 50%

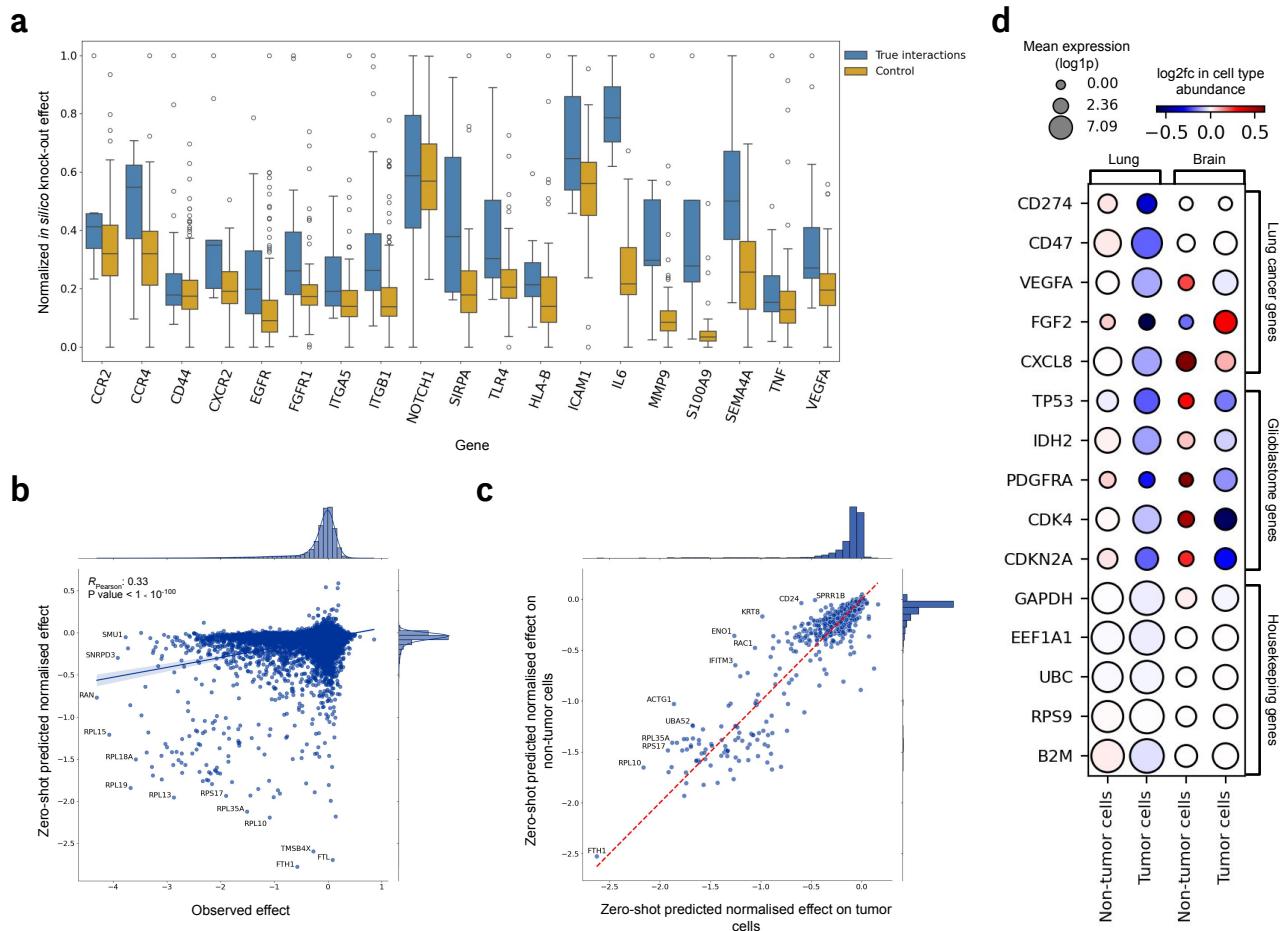


FIGURE 4. *In silico* genetic perturbation effect estimation at cell population level. (a) *In silico* gene knock-out effect on a set of true interacting genes and control genes. The control genes were randomly selected (Methods). The knock-out effect is quantified by the maximum log-fold change in gene's logits (Methods). The predictions were min max normalized within a gene. The box represents the interquartile range (IQR), horizontal like marks the median, and whiskers extend to farthest data points within 1.5 times the IQR. (b) The zero-shot predicted normalized effect on tumor expression using Multicellular Transformer vs the observed effect on tumor across lung cancer cell lines from DepMap data (Tsherniak et al. 2017). The predicted effect from Multicellular Transformer was computed by calculating the change in expression in the sample between perturbed vs control and normalizing it by control expression. (c) The zero-shot predicted normalized effect on non-tumor cells vs on tumor cells. The predicted effect was calculated same as in (c). (d) The predicted gene knock-out effect on tumor and non-tumor cells on lung cancer (Salcher et al. 2022) and glioblastoma (Ruiz-Moreno et al. 2022) across multiple validation samples on a set of genes specific for lung cancer, glioblastoma and housekeeping genes. The dot size represents the mean expression in the sample, ensuring the gene is expressed in both tumor and non-tumor cells. The dot color signifies the change in relative cell type abundance between the reconstructed expression with vs without the knock-out. The knock-out effect was calculated by zeroing out the gene expression at input and comparing the log2fc in cell type abundance compared to the baseline sample without the gene knock-out. The negative enrichment in reconstructed tumor cell expression compared to non-tumor cells shows how MCT captures tissue-specific tumor-selective dependencies.

overlap (Supp. Fig. S12c). While substantial, this is expected given that predictions were made in a zero-shot fashion. We further examined the pairwise correlation between MCT predictions and DepMap data (Fig. S4b). For the *SW 1573* cell line, the Pearson correlation between the two was 0.33 (p -value $< 1 \times 10^{-10}$). Notably, many of the genes with high effects in both MCT and DepMap were ribosomal proteins (Fig. 4b), which is unsurprising since knocking out ribosomal proteins can lead to cell death across cell types. To identify targets that selectively affect tumor cells, we calculated the predicted effect on non-tumor cells and compared it to the effect on tumor cells (Methods). This allowed us to find genes that have a high impact on tumor cells while having a low effect on non-tumor cells (Fig. 4c), potentially indicating safer therapeutic targets. The genes identified by this approach were significantly less enriched in ribosomal proteins.

A crucial step in identifying promising therapeutic targets is selecting an experimental model that is physiologically relevant and translatable to real-world systems. Since DepMap includes hundreds of cell lines, we identified those with the highest correlation to MCT's predicted effects by calculating the Pearson correlation between MCT's predictions and the viability readouts. This can help select cell lines that best mimic the real-world settings modeled by MCT. The top cell lines for human lung cancer are shown in Fig. (Fig. S12d).

Having validated that MCT can predict the effects of genetic perturbations in cell lines, we next performed an *in silico* knock-out of disease-specific genes in lung cancer (Salcher et al. 2022) and glioblastoma (Ruiz-Moreno et al. 2022) across patient samples. We selected a set of representative genes for lung cancer and glioblastoma, ensuring the genes are expressed in both tumor and non-tumor cells. We then compared the resulting changes in relative cell type abundance between tumor and non-tumor populations. We also tested a set of housekeeping genes, which are stably expressed across tissues and generally not implicated in driving tumorigenesis. The knock-out effects predicted by MCT consistently reduced tumor populations for lung cancer- and glioblastoma-specific genes, while having a lesser impact on non-tumor cells and other tissue types (Fig. 4d). We have also noticed that tumor-specific genes had an effect across both brain and lung tissues, demonstrating shared oncogenic dependencies that transcend tissue boundaries. However, the knock-out effect was more consistent for the tumor genes specific to the tissue. Finally, knocking out housekeeping genes had minimal impact on relative cell type abundance despite their high expression, affecting all non-tumor and tumor cells roughly equally. Thus, demonstrating how MCT can leverage the contextual tissue context to distinguish tissue-specific genes crucial basic cellular function (i.e. housekeeping genes) from those driving malignancy.

2.5. Multicellular Transformer recovers spatial co-localization patterns from scRNA-seq data

To investigate whether the gene-gene interactions learned by MCT using scRNA-seq data only recover the spatial co-localization patterns, we analyzed whether the gene-gene similarities extracted from MCT's embeddings correspond to spatial co-localization patterns observed in spatial transcriptomics data.

To do this, we evaluated whether the pretrained gene function embeddings which are close in the embedding space also show high spatial co-localization. We extracted gene pairs with high cosine similarity from the MCT embeddings in the lung cancer atlas (Salcher et al. 2022) and a glioblastoma study (Ruiz-Moreno et al. 2022). We then visualized these gene pairs using corresponding spatial transcriptomics datasets—Xenium lung cancer and Visium glioblastoma—which MCT was not trained on, making this a

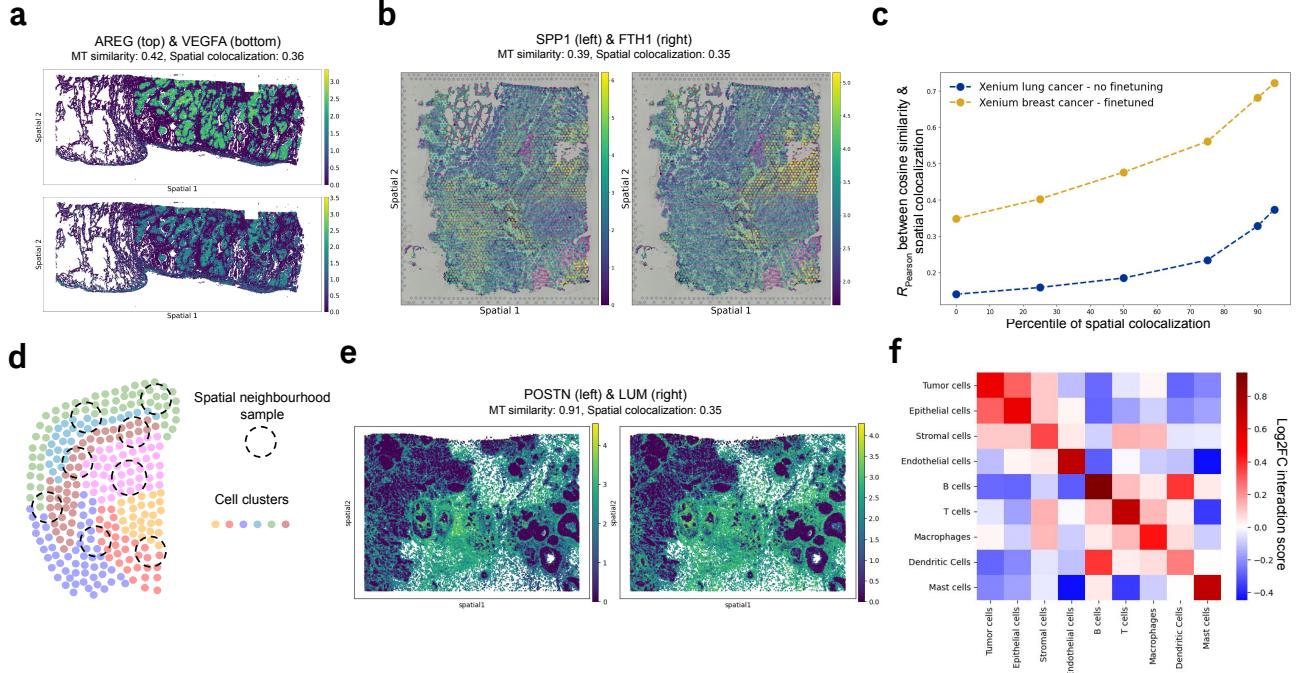


FIGURE 5. Spatial patterns across datasets recovered by the Multicellular Transformer **(a)** Spatial co-localization of *AREG* & *VEGFA* genes in Xenium human lung cancer dataset, recovered by the Multicellular Transformer in a zero-shot fashion by assessing genes with highest cosine similarity in gene embeddings for the *AREG* gene. The spatial co-localization was measured by computing bivariate Moran's I between two genes. **(b)** Spatial co-localization of *SPP1* & *FTH1* genes in Visium glioblastoma dataset, recovered by the Multicellular Transformer in a zero-shot fashion by assessing genes with highest cosine similarity in gene embeddings for the *SPP1* gene. The spatial co-localization was measured by computing bivariate Moran's I between two genes. **(c)** The correlation between spatial co-localization and the cosine similarity computed using Multicellular's Transformer gene embeddings across different percentiles of absolute spatial co-localization (the higher the percentile, the higher the absolute spatial co-localization). The correlation was computed by computing spatial co-localization using bivariate Moran's I, and cosine similarity using gene embeddings from Multicellular Transformer. **(d)** A visualization showing spatial samples used for fine-tuning the Multicellular Transformer model on spatial transcriptomics data. A spatial sample is defined by taking the cells within a predefined radius of a cell of interest. **(e)** Spatial co-localization of *POSTN* & *LUM* genes in Xenium breast cancer dataset, recovered by the Multicellular Transformer fine-tuned on the dataset by assessing genes with highest cosine similarity in the gene embedding space. **(f)** Interaction matrix between cell types computed using the cosine similarity between Multicellular Transformer's gene embeddings of marker genes. The Multicellular Transformer model used here was fine-tuned on the Xenium human breast cancer dataset and the marker genes were annotated using the same dataset. The interaction score was computed by calculating the \log_2 fold change between the gene pair's cosine similarity divided by its average cosine similarity across all genes and taking the average across gene pairs for each cell type-cell type pair.

zero-shot evaluation. Our results showed that gene pairs with high cosine similarity in MCT embeddings also exhibited high spatial co-localization, as measured by bivariate Moran's I (Methods), when visualized on spatial profiling slides (Fig. 5a-b and Supp. Fig. S13a-b). Moreover, the Pearson correlation between MCT's cosine similarity and spatial co-localization increased when focusing on gene pairs within higher spatial co-localization percentiles (Fig. 5c). This indicates that MCT can accurately capture highly colocalized gene pairs from scRNA-seq data, demonstrating the model's ability to infer spatial patterns through multicellular contextual embeddings. When investigating the reason behind low correlation at lower percentile values, we discovered that the majority of gene pairs have spatial co-localization close to 0 (Supp. Fig. S13e). This means that for most gene pairs, there is minimal spatial overlap, making it difficult to establish a meaningful correlation between MCT's embedding similarities and spatial co-localization at lower percentiles.

Fine-tuning on spatial transcriptomics data. Interactions between cells are particularly significant in spatial niches, where physically proximal cells are more likely to communicate (Morrison and Spradling 2008). To enhance MCT's ability to capture these spatial interactions, we fine-tuned the model on spatial transcriptomics data by defining samples as sets of cells within a defined radius (Fig. 5d, Methods). We utilized the Xenium human breast cancer dataset to fine-tune the pretrained MCT, using the same training objective as during pretraining.

After fine-tuning, the model showed a significantly higher correlation between MCT's cosine similarity and spatial co-localization (Fig. 5c) and identified highly colocalized gene pairs (Fig. 5e and Supp. Fig. S13c), illustrating the benefit of incorporating spatial information. A key application of spatial data is identifying interactions between different cell types within tissue. We applied the fine-tuned model to identify interactions between marker genes of distinct cell types, such as tumor cells and T cells, in the Xenium breast cancer data. MCT identified several interacting genes across different cell types (Supp. Fig. S13d).

To further analyze cell-type interactions, we aggregated the results by computing cosine similarities between marker genes to construct an interaction matrix. We assumed that higher similarity between marker genes of cell types A and B indicates interaction between those cell types. MCT accurately captured relationships between cell types, showing high interaction scores between immune cells (T cells, B cells, macrophages, dendritic cells) and between tumor, epithelial, and stromal cells (Fig. 5f). In contrast, applying the same approach using gene co-expression resulted in noisy and less interpretable results (Supp. Fig. S13f). This experiment demonstrates how MCT, through its multicellular contextual embeddings, can be effectively applied to spatial data and fine-tuned to enhance performance on specific datasets.

3. Discussion

In this study, we introduced the Multicellular Transformer (MCT), the first large-scale pretrained model that explicitly models gene interactions within a multicellular environment. By considering the context of multiple cells in a tissue, MCT learns meaningful gene embeddings that capture both intercellular and intracellular relationships. By using multiple cells as input, we created context-aware gene embeddings that reflect interactions within the entire cellular environment, rather than focusing solely on individual cells.

We demonstrated the versatility of MCT by showing its effectiveness both in zero-shot applications and after fine-tuning. We validated that the model learns gene co-regulation by using pretrained gene

embeddings to identify receptor-ligand pairs, uncover gene sets involved in cell signaling and recover spatial co-localization patterns between genes, all without additional training. Moreover, we used MCT to develop novel capabilities. We utilized MCT’s highly contextual gene embeddings to reveal tissue- and disease-specific genes based on its neighborhood, discovering genes in systemic lupus erythematosus which were missed using traditional differential expression analysis. Finally we demonstrated how MCT can perform *in silico* genetic perturbations at the cell population level, opening avenues for evaluating the effects of potential treatments on the entire cellular environment rather than just individual cells. We validated these predictions using CRISPR knock-out data from the DepMap consortium (Tsherniak et al. 2017) as well as different tissues and diseases.

An exciting future direction is scaling up the model, particularly by increasing and diversifying the pretraining corpus. While we focused on scRNA-seq and spatial transcriptomics data, MCT can be applied to any type of single-cell data, including multimodal datasets, and bulk RNA-seq data. Moreover, fine-tuning the model on tasks such as perturbation effect prediction at the cell population level offers a promising avenue for assessing the impact of treatments on the whole environment, thereby informing us about the efficacy and safety of interventions. Incorporating protein-level information. Finally, incorporating information on protein-protein interactions which may be modulating intercellular interactions, for example by jointly optimizing the model to learn interactions at transcriptomic and protein-level could better capture the interactions between genes. Overall, by capturing gene relationships in a multicellular context, the Multicellular Transformer is a unique tool that advances our understanding of gene interactions beyond the single-cell level.

4. Code and data availability

We include the code used to perform pretraining and the experiments as part of the submission. We also include a number of tutorials on using Multicellular Transformer.

The details of the datasets used for pretraining can be found in Supp. Table S1. The details of other datasets used in the manuscript can be found in Methods, together with the preprocessing setup. We used the CellArr (<https://github.com/BiocPy/cellarr>) package which allows for storing large collections of genomic data for storing, accessing and sampling from the pretraining corpus.

References

- Abdi, Hervé, and Lynne J Williams. 2010. “Principal component analysis.” *Wiley interdisciplinary reviews: computational statistics* 2 (4): 433–459.
- Abramson, Josh, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick et al. 2024. “Accurate structure prediction of biomolecular interactions with AlphaFold 3.” *Nature*: 1–3.
- Agarap, AF. 2018. “Deep learning using rectified linear units (relu).” *arXiv preprint arXiv:1803.08375*.
- Aibar, Sara, Carmen Bravo González-Blas, Thomas Moerman, VÂN ANH HUYNH-THU, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts et al. 2017. “SCENIC: single-cell regulatory network inference and clustering.” *Nature methods* 14 (11): 1083–1086.
- AlFadhli, Suad, Mashaal Al-Mutairi, Bader Al Tameemi, and Rasheeba Nizam. 2016. “Influence of MX1 promoter rs2071430 G/T polymorphism on susceptibility to systemic lupus erythematosus.” *Clinical rheumatology* 35: 623–629.

- Armingol, Erick, Adam Officer, Olivier Harismendy, and Nathan E. Lewis. 2020. "Deciphering cell-cell interactions and communication from gene expression." *Nature Reviews Genetics* 22 (2): 71–88. <http://dx.doi.org/10.1038/s41576-020-00292-x>.
- Ashburner, Michael, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig et al. 2000. "Gene ontology: tool for the unification of biology." *Nature genetics* 25 (1): 25–29.
- Ba, Jimmy Lei. 2016. "Layer normalization." *arXiv preprint arXiv:1607.06450*.
- Banday, A Rouf, Megan L Stanifer, Oscar Florez-Vargas, Olusegun O Onabajo, Brenen W Papenberg, Muhammad A Zahoor, Lisa Mirabello, Timothy J Ring, Chia-Han Lee, Paul S Albert et al. 2022. "Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19 hospitalization in patients of European and African ancestries." *Nature genetics* 54 (8): 1103–1116.
- Belardelli, FILIPPO. 1995. "Role of interferons and other cytokines in the regulation of the immune response." *Apmis* 103 (1-6): 161–179.
- Birk, Sebastian, Irene Bonafonte-Pardàs, Adib Miraki Feriz, Adam Boxall, Eneritz Agirre, Fani Memi, Anna Maguza, Rong Fan, Gonçalo Castelo-Branco, Omer Ali Bayraktar et al. 2024. "Large-scale characterization of cell niches in spatial atlases using bio-inspired graph learning." *bioRxiv*: 2024-02.
- Cang, Zixuan, Yanxiang Zhao, Axel A Almet, Adam Stabell, Raul Ramos, Maksim V Plikus, Scott X Atwood, and Qing Nie. 2023. "Screening cell-cell communication in spatial transcriptomics via collective optimal transport." *Nature methods* 20 (2): 218–228.
- Carson-Walter, Eleanor B, D Neil Watkins, Akash Nanda, Bert Vogelstein, Kenneth W Kinzler, and Brad St. Croix. 2001. "Cell surface tumor endothelial markers are conserved in mice and humans." *Cancer research* 61 (18): 6649–6655.
- Cecchi, Fabiola, Daniel C Rabe, and Donald P Bottaro. 2010. "Targeting the HGF/Met signalling pathway in cancer." *European journal of cancer* 46 (7): 1260–1270.
- Cheng, Guoyan, Aixin Yu, and Thomas R Malek. 2011. "T-cell tolerance and the multi-functional role of IL-2R signaling in T-regulatory cells." *Immunological reviews* 241 (1): 63–76.
- Ciesielska, Anna, Marta Matyjek, and Katarzyna Kwiatkowska. 2021. "TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling." *Cellular and molecular life sciences* 78: 1233–1261.
- Colotta, Francesco, Fabio Re, Marta Muzio, Riccardo Bertini, Nadia Polentarutti, Marina Sironi, Judith G Giri, Steven K Dower, John E Sims, and Alberto Mantovani. 1993. "Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4." *Science* 261 (5120): 472–475.
- Consortium*, The Tabula Sapiens, Robert C Jones, Jim Karkanias, Mark A Krasnow, Angela Oliveira Pisco, Stephen R Quake, Julia Salzman, Nir Yosef, Bryan Bulthaup, Phillip Brown et al. 2022. "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans." *Science* 376 (6594): eabl4896.
- Cui, Haotian, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. 2024. "scGPT: toward building a foundation model for single-cell multi-omics using generative AI." *Nature Methods*: 1–11.
- Daneshpour, Hirad, and Hyun Youk. 2019. "Modeling cell-cell communication for immune systems across space and time." *Current Opinion in Systems Biology* 18: 44–52.
- Dao, Tri, Dan Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. 2022. "Flashattention: Fast and memory-efficient exact attention with io-awareness." *Advances in Neural Information Processing Systems* 35: 16344–16359.
- Devlin, Jacob. 2018. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805*.
- Dimitrov, Daniel, Philipp Sven Lars Schäfer, Elias Farr, Pablo Rodriguez-Mier, Sebastian Lobentanzer, Pau Badia-i Mompel, Aurelien Dugourd, Jovan Tanevski, Ricardo Omar Ramirez Flores, and Julio Saez-Rodriguez. 2024. "LIANA+ provides an all-in-one framework for cell-cell communication inference." *Nature Cell Biology*: 1–10.
- Dinarello, Charles A. 2000. "Proinflammatory cytokines." *Chest* 118 (2): 503–508.
- Dissanayake, Samudra K, Michael Wade, Carrie E Johnson, Michael P O'Connell, Poloko D Leotlela, Amanda D French, Kavita V Shah, Kyle J Hewitt, Devin T Rosenthal, Fred E Indig et al. 2007. "The Wnt5A/protein kinase C pathway

- mediates motility in melanoma cells via the inhibition of metastasis suppressors and initiation of an epithelial to mesenchymal transition." *Journal of Biological Chemistry* 282 (23): 17259–17271.
- Ebert, Dustin, Marc Feuermann, Pascale Gaudet, Nomi L Harris, David P Hill, Raymond Lee, Huaiyu Mi, Sierra Moxon, Christopher J Mungall, Anushya Muruganugan et al. 2023. "The gene ontology knowledgebase in 2023." *Genetics* 224 (10.1093).
- Elmentait, Rasa, Natsuhiko Kumasaka, Kenny Roberts, Aaron Fleming, Emma Dann, Hamish W King, Vitalii Kleshchevnikov, Monika Dabrowska, Sophie Pritchard, Liam Bolt et al. 2021. "Cells of the human intestinal tract mapped across space and time." *Nature* 597 (7875): 250–255.
- ESM Team. 2024. "ESM Cambrian: Revealing the mysteries of proteins with unsupervised learning." EvolutionaryScale Website. <https://evolutionyscale.ai/blog/esm-cambrian>. Accessed: Month Day, Year (optional).
- Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, Bijay Jassal, Florian Korninger, Bruce May et al. 2018. "The reactome pathway knowledgebase." *Nucleic acids research* 46 (D1): D649–D655.
- Fang, Zhuoqing, Xinyuan Liu, and Gary Peltz. 2023. "GSEApY: a comprehensive package for performing gene set enrichment analysis in Python." *Bioinformatics* 39 (1): btac757.
- Gao, Fei, Yuan Tan, and Hong Luo. 2020. "MALAT1 is involved in type I IFNs-mediated systemic lupus erythematosus by up-regulating OAS2, OAS3, and OASL." *Brazilian Journal of Medical and Biological Research* 53: e9292.
- Guimarães, Gabriela Rapozo, Giovanna Resk Maklouf, Cristiane Esteves Teixeira, Leandro de Oliveira Santos, Nayara Gusmão Tessarollo, Nayara Evelin de Toledo, Alessandra Freitas Serain, Cristóvão Antunes de Lanna, Marco Antônio Pretti, Jéssica Gonçalves Vieira da Cruz et al. 2024. "Single-cell resolution characterization of myeloid-derived cell states with implication in cancer outcome." *Nature Communications* 15 (1): 5694.
- Hao, Minsheng, Jing Gong, Xin Zeng, Chiming Liu, Yucheng Guo, Xingyi Cheng, Taifeng Wang, Jianzhu Ma, Xuegong Zhang, and Le Song. 2024. "Large-scale foundation model on single-cell transcriptomics." *Nature Methods*: 1–11.
- Heimberg, Graham, Tony Kuo, Daryle DePianto, Tobias Heigl, Nathaniel Diamant, Omar Salem, Gabriele Scalia, Tommaso Biancalani, Shannon Turley, Jason Rock et al. 2023. "Scalable querying of human cell atlases via a foundational model reveals commonalities across fibrosis-associated macrophages." *BioRxiv*: 2023–07.
- Hong, Jiwon, Hyo Joon Jin, Mi Ran Choi, Darren Wan-Teck Lim, Jong-Eun Park, You-Sun Kim, and Su Bin Lim. 2024. "Matrisomics: Beyond the extracellular matrix for unveiling tumor microenvironment." *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*: 189178.
- Jeffries, Matlock, Mikhail Dozmorov, Yuhong Tang, Joan T Merrill, Jonathan D Wren, and Amr H Sawalha. 2011. "Genome-wide DNA methylation patterns in CD4+ T cells from patients with systemic lupus erythematosus." *Epigenetics* 6 (5): 593–601.
- Jerby-Arnon, Livnat, and Aviv Regev. 2022. "DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data." *Nature biotechnology* 40 (10): 1467–1477.
- Ji, Andrew L, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhl et al. 2020. "Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma." *Cell* 182 (2): 497–514.
- Kalfon, Jérémie, Jules Samaran, Gabriel Peyré, and Laura Cantini. 2025. "scPRINT: pre-training on 50 million cells allows robust gene network predictions." *Nature Communications* 16 (1): 3607.
- Kamimoto, Kenji, Blerta Stringa, Christy M Hoffmann, Kunal Jindal, Lilianna Solnica-Krezel, and Samantha A Morris. 2023. "Dissecting cell identity via network inference and in silico gene perturbation." *Nature* 614 (7949): 742–751.
- Kiener, Hans P, David M Lee, Sandeep K Agarwal, and Michael B Brenner. 2006. "Cadherin-11 induces rheumatoid arthritis fibroblast-like synoviocytes to form lining layers in vitro." *The American journal of pathology* 168 (5): 1486–1499.
- Kingma, Diederik P. 2014. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*.
- Kock, Kian Hong, Le Min Tan, Kyung Yeon Han, Yoshinari Ando, Damita Jevapatarakul, Ankita Chatterjee, Quy Xiao Xuan Lin, Eliora Violain Buyamin, Radhika Sonthalia, Deepa Rajagopalan et al. 2024. "Single-cell analysis of human diversity in circulating immune cells." *BioRxiv*: 2024–06.

- Kolodziejczyk, Aleksandra A, Jong Kyoung Kim, Valentine Svensson, John C Marioni, and Sarah A Teichmann. 2015. "The technology and biology of single-cell RNA sequencing." *Molecular cell* 58 (4): 610–620.
- Kuribayashi, Kageaki, Gabriel Kriegsfeld, Wenge Wang, Jing Xu, Patrick A Mayes, David T Dicker, Gen Sheng Wu, and Wafik S El-Deiry. 2008. "TNFSF10 (TRAIL), a p53 target gene that mediates p53-dependent cell death." *Cancer biology & therapy* 7 (12): 2034–2038.
- Lai, Jin-Ping, Dalbir S Sandhu, Abdirashid M Shire, and Lewis R Roberts. 2008. "The tumor suppressor function of human sulfatase 1 (SULF1) in carcinogenesis." *Journal of gastrointestinal cancer* 39: 149–158.
- Lavin, Yonit, Deborah Winter, Ronnie Blecher-Gonen, Eyal David, Hadas Keren-Shaul, Miriam Merad, Steffen Jung, and Ido Amit. 2014. "Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment." *Cell* 159 (6): 1312–1326.
- Lee, S, H Kaneko, I Sekigawa, Y Tokano, Y Takasaki, and H Hashimoto. 1998. "Circulating interleukin-16 in systemic lupus erythematosus." *British journal of rheumatology* 37 (12): 1334–1337.
- Liekens, Sandra, Dominique Schols, and Sigrid Hatse. 2010. "CXCL12-CXCR4 axis in angiogenesis, metastasis and stem cell mobilization." *Current pharmaceutical design* 16 (35): 3903–3920.
- Litviňuková, Monika, Carlos Talavera-López, Henrike Maatz, Daniel Reichart, Catherine L Worth, Eric L Lindberg, Masatoshi Kanda, Krzysztof Polanski, Matthias Heinig, Michael Lee et al. 2020. "Cells of the adult human heart." *Nature* 588 (7838): 466–472.
- Lotfollahi, Mohammad, Anna Klimovskaia Susmelj, Carlo De Donno, Leon Hetzel, Yuge Ji, Ignacio L Ibarra, Sanjay R Srivatsan, Mohsen Naghipourfar, Riza M Daza, Beth Martin et al. 2023. "Predicting cellular responses to complex perturbations in high-throughput screens." *Molecular systems biology* 19 (6): e11517.
- Ma, Chi, Justin McCallen, John C McVey, Rajiv Trehan, Kylynda Bauer, Qianfei Zhang, Benjamin Ruf, Sophie Wang, Chunwei Walter Lai, Giorgio Trinchieri et al. 2023. "CSF-1R+ Macrophages Control the Gut Microbiome-Enhanced Liver Invariant NKT Function through IL-18." *The Journal of Immunology*.
- McInnes, Leland, John Healy, and James Melville. 2018. "Umap: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint arXiv:1802.03426*.
- Morrison, Sean J, and Allan C Spradling. 2008. "Stem cells and niches: mechanisms that promote stem cell maintenance throughout life." *Cell* 132 (4): 598–611.
- Müller-Dott, Sophia, Eirini Tsirvouli, Miguel Vazquez, Ricardo O Ramirez Flores, Pau Badia-i Mompe, Robin Fallegger, Dénes Türei, Astrid Lægreid, and Julio Saez-Rodriguez. 2023. "Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities." *Nucleic acids research* 51 (20): 10934–10949.
- Nakahama, Ken-ichi. 2010. "Cellular communications in bone homeostasis and repair." *Cellular and molecular life sciences* 67: 4001–4009.
- Nguyen, Tram TT, Hee Kyeong Yoon, Yoon Tae Kim, Yun Hui Choi, Won-Kyu Lee, and Mirim Jin. 2020. "Tryptophanyl-tRNA synthetase 1 signals activate TREM-1 via TLR2 and TLR4." *Biomolecules* 10 (9): 1283.
- Palla, Giovanni, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L Ibarra, Olle Holmberg, Isaac Virshup et al. 2022. "Squidpy: a scalable framework for spatial omics analysis." *Nature methods* 19 (2): 171–178.
- Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga et al. 2019. "Pytorch: An imperative style, high-performance deep learning library." *Advances in neural information processing systems* 32.
- Perez, Richard K, M Grace Gordon, Meena Subramaniam, Min Cheol Kim, George C Hartoularios, Sasha Targ, Yang Sun, Anton Ogorodnikov, Raymund Bueno, Andrew Lu et al. 2022. "Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus." *Science* 376 (6589): eabf1970.
- Postal, Mariana, and Simone Appenzeller. 2011. "The role of Tumor Necrosis Factor-alpha (TNF- α) in the pathogenesis of systemic lupus erythematosus." *Cytokine* 56 (3): 537–543.
- Qi, Jingjing, Hongxiang Sun, Yao Zhang, Zhengting Wang, Zhenzhen Xun, Ziyi Li, Xinyu Ding, Rujuan Bao, Liwen Hong,

- Wenqing Jia et al. 2022. "Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer." *Nature communications* 13 (1): 1742.
- Reed, Austin D, Sara Pensa, Adi Steif, Jack Stenning, Daniel J Kunz, Linsey J Porter, Kui Hua, Peng He, Alecia-Jane Twigger, Abigail JQ Siu et al. 2024. "A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast." *Nature Genetics* 56 (4): 652–662.
- Ren, Xianwen, Wen Wen, Xiaoying Fan, Wenhong Hou, Bin Su, Pengfei Cai, Jiesheng Li, Yang Liu, Fei Tang, Fan Zhang et al. 2021. "COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas." *Cell* 184 (7): 1895–1913.
- Ring, Larisa, Peter Neth, Christian Weber, Sabine Steffens, and Alexander Faussner. 2014. "β-Catenin-dependent pathway activation by both promiscuous "canonical" WNT3a-, and specific "noncanonical" WNT4-and WNT5a-FZD receptor combinations with strong differences in LRP5 and LRP6 dependency." *Cellular Signalling* 26 (2): 260–267.
- Rodríguez-Carrio, Javier, Patricia López, Mercedes Alperi-López, Luis Caminal-Montero, Francisco J Ballina-García, and Ana Suárez. 2019. "IRF4 and IRGs delineate clinically relevant gene expression signatures in systemic lupus erythematosus and rheumatoid arthritis." *Frontiers in immunology* 9: 3085.
- Roohani, Yusuf, Kexin Huang, and Jure Leskovec. 2024. "Predicting transcriptional outcomes of novel multigene perturbations with GEARS." *Nature Biotechnology* 42 (6): 927–935.
- Ruiz-Moreno, Cristian, Sergio Marco Salas, Erik Samuelsson, Sebastian Brandner, Mariette EG Kranendonk, Mats Nilsson, and Hendrik G Stunnenberg. 2022. "Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma." *BioRxiv*: 2022–08.
- Salcher, Stefan, Gregor Sturm, Lena Horvath, Gerold Untergasser, Christiane Kuempers, Georgios Fotakis, Elisa Panizzolo, Agnieszka Martowicz, Manuel Trebo, Georg Pall et al. 2022. "High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer." *Cancer cell* 40 (12): 1503–1520.
- Van de Sande, Bram, Christopher Flerin, Kristofer Davie, Maxime De Waegeneer, Gert Hulselmans, Sara Aibar, Ruth Seurinck, Wouter Saelens, Robrecht Cannoodt, Quentin Rouchon et al. 2020. "A scalable SCENIC workflow for single-cell gene regulatory network analysis." *Nature protocols* 15 (7): 2247–2276.
- Scandura, Joseph M, Piernicola Boccuni, Joan Massagué, and Stephen D Nimer. 2004. "Transforming growth factor β-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation." *Proceedings of the National Academy of Sciences* 101 (42): 15231–15236.
- Schaar, Anna Christina, Alejandro Tejada-Lapuerta, Giovanni Palla, Robert Gutgesell, Lennard Halle, Mariia Minaeva, Larsen Vornholz, Leander Dony, Francesca Drummer, Mojtaba Bahrami et al. 2024. "Nicheformer: a foundation model for single-cell and spatial omics." *bioRxiv*: 2024–04.
- Sikkema, Lisa, Ciro Ramírez-Suástegui, Daniel C Strobl, Tessa E Gillett, Luke Zappia, Elo Madissoon, Nikolay S Markov, Laure-Emmanuelle Zaragozi, Yuge Ji, Meshal Ansari et al. 2023. "An integrated cell atlas of the lung in health and disease." *Nature medicine* 29 (6): 1563–1577.
- Tomaz da Silva, Pedro, Alexander Karollus, Johannes Hingerl, Gihanna Galindez, Nils Wagner, Xavier Hernandez-Alias, Danny Incarnato, and Julien Gagneur. 2024. "Nucleotide dependency analysis of DNA language models reveals genomic functional elements." *bioRxiv*: 2024–07.
- Song, Gwan Gyu, Jae-Hoon Kim, Young Ho Seo, Sung Jae Choi, Jong Dae Ji, and Young Ho Lee. 2014. "Associations between interleukin 1 polymorphisms and susceptibility to systemic lupus erythematosus: a meta-analysis." *Human immunology* 75 (1): 105–112.
- Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. "Dropout: a simple way to prevent neural networks from overfitting." *The journal of machine learning research* 15 (1): 1929–1958.
- Stephenson, Emily, Gary Reynolds, Rachel A Botting, Fernando J Calero-Nieto, Michael D Morgan, Zewen Kelvin Tuong, Karsten Bach, Waradon Sungnak, Kaylee B Worlock, Masahiro Yoshida et al. 2021. "Single-cell multi-omics analysis of the immune response in COVID-19." *Nature medicine* 27 (5): 904–916.
- Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. 2017. "Axiomatic attribution for deep networks." In *International conference on machine learning*,: 3319–3328, PMLR.

- Suo, Chenqu, Emma Dann, Issac Goh, Laura Jardine, Vitalii Kleshchevnikov, Jong-Eun Park, Rachel A Botting, Emily Stephenson, Justin Engelbert, Zewen Kelvin Tuong et al. 2022. "Mapping the developing human immune system across organs." *Science* 376 (6597): eab0510.
- Tang, Fuchou, Catalin Barbacioru, Yangzhou Wang, Ellen Nordman, Clarence Lee, Nanlan Xu, Xiaohui Wang, John Bodeau, Brian B Tuch, Asim Siddiqui et al. 2009. "mRNA-Seq whole-transcriptome analysis of a single cell." *Nature methods* 6 (5): 377–382.
- Theodoris, Christina V, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, X Shirley Liu et al. 2023. "Transfer learning enables predictions in network biology." *Nature* 618 (7965): 616–624.
- Traag, Vincent A, Ludo Waltman, and Nees Jan Van Eck. 2019. "From Louvain to Leiden: guaranteeing well-connected communities." *Scientific reports* 9 (1): 1–12.
- Tsherniak, Aviad, Francisca Vazquez, Phil G Montgomery, Barbara A Weir, Gregory Kryukov, Glenn S Cowley, Stanley Gill, William F Harrington, Sasha Pantel, John M Krill-Burger et al. 2017. "Defining a cancer dependency map." *Cell* 170 (3): 564–576.
- Türei, Dénes, Tamás Korcsmáros, and Julio Saez-Rodriguez. 2016. "OmniPath: guidelines and gateway for literature-curated signaling pathway resources." *Nature methods* 13 (12): 966–967.
- UniProt Consortium. 2019. "UniProt: a worldwide hub of protein knowledge." *Nucleic acids research* 47 (D1): D506–D515.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention is All you Need." In *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, vol. 30: Curran Associates, Inc.. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fdb053c1c4a845aa-Paper.pdf.
- Vázquez-García, Ignacio, Florian Uhlig, Nicholas Ceglia, Jamie LP Lim, Michelle Wu, Neeman Mohibullah, Juliana Niyazov, Arvin Eric B Ruiz, Kevin M Boehm, Viktoria Bojilova et al. 2022. "Ovarian cancer mutational processes drive site-specific immune evasion." *Nature* 612 (7941): 778–786.
- Wang, Chao, Chen-Chen Feng, Hai-Feng Pan, De-Guang Wang, and Dong-Qing Ye. 2013. "Therapeutic potential of SIGIRR in systemic lupus erythematosus." *Rheumatology International* 33: 1917–1921.
- Waring, Michael J, John Arrowsmith, Andrew R Leach, Paul D Leeson, Sam Mandrell, Robert M Owen, Garry Paireaudieu, William D Pennie, Stephen D Pickett, Jibo Wang et al. 2015. "An analysis of the attrition of drug candidates from four major pharmaceutical companies." *Nature reviews Drug discovery* 14 (7): 475–486.
- Weodon, MN. 2007. "The importance of TCF7L2." *Diabetic Medicine* 24 (10): 1062–1066.
- Wen, Hongzhi, Wenzhuo Tang, Xinnan Dai, Jiayuan Ding, Wei Jin, Yuying Xie, and Jiliang Tang. 2023. "CellPLM: pre-training of cell language model beyond single cells." *bioRxiv*: 2023–10.
- Westmeier, Jaana, Annika Brochtrup, Krystallenia Paniskaki, Zehra Karakose, Tanja Werner, Kathrin Sutter, Sebastian Dolff, Andreas Limmer, Daniela Mittermüller, Jia Liu et al. 2023. "Macrophage migration inhibitory factor receptor CD74 expression is associated with expansion and differentiation of effector T cells in COVID-19 patients." *Frontiers in Immunology* 14: 1236374.
- Wicki, A, and G Christofori. 2007. "The potential role of podoplanin in tumour invasion." *British journal of cancer* 96 (1): 1–5.
- Wolf, F Alexander, Philipp Angerer, and Fabian J Theis. 2018. "SCANPY: large-scale single-cell gene expression data analysis." *Genome biology* 19: 1–5.
- Yazar, Seyhan, Jose Alquicira-Hernandez, Kristof Wing, Anne Senabouth, M Grace Gordon, Stacey Andersen, Qinyi Lu, Antonia Rowson, Thomas RP Taylor, Linda Clarke et al. 2022. "Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease." *Science* 376 (6589): eabf3041.

Learning gene co-regulation across cells with Multicellular Transformer

Supplementary Material

Contents

1	Introduction	1
2	Results	4
2.1	Learning gene co-regulation across cells with Multicellular Transformer	4
2.2	Multicellular Transformer uncovers gene interactions involved in cell-cell communication	5
2.3	Multicellular Transformer discovers disease-specific genes with context-aware gene embeddings	9
2.4	Multicellular Transformer allows for <i>in silico</i> zero-shot perturbation prediction at cell population level	12
2.5	Multicellular Transformer recovers spatial co-localization patterns from scRNA-seq data	14
3	Discussion	16
4	Code and data availability	17
A	Methods	28
A.1	Model architecture & training	28
A.1.1	Multicellular gene embeddings	28
A.1.2	Transformer block	29
A.1.3	Training objective	30
A.1.4	Training details	31
A.2	Data preprocessing & training data augmentation	32
A.3	Downstream tasks & analysis	33
A.3.1	Receptor-Ligand identification	33
A.3.2	Gene set extraction & analysis	35
A.3.3	Context-aware gene embeddings analysis	36

†Corresponding author. Aïcha BenTaieb (bentaieb.aicha@gene.com)

A.3.4	Sample-level clustering	37
A.3.5	Recovering spatial patterns from scRNA-seq	37
A.3.6	<i>In silico</i> perturbation prediction at cell population level	38
A.3.7	Fine-tuning on lung cancer atlas	39
A.4	Datasets	39
A.4.1	Pretraining corpus	39
A.4.2	Lung cancer atlas	39
A.4.3	Spatial transcriptomics datasets	40
A.4.4	DepMap	40
B	Supplementary Figures & Tables	40

Appendix A. Methods

A.1. Model architecture & training

Multicellular Transformer takes as input a set of cells originating from the same sample, allowing it to model the interactions between genes across cells. Therefore, in contrary to a number of existing single-cell foundational models (Cui et al. 2024; Theodoris et al. 2023; Hao et al. 2024; Heimberg et al. 2023), a single input example is not a single cell, but a cell-by-gene matrix $X \in \mathbb{R}^{M \times N}$ originating from the same sample, where M is the number of cells and N is the number of genes. Here, the sample means cells coming from the same donor, tissue and sequencing batch. Therefore, the cells which may interact with each other. This matrix is then used for modeling the contextual interactions between genes in a multicellular environment.

A.1.1. Multicellular gene embeddings

The multicellular gene embeddings, used as input to the transformer, consist of two parts: the gene expression latent vector and the gene identity embeddings. The first represents the context-dependent activity of the gene across cells in the sample, while the second represents the context-independent function of the gene, which is the same across all samples. The two are computed separately and then combined by summation.

Gene expression embeddings. Given an expression matrix $X \in \mathbb{R}^{N \times M}$, where N is the number of genes in a sample and M is the number of cells, each element $X_{i,j} \in \mathbb{R}^+$ represents the normalized expression value (after \log_{10} transformation) of gene g_i in cell c_j .

For each gene g_i , we define its expression vector $\mathbf{x}_{g_i} \in \mathbb{R}^M$ as the i -th row of X :

$$\mathbf{x}_{g_i} = [X_{i,1}, X_{i,2}, \dots, X_{i,M}]$$

This vector represents the expression of gene g_i across all M cells in the sample.

The feed-forward neural network (FFNN) MLP_e is a 2-layer multilayer perceptron (MLP) which transforms the gene expression vector across cells into a latent vector of dimension D :

$$GEXEmb_{g_i} = MLP_e(\mathbf{x}_{g_i})$$

The MLP_e is shared across all genes to prevent overfitting, therefore, it is not gene-specific.

Importantly, the input units of MLP_e of dimension M are associated with certain cell types during training. Specifically, given a set of cell types $\{t_1, t_2, \dots, t_T\}$, where $T \leq M$, the input vector \mathbf{x}_{g_i} can be partitioned according to these cell types:

$$\mathbf{x}_{g_i} = [\mathbf{x}_{g_i}^{t_1}, \mathbf{x}_{g_i}^{t_2}, \dots, \mathbf{x}_{g_i}^{t_T}]$$

where $\mathbf{x}_{g_i}^{t_k}$ contains the expression values of gene g_i in cells of type t_k , such as *CD4 T cell*. Fixing input units to certain cell types allows the model to learn cell type-specific patterns, instead of just learning average gene loadings, and is crucial for the model to learn meaningful interactions.

It is also possible to use a pseudobulked input vector as input to the model. In this case, we aggregate the expression values of each gene across all cells of the same cell type, resulting in a reduced expression vector that captures the total expression per cell type. Formally, for each gene g_i , we compute a pseudobulked

expression vector $\mathbf{p}_{g_i} \in \mathbb{R}^T$, where T is the number of cell types ($T \leq M$). Each element $p_{g_i}^{t_k}$ in \mathbf{p}_{g_i} represents the sum of normalized expression values of gene g_i across all cells belonging to the cell type t_k :

$$p_{g_i}^{t_k} = \sum_{j \in C_{t_k}} X_{i,j}$$

where C_{t_k} is the set of cell indices corresponding to cell type t_k . In this pseudobulked approach, the input dimension to the gene expression network MLP_e becomes T instead of M , reducing the input size and focusing on aggregated expression patterns per cell type. We have trained models using both the original per-cell expression vectors $\mathbf{x}_{g_i} \in \mathbb{R}^M$ and the pseudobulked expression vectors $\mathbf{p}_{g_i} \in \mathbb{R}^T$, with both approaches performing competitively. Model weights for both versions are made available.

We have also experimented with training the model without utilizing cell type labels, effectively treating all cells as a single group. The resulting model significantly underperformed compared to the one using cell type labels, highlighting the importance of incorporating cell type information.

Finally, during pretraining, we used cell type labels to guide the partitioning and aggregation of expression vectors. However, it is also possible to use any mutually exclusive cell labels, such as clusters derived from unsupervised methods like Leiden clustering. We demonstrated this when fine-tuning the Multicellular Transformer on spatial transcriptomics data, where we utilized labels from Leiden clustering performed on the dataset.

Gene identity embeddings. Given a predefined set of genes G , we assign each gene g_i a unique identifier, which forms a vocabulary of tokens used in the Multicellular Transformer. This allows us to harmonize different sets of genes across different studies, rather than taking the intersection of genes across them. We use these identifiers to transform a gene into a multidimensional vector of dimension D , using a gene embedding layer function emb_g , which is learned during pretraining. The output of the gene embedding layer emb_g is the same size as the output of the gene expression MLP_e .

Combining all of the above, for a gene g_i with expression vector $\mathbf{x}_{g_i} \in \mathbb{R}^M$, the multicellular gene embedding is defined as:

$$\text{Embedding}_{g_i} = \text{emb}_g(g_i) + \text{MLP}_e(\mathbf{x}_{g_i})$$

Across a set of genes present in the sample, the input to the transformer is:

$$\text{Input} = [\text{Embedding}_{g_1}, \text{Embedding}_{g_2}, \dots, \text{Embedding}_{g_L}],$$

where L denotes a predefined maximum input length. This setup allows us to: (a) account for multiple cells in the environment and the interactions between them, (b) consider the context-dependent gene expression of the genes present in the sample, (c) represent gene function independent of expression across samples.

A.1.2. Transformer block

We leverage the transformer (Vaswani et al. 2017) architecture to compute contextual representations of multicellular gene embeddings. Given the expression matrix $X \in \mathbb{R}^{M \times N}$, we first compute the multicellular

gene embeddings for each gene g_i as previously defined:

$$\text{Embedding}_{g_i} = \text{emb}_g(g_i) + \text{MLP}_e(\mathbf{x}_{g_i})$$

where $\mathbf{x}_{g_i} \in \mathbb{R}^M$ is the expression vector of gene g_i across all M cells in the sample.

We collect these embeddings into a sequence:

$$\mathbf{E} = [\text{Embedding}_{g_1}, \text{Embedding}_{g_2}, \dots, \text{Embedding}_{g_L}]$$

where L is the number of genes considered (up to a predefined maximum input length).

We then compute their contextual representations $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_L]$ by passing them through multiple transformer layers:

$$\mathbf{H} = \text{Transformer}(\mathbf{E})$$

To improve the efficiency of the self-attention mechanism, we use Flash Attention ([Dao et al. 2022](#)). Additionally, we include special tokens [CLS] and [SEP] to mark the beginning and end of the sequence.

A.1.3. Training objective

The model is optimized with two objectives: (1) a masked modeling objective ([Devlin 2018](#)), and (2) a gene expression reconstruction objective.

In the first objective, we randomly mask 15% of the genes in the input sequence. Specifically, we design our own masking strategy, where out of the 15% of masked genes:

- 70% are masked gene identity embeddings: for gene g_i , we replace its gene identity embedding $\text{emb}_g(g_i)$ with a special token embedding [MASK].
- 10% are masked gene expression embeddings, where we replace the gene expression embedding with [MASK] vector,
- 10% where random gene identity embedding is swapped with the true gene identity embedding,
- 10% where neither gene identity nor expression is masked, leaving the embeddings unchanged.

This strategy is inspired by the masking strategy in BERT ([Devlin 2018](#)) and allows the model to learn robust and context-aware representations by predicting masked gene identities and expressions. By varying the masking types—including masking gene identities, gene expressions, swapping gene identities, and leaving some unchanged—the model is encouraged to leverage both gene identity and expression information. This helps the model capture complex dependencies among genes and understand the impact of gene expression patterns across different cells, thereby enhancing its ability to model gene interactions in a multicellular environment. The gene probabilities for each masked gene are computed by passing the output of a transformer $\mathbf{h}_i \in \mathbb{R}^D$ of a masked gene g_i through a classifier head which is weight matrix $W \in \mathbb{R}^{D \times G}$, where D is the embedding dimension and G is the total number of genes in the pretraining corpus. W is learned over the course of pretraining and used for analysis on downstream tasks. The masked modeling loss $\mathcal{L}_{\text{mask}}$ is defined as:

$$\mathcal{L}_{\text{mask}} = -\frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \log P(g_i | \mathbf{E}_{\text{masked}})$$

where \mathcal{M} is the set of positions of masked gene identities, g_i is the true gene identity at position i , and $\mathbf{E}_{\text{masked}}$ represents the input embeddings with applied masking.

For the gene expression reconstruction objective, we use the last hidden state from the transformer block to reconstruct the input normalized gene expression for each gene. This is performed for all genes in a sample, including genes with masked gene identity as well as masked gene expression. To reconstruct the gene expression, we learn a 2-layer MLP, MLP_r , which, given a D -dimensional vector, projects it into an M -dimensional vector equal to the input gene expression vector:

$$\hat{\mathbf{x}}_{g_i} = \text{MLP}_r(\mathbf{h}_i)$$

where $\mathbf{h}_i \in \mathbb{R}^D$ is the hidden representation of gene g_i from the transformer, and $\hat{\mathbf{x}}_{g_i} \in \mathbb{R}^M$ is the reconstructed gene expression vector. The reconstruction loss $\mathcal{L}_{\text{recon}}$ is defined as the mean squared error between the reconstructed and true gene expression vectors:

$$\mathcal{L}_{\text{recon}} = \frac{1}{L} \sum_{i=1}^L \|\hat{\mathbf{x}}_{g_i} - \mathbf{x}_{g_i}\|^2$$

where L is the total number of genes in the input sequence.

The final loss function consists of the two losses described above, weighted by α (masked cross-entropy loss) and $1 - \alpha$ (reconstruction loss):

$$\mathcal{L} = \alpha \mathcal{L}_{\text{mask}} + (1 - \alpha) \mathcal{L}_{\text{recon}}$$

we experimented with different values of α and used $\alpha = 0.99$ in the best performing pretraining run. Large value of α is due to the reconstruction loss having significantly larger absolute values. By optimizing this combined loss function, the model learns to predict masked gene identities and accurately reconstruct gene expression profiles, capturing intricate patterns and dependencies within the multicellular context. We have found that the masked learning objective is the most important to learning gene-gene interactions which we are interested in. However, the gene expression reconstruction improves the stability of training and allows us to use it for downstream tasks, such as *in silico* perturbation effect prediction at cell population level.

A.1.4. Training details

The transformer ([Vaswani et al. 2017](#)) block consists of 12 transformer layers with self-attention. The gene expression embedding MLP_e and gene expression reconstruction MLP_r are both 2-layer MLPs, interleaved with *ReLU* activation function ([Agarap 2018](#)), layer normalization ([Ba 2016](#)) and dropout ([Srivastava et al. 2014](#)) equal to 0.2. The gene identity embedding is an embedding layer $emb_g \in \mathbb{R}^{G \times M}$. All parameters are learned from scratch. Total nr of parameters equals $\approx 100M$. The model is implemented in PyTorch ([Paszke et al. 2019](#)).

We have performed random split on sample-level, keeping 0.95 for training and 0.05 for validation. We did not include a test set as we do not have a downstream task which directly matches our pretraining scenario. During pretraining, we set the number of cells in an input example of the model to 1,024 and filter genes which are expressed (normalized expression > 0) in less than 1% of the sampled cells. The maximum context size, which here is the maximum nr of genes in a sample, was set to 12,000.

We have trained the model with the learning rate 0.00015, *Adam* optimizer (Kingma 2014) with weight decay and a linear warmup of 10,000 steps. We have performed validation every 10,000 steps, monitoring the validation loss. We used the batch size of 4 while accumulating gradients for 4 batches before performing a backward pass. The maximum nr of training steps has been set to 500,000. The training has been performed on a single NVIDIA A100 80GB GPU for approximately 14 days.

A.2. Data preprocessing & training data augmentation

scRNA-seq preprocessing. Multicellular Transformer takes as input a gene-by-cell expression matrix $X \in \mathbb{R}^{N \times M}$, where N is the number of genes and M is the number of cells, each element $X_{i,j} \in \mathbb{R}^+$ is the normalized expression value. The normalization is performed by normalizing each cell by total counts over all genes, so that every cell has the same total count after normalization. We set the target counts sum to 10,000. Then, \log_{10} normalization is applied. The preprocessing was performed using [scipy](#) package (Wolf, Angerer, and Theis 2018).

Cell type harmonization. Multicellular Transformer learns interactions between genes across cells, therefore, in contrary to existing single-cell foundational models, the input is not the expression of a gene in a single-cell, but the expression of a gene across distinct cells. The cells within and across samples differ, and we cannot rely on the same genes as features for every sample, therefore, to ensure consistency, the input units of the gene expression embedding, MLP_e are reserved for certain cell types during pretraining (see *Gene expression embeddings* section above). However, cell types across datasets often differ, therefore, we harmonize more than 300 cell types present in the datasets used for pretraining to X cell types. We recalculate the relative number of cells for each cell type, and reserve a number of input units in the MLP_e proportional to the relative count of the cell type. Finally, we remove the cells belong to cell types with less than 1% of relative counts.

Training data augmentation. An input example to the Multicellular Transformer is a sample, and our pretraining corpus has around 4,200 distinct samples, which is a low number for training a model of this size, even considering the large diversity of the samples. Moreover, Multicellular Transformers requires M cells as input in every sample, which is used as input to gene expression embedding MLP, MLP_e . Therefore, we employ a training data augmentation technique by iteratively sampling M cells from each sample. This prevents the model from simply remembering, thus overfitting to each sample, as every time the set of sampled cells is different, and many samples have more than 10,000 cells. We sample the cells from each sample using stratified sampling, where the class is a cell type label. This ensures consistency over various samples. The cell type counts used for stratified sampling are precomputed prior to training. In some samples, there may not be enough of cell types of a particular class, for example, sample *A* has only 70 *Macrophages*, while the MLP_e required 80. In this case, we pad the input with zeros. The training data augmentation played a crucial role in our pretraining, with the model without it performing significantly worse.

Gene filtering. Across cells, the majority of the genes will have scRNA-seq counts equal to 0. This is due to the presence of dropout in the transcriptomic data. Existing single-cell foundational models, such as scGPT drop the genes with 0 counts, as this would make predicting gene identity challenging, causing the model to fail to converge, as the huge majority of genes will have the same expression level, equal to 0. This is a severe

limitation as it makes the model only learn the interactions between the genes with expression higher than 0, which is usually around 1,000 genes.

Multicellular Transformer takes as input a set of cells and their corresponding scRNA-seq counts across genes. Therefore, instead of filtering genes at a single-cell level, we filter genes at the sample-level. Specifically, during pretraining we filter out genes which are expressed (normalized expression > 0) in less than 1% of cells. This allows us to significantly increase our gene context size and take as input $\approx 12,000$ genes, rather than $\approx 1,000$ genes in a single-cell scenario. Our results suggest, that the context size plays a crucial role on downstream tasks, such as gene set extraction and receptor-ligand identification. We have also experimented with more or less stringent filtering and found that more stringent filtering leads to a stable training, however, worse results on downstream tasks, while the less stringent filtering leads to less stable training and often similar or worse performance on downstream tasks. The context size used during our pretraining was set to 12,000 genes.

Sampling robustness analysis. During training and validation, Multicellular Transformer samples M cells from every sample, where M is a parameter set prior to training ($M = 1,024$ during pretraining). This serves as a training data augmentation technique, as the number of samples is relatively low, but each sample contains a different number of cells. Therefore, by sampling from each sample, we increase the effective training data size and introduce variability, helping the model generalize better across distinct cell populations. However, by sampling from each sample, we may also introduce noise if the result of sampling from the same sample varies too much. To assess it, we have run sampling on 100 randomly selected samples from the validation set of the pretraining corpus, sampling each sample 5 times. We have then extracted 50 nearest neighbors for each gene across samples, and evaluated the overlap between the gene neighborhoods within the same sample as well as across samples. The overlap within a sample across genes is significantly higher than the one across samples, achieving the median of 0.2 Jaccard similarity, compared to 0.08 between distinct samples. Therefore, showing that MT can robustly account for the variability within the same sample, maintaining relatively stable gene neighborhoods compared to the gene neighborhoods across distinct samples.

A.3. Downstream tasks & analysis

A.3.1. Receptor-Ligand identification

To analyse the interactions between receptor and ligand genes learned over the course of pretraining, we extracted the pretrained gene embedding matrix from the masked learning classifier, resulting in a $X \in \mathbb{R}^{G \times D}$ matrix, where G is the total number of genes present in the pretraining corpus and D is the dimensionality. Using a predefined set of true and false receptor-ligand pairs, we have computed the cosine similarity value for each pair and employed it to calculate the area under the ROC curve (AUC).

Dataset. To evaluate the performance on the receptor-ligand identification task, we have used a set of known consensus receptor-ligand pairs from *LIANA* (Dimitrov et al. 2024). Importantly, not all of these pairs have been experimentally validated and may not be physically present on the cell surface for the receptors and secreted for the ligands. However, they represent a set of consensus pairs across multiple receptor-ligand sources and have been shown to play a crucial role in cell-cell communication. Therefore, we use it for evaluation.

To establish negative pairs, which would complement the set of positive receptor-ligand pairs described above, we randomly created negatives using genes present in all the benchmarks in a ratio 1:10 for positives to negatives. Importantly, to avoid only including easy negatives with no overlap with the positive pairs, half of the negatives were created by using a receptor present in the true positives. This is to avoid a situation, where the positives and negatives have no overlap in the genes, making it potentially significantly easier for the model and inflating the results. In our case, we wanted to evaluate whether the model can identify the correct ligands for a receptor from a large pool of potential ligands. Finally, our setup does not account for false negatives, which may arise as the space of receptor-ligand remains unexplored. However, the statistical chances of it are small and the evaluation is the same for all models, therefore making it comparable.

For fair comparison, we only considered the receptor-ligand pair which were present in all benchmarks used for the analysis. Additionally, as our analysis was transcriptomics-based, we excluded pairs where either a receptor or ligand are a protein complex created from more than one gene product.

Benchmarks. We followed the same approach for extracting gene embeddings for scGPT and CellPLM as in the gene set extraction analysis. We then computed the cosine similarity for all pairs using both benchmarks and used this score for computing metrics. For co-expression, we followed the same steps as in the section on gene set extraction analysis to get a matrix $E \in \mathbb{R}^{G \times S}$, where S represents the number of samples in the pretraining corpus, and G total number of genes. We then computed the *Pearson correlation* coefficient for all pairs of genes in the dataset and used the resulting coefficient for computing metrics. Finally, for context size impact analysis. We have used Multicellular Transformer models pretrained with different context size.

Integrating multicellular and protein-level information. To complement MCT with protein-level context, we employed the ESMC protein language model ([ESM Team 2024](#)), a state-of-the-art approach for learning protein representations. Using UniProt, we mapped the *Ensembl* gene IDs in the Liana consensus receptor-ligand set ([Dimitrov et al. 2024](#)) to their corresponding UniProt protein IDs and retrieved the associated amino-acid sequences. Each sequence was embedded with the open-source 300 M-parameter ESMC model, averaging the residue embeddings to obtain a single vector for every protein. We then computed the cosine similarity between receptor and ligand vectors to produce a scalar interaction score. This ESMC score was concatenated with the MCT score to form a two-dimensional feature vector, which served as input to an XGBoost classifier. We performed five-fold cross-validation and report the mean performance across folds.

Identifying protein complexes using contextual gene embeddings. We followed the same approach to compute cosine similarities between contextual gene embeddings as in the paragraph above. However, we did not limit ourselves to receptor-ligand pairs and compute cosine similarity across both disease and healthy together. We focused on cytokines, as they play key part in immune system signaling. We then computed cosine similarities across all cytokines and filtered pairs which appeared in less than 3 samples. Finally, we only considered pairs where one of the pairs was secreted from the cell and the other was a transmembrane protein. We have retrieved this information from UniProt ([UniProt Consortium 2019](#)). The top 70 pairs were screened using AlphaFold3 ([Abramson et al. 2024](#)) server.

Analysing receptor-ligand interactions in healthy vs disease. We used the systemic lupus erythematosus (SLE) study ([Perez et al. 2022](#)), and ran Multicellular Transformer across all samples. For each sample, we have computed the cosine similarity using contextual embedding of receptor-ligand pairs from Liana

consensus set (Dimitrov et al. 2024) present in the sample, keeping track of whether the sample has been from disease or healthy patient. After collecting this data, we have excluded the receptor-ligand pairs which appeared less than 3 times in either healthy or disease to remove pairs for which the resulting analysis would not be interpretable due to low sample size. Then, we calculated the relative change in the cosine similarity value in disease vs healthy. Let X represent the cosine similarity of the receptor-ligand pair in disease samples, and Y represent the cosine similarity in healthy samples. Then, the relative fold change in cosine similarity Δ_{rel} is given by $\Delta_{\text{rel}} = \frac{X-Y}{Y}$.

A.3.2. Gene set extraction & analysis

Similarly as with receptor-ligand analysis, to analyse the interactions between genes learned by the model, we have extracted the pretrained gene embedding matrix from the masked learning classifier, resulting in a $X \in \mathbb{R}^{G \times D}$ matrix, where G is the total number of genes present in the pretraining corpus and D is the dimensionality. We have also experimented with using the gene identity embeddings, but found the former to perform better across different tasks. We have followed the setup for gene set extraction used in scGPT (Cui et al. 2024) for comparability. The gene embeddings were used to compute the gene-gene similarity matrix. We have then used the gene-gene similarity matrix to cluster genes into gene sets by applying Leiden clustering (Traag, Waltman, and Van Eck 2019). The gene sets are mutually exclusive, meaning if a gene was present in gene set A, it could not be present in any other gene set. We have run the Leiden clustering across different resolutions, setting minimum resolution at 10, as lower resolutions led to very large gene sets, which were not interpretable (Supp. Fig. S2c). Following clustering, we have excluded the gene sets with fewer than 4 genes.

Cell-cell communication pathways enrichment. To assess whether the extracted gene sets were enriched in cell-cell communication related pathways, we selected the pathways from GO (Ashburner et al. 2000) and Reactome (Fabregat et al. 2018) and filtered them for gene sets which included either of predefined keywords: [receptor, ligand, signal transduction, cell signaling, signaling pathway, signaling pathway, communication]. The extracted pathways were then manually verified. This left us with 451 CCC related GO pathways and 137 CCC related Reactome pathways. We used GSEAp (Fang, Liu, and Peltz 2023) to identify pathways enriched in gene sets, and kept only the significantly enriched pathways. We performed enrichment using both GO and Reactome libraries.

We used the highly variable genes from the tumor microenvironment atlas, enriched with receptor and ligand genes, setting the total number of genes to 8,000. We only considered the genes which were present in all benchmarked methods. The tumor microenvironment atlas (Guimarães et al. 2024) was selected as (1) the multicellular interactions are particularly important in cancer, (2) none of the methods have been trained on the dataset, therefore the setup was zero-shot.

Benchmarks. For *scGPT* (Cui et al. 2024) and *CellPLM* (Wen et al. 2023), we used the latest recommended checkpoints to extract gene embeddings (whole human, pretrained on 33M human cells for scGPT and checkpoint 20230926_85M for CellPLM). We then followed the same framework for gene set extraction and analysis as for the Multicellular Transformer described above. For co-expression, we have aggregated the normalized expression across samples, summing over the expression over cells for each gene in the sample. Therefore, focusing on the intercellular, rather than intracellular expression achieved by calculating the

co-expression at cell-level. This results in a matrix $E \in \mathbb{R}^{G \times S}$, where S represents the number of samples in the pretraining corpus, and G total number of genes. We then computed the PCA ([Abdi and Williams 2010](#)) on matrix E , resulting in a matrix $E_{PCA} \in \mathbb{R}^{G \times P}$, where P is the number of principal components (here, 50). The E_{PCA} matrix was used to cluster genes with Leiden clustering, same as for the methods above.

We provide a jupyter notebook outlining how to extract gene sets given multicellular gene embeddings.

A.3.3. Context-aware gene embeddings analysis

For context-aware gene embeddings analysis, we used the output representations from the final transformer layer. We wanted to identify condition-specific genes, which we defined as genes with consistent neighborhood within a condition, while simultaneously having a different neighborhood across conditions.

To calculate this, we retrieved the 50 nearest neighbors for each gene across all samples from three datasets: 1) immune system across organs ([Suo et al. 2022](#)), 2) systemic lupus erythematosus ([Perez et al. 2022](#)), and 3) COVID-19 ([Stephenson et al. 2021](#)). We then computed the mean overlap across neighborhoods within a condition by calculating the Jaccard similarity between the neighborhoods of each gene across samples with the same condition (e.g., from the same tissue or disease) using the 20 nearest neighbors.

For each gene g , let $N_g^{(s,c)}$ denote the set of 20 nearest neighbors in sample s with condition c . The mean within-condition Jaccard similarity $J_{\text{within}}^{(c)}(g)$ for gene g in condition c is computed as follows:

$$J_{\text{within}}^{(c)}(g) = \frac{1}{|S_c| \cdot (|S_c| - 1)} \sum_{\substack{s, s' \in S_c \\ s \neq s'}} \frac{|N_g^{(s,c)} \cap N_g^{(s',c)}|}{|N_g^{(s,c)} \cup N_g^{(s',c)}|}$$

where S_c is the set of samples under condition c .

Next, we computed the Jaccard similarity $J_{\text{across}}^{(c_1, c_2)}(g)$ between each pair of conditions c_1 and c_2 for gene g :

$$J_{\text{across}}^{(c_1, c_2)}(g) = \frac{|N_g^{(c_1)} \cap N_g^{(c_2)}|}{|N_g^{(c_1)} \cup N_g^{(c_2)}|}$$

where $N_g^{(c)}$ represents the most common 40 neighbors of gene g within condition c .

Calculating the overlap between samples for each gene is computationally expensive, requiring $G \times |S| \times |S|$ Jaccard similarity calculations, where G is the number of genes and S is the number of samples per condition. To reduce this, we extracted the most common 40 neighbors for each gene within a condition and used these to calculate the overlap across conditions.

We excluded conditions with too few samples to allow for meaningful analysis. Finally, we calculated the fold change between the mean Jaccard similarity within a condition vs. across conditions:

$$\Delta J(g) = \frac{J_{\text{within}}^{(c)}(g)}{J_{\text{across}}^{(c)}(g)}$$

We used this fold change $\Delta J(g)$ to identify condition-specific genes.

A.3.4. Sample-level clustering

To extract sample-level embeddings for clustering. We ran Multicellular Transformer in an inference mode and computed sample-level embeddings by averaging out all representations from the final layer of transformer. This results in a matrix $X \in \mathbb{R}^{S \times D}$, where S is the number of samples and D is the dimensionality of the representations. We then ran UMAP ([McInnes, Healy, and Melville 2018](#)) on the relevant samples with the number of neighbors varying from 15 to 30 depending on the dataset the samples originated from.

To calculate clustering metrics, we have performed Leiden clustering ([Traag, Waltman, and Van Eck 2019](#)) at the resolution of 1.0 and computed the Adjusted Rand Index (ARI) as well as Normalized Mutual Information (NMI). We have also experimented with different levels of Leiden clustering resolution, with the 1.0 consistently showing the best results across methods and datasets.

A.3.5. Recovering spatial patterns from scRNA-seq

Spatial colocalization analysis. We computed the spatial colocalization by calculating the bivariate Moran's I . The bivariate Moran's I is given as:

$$I_{XY} = \frac{N \sum_i \sum_j w_{ij}(X_i - \bar{X})(Y_j - \bar{Y})}{\sum_i (X_i - \bar{X})^2 \sum_i \sum_j w_{ij}}$$

where N is the total number of observations (cells or spots), X_i and Y_j are the values of the two variables X and Y for cells i and j , \bar{X} and \bar{Y} are the means of X and Y , w_{ij} is the spatial weight between observations i and j , where $w_{ij} = 1$ if i and j are neighbors and $w_{ij} = 0$ otherwise.

For the Xenium data, we defined the neighborhood as cells within the radius of 50 μm of the target cell, while for Visium data, we defined it as 6 neighboring spots in the grid. We used the Squidpy package ([Palla et al. 2022](#)) to perform the analysis. We calculated the bivariate Moran's I for a random subset of 200 genes, meaning we calculated the pairwise Moran's I index for 200×200 pairs. This covers the majority of the genes as the gene panels in the Xenium datasets used do not exceed 400 genes after filtering out lowly expressed genes. The cosine similarity has been calculated by extracting the similarity using the contextual gene embeddings between genes across relevant samples. Therefore, for Xenium lung cancer the cosine similarities were computed by running the Multicellular Transformer model in the inference mode across samples from lung cancer atlas ([Salcher et al. 2022](#)), and for Xenium breast cancer it was performed by running the Multicellular Transformer fine-tuned on the same dataset across samples in the inference mode. For calculating the correlation between cosine similarity across percentiles, for each percentile, we limited the pairs of genes to values above the percentile and then recalculated the correlation coefficient only for these pairs.

Fine-tuning on spatial transcriptomics data. We used the Xenium breast cancer dataset to perform fine-tuning on the spatial transcriptomics data. Due to the lack of annotated cell types, we performed Leiden clustering using default parameters in the scanpy package. We then used the resulting annotations and used them as input to the model treating them as cell type labels. Therefore, showing how one can leverage the model without available cell type labels. As a sample, we used the neighboring cells of within a radius of 350 μm of the cell of interest. We then fine-tuned the model using the pretrained objective, leveraging the pretrained gene identity embeddings and transformer block, while learning new MLP_e and MLP_r , allowing

the model to adapt to the new dataset. We have run the fine-tuning with the same hyperparameters, except the number of cells per sample M being set to 512.

Cell type-Cell type interactions. We leveraged the gene annotations from the 10xGenomics dataset explorer¹ to associate genes with particular cell states. A table with a set of genes per cell type can be found in Supp. Table S2. To calculate the interactions between cell types, we calculated the pairwise similarity between all genes associated with a cell state. Then, to account for the mean cosine similarity of each gene, we divided the resulting cosine similarities by the mean cosine similarities of each gene. Therefore, for genes g_x and g_y , the normalized cosine similarity is given as

$$2 \frac{g_{xy}}{\bar{g}_x + \bar{g}_y}$$

where g_{xy} represents the cosine similarity between g_x and g_y , and \bar{g}_x and \bar{g}_y are the mean cosine similarities of g_x and g_y across all genes, respectively. We have then aggregated the results for each cell type-cell type pair, taken a mean of it and applied \log_2 transformation.

We have followed the same approach for calculating cell type-cell type interaction matrix using the spatial colocalization, using spatial colocalization instead of cosine similarity as the measure of similarity.

A.3.6. *In silico* perturbation prediction at cell population level

Calculating gene-gene relationships upon *in silico* genetic perturbations. We calculated the perturbation effect on other genes by removing the gene from the input genes and evaluating the impact of the gene removal on the predicted gene logits. Specifically, we followed a similar approach to the one proposed in Silva et al. (2024). For a vector of predicted gene logits b_i for a gene i without a perturbed gene and a vector of predicted gene logits p_i with a perturbation, where both logits vectors are of dimension G , equal to the total nr of genes, the predicted perturbation effect pef_i is given as:

$$pef_i = \max \left(\log_2 \left(\frac{b_i}{p_i} \right) \right).$$

Note that removing the gene from the input is equivalent to zeroing out the gene expression in the Multicellular Transformer setup, as not expressed genes are removed from the input. We ran the *in silico* perturbation across validation samples from the lung cancer atlas, performing the knock-out on each sample where the gene of interest was present, using the Multicellular Transformer model fine-tuned on lung cancer atlas. We then took the mean perturbation effect for true interacting genes and control for each perturbed gene and normalized the results for each knocked-out gene with min max normalization for consistency.

We have used the same dataset as for receptor-ligand task, with the difference that we added negatives following the same approach for the ligands as well. Here the negative pairs served as control. We then selected a subset of the genes to perform the *in silico* gene perturbation.

Predicting the effect of genetic perturbation across cell types. Using trained Multicellular Transformer model, we measured the effect of a gene knock-out on a cell type by removing the gene from the input sample and measuring the change in reconstructed gene expression across cell types vs the baseline reconstructed

¹<https://www.10xgenomics.com/products/xenium-in-situ/human-breast-dataset-explorer>

gene expression. Specifically, for a baseline reconstructed normalized expression in cell type t , b_t , and reconstructed normalized expression with an *in silico* knocked-out gene, p_t , the perturbation effect for a cell type t is given as $pef_t = \frac{p_t - b_t}{b_t}$. For this, we used Multicellular Transformer fine-tuned on the lung cancer atlas (Salcher et al. 2022) with normalized expression pseudobulked at cell type level. We have decided to use this model as this makes measuring the change in the normalized expression of a cell type easier.

A.3.7. Fine-tuning on lung cancer atlas

Multicellular Transformer can be fine-tuned allowing for dataset-specific analysis. We fine-tuned Multicellular Transformer on a lung cancer atlas (Salcher et al. 2022). The H5AD object was accessed from <https://datasets.cellxgene.cziscience.com/08adcdb7-1f28-42f6-8969-f8f13a2f82f5.h5ad> and split randomly by sample into train and validation following a 80%, 20% ratio. During fine-tuning, we leverage the pretrained gene embeddings and transformer block and train the gene expression embedding MLP_e , as well as gene expression reconstruction MLP_r from scratch. This allows the model to adapt to the cell types present in the specific dataset, while reusing the information learned over the course of pretraining by leveraging pretrained gene identity embeddings, transformer block and masked gene classifier.

We have followed the same scRNA-seq normalization procedure as for pretraining corpus. We have set the nr of cells sampled from each sample to 512, therefore, the first layer of the MLP_e used for fine-tuning has been also set to 512. Samples with less than 512 cells have been filtered out, and cell types which make up of less than 0.2% cell types were also filtered out. We have used the cell type labels present in the H5AD object downloaded from the CELLxGENE portal. The pretraining has been done using the same hyperparameters as during pretraining.

A.4. Datasets

A.4.1. Pretraining corpus

We have collected a corpus using 13 publicly available datasets from CELLxGENE (<https://cellxgene.cziscience.com/>). The dataset spanned more than 14M cells, including more than 3k donors and 4.2k samples. We have filtered out genes with less than 5 counts across cells and cells with less than 100 genes expressed using scanpy package (Wolf, Angerer, and Theis 2018). We have also filtered out samples with less than 1,024 cells, which is the number of cells sampled from each scRNA-seq sample. We took the union of all genes present across datasets, resulting in almost 60k genes and did not perform any highly variable genes selection, leveraging all genes. All of the data has been sequenced from *Homo Sapiens*. The corpus includes various tissues and diseases (Supp. Fig. S1) to allow for learning across diverse contexts. Exact list of datasets used and their statistics can be found in Supp. Table S1. We have harmonised the cell types across datasets and filtered out the cell types which made up less than 0.1% of all cells.

We have used the *CellArr* package (<https://github.com/BiocPy/cellarr>) to combine multiple datasets into one and allow seamless sampling of cells from the same sample during training and evaluation.

A.4.2. Lung cancer atlas

The lung cancer atlas (Salcher et al. 2022) has been accessed from the CELLxGENE portal (<https://datasets.cellxgene.cziscience.com/08adcdb7-1f28-42f6-8969-f8f13a2f82f5.h5ad>). We

have filtered out genes and cells with less than 5 and 100 counts respectively using scanpy package. Additionally, we filtered out samples with less than 512 cells. We have used the cell type annotations provided in the dataset, filtering out cell types which made up less than 0.1% of all cells. We used the dataset for fine-tuning.

A.4.3. Spatial transcriptomics datasets

Xenium human lung & breast cancer. Both datasets have been downloaded from the 10x genomics website (lung cancer: <https://www.10xgenomics.com/datasets/ffpe-human-lung-cancer-data-with-human-immuno-oncology-profiling-panel-and-custom-add-on-1-standard>, breast cancer: <https://www.10xgenomics.com/products/xenium-in-situ/preview-dataset-human-breast>). We have filtered out genes and cells with less than 5 and 10 counts respectively using scanpy package.

For fine-tuning on the breast cancer dataset, we performed Leiden clustering ([Traag, Waltman, and Van Eck 2019](#)) using scanpy package ([Wolf, Angerer, and Theis 2018](#)) to annotate each cell with a label. We have then used these cluster labels as the equivalent of cell type labels during fine-tuning. Therefore, showing how one can use alternative labels to cell type labels for training the model.

Visium glioblastoma. The Visium glioblastoma dataset has been accessed using scanpy package (sample id: Parent_Visium_Human_Glioblastoma). We have filtered out genes with less than 10 counts across cells and cells with less than 1000 or more than 35000 counts respectively using scanpy.

A.4.4. DepMap

We have extracted the publicly available viability scores for each gene from the DepMap consortium ([Tsherniak et al. 2017](#)) website (https://depmap.org/portal/data_page/). We leveraged the CRISPRGeneEffect.csv file to extract gene effect estimates for all cell lines. The gene effect estimates have been already copy number corrected, scaled, and screen quality corrected. We filtered to relevant lung cancer cell lines present in both DepMap as well as lung cancer atlas, specifically *lung adenocarcinoma*, *squamous cell lung carcinoma* and *non-small cell lung carcinoma*, leaving us with gene effect estimates for 81 cell lines across 18443 genes.

Appendix B. Supplementary Figures & Tables

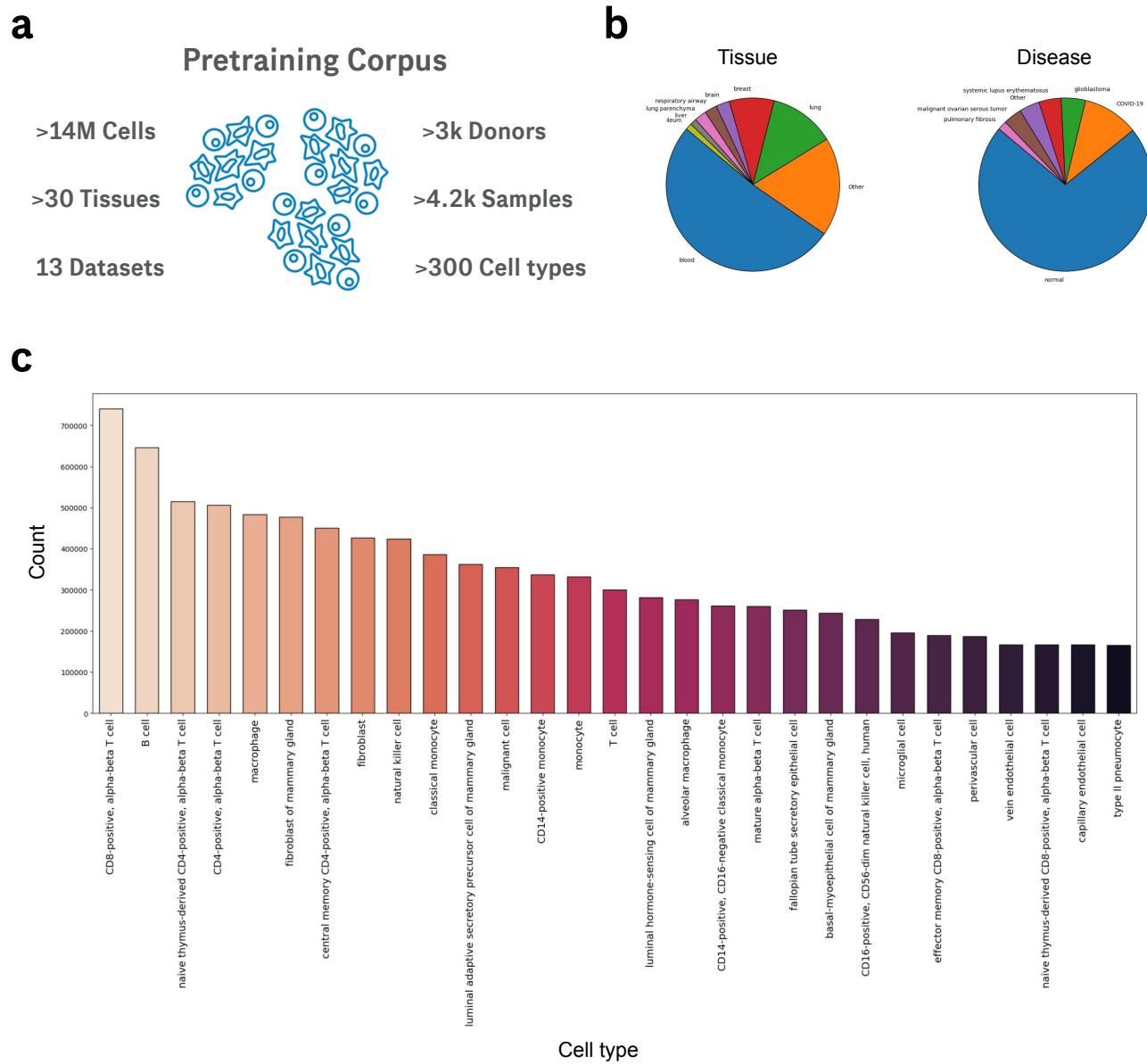


FIGURE S1. Pretraining corpus statistics **(a)** Overall statistics of the pretraining corpus. **(b)** Distribution of tissues and diseases in the pretraining corpus. **(c)** Most common cell type counts present in the pretraining corpus across datasets.

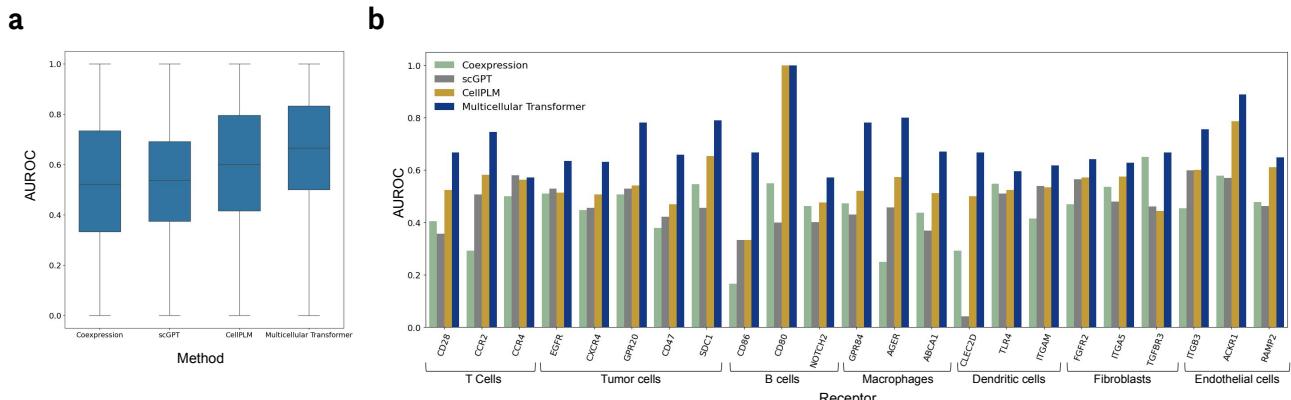


FIGURE S2. Additional results on receptor-ligand identification task **(a)** AUROC across receptors using different methods. We have filtered out receptors with less than 3 ligands to ensure that there is sufficient data for reliable statistical analysis and meaningful comparisons. **(b)** Receptor-ligand prediction performance across a set of diverse receptors specific to different cell types.

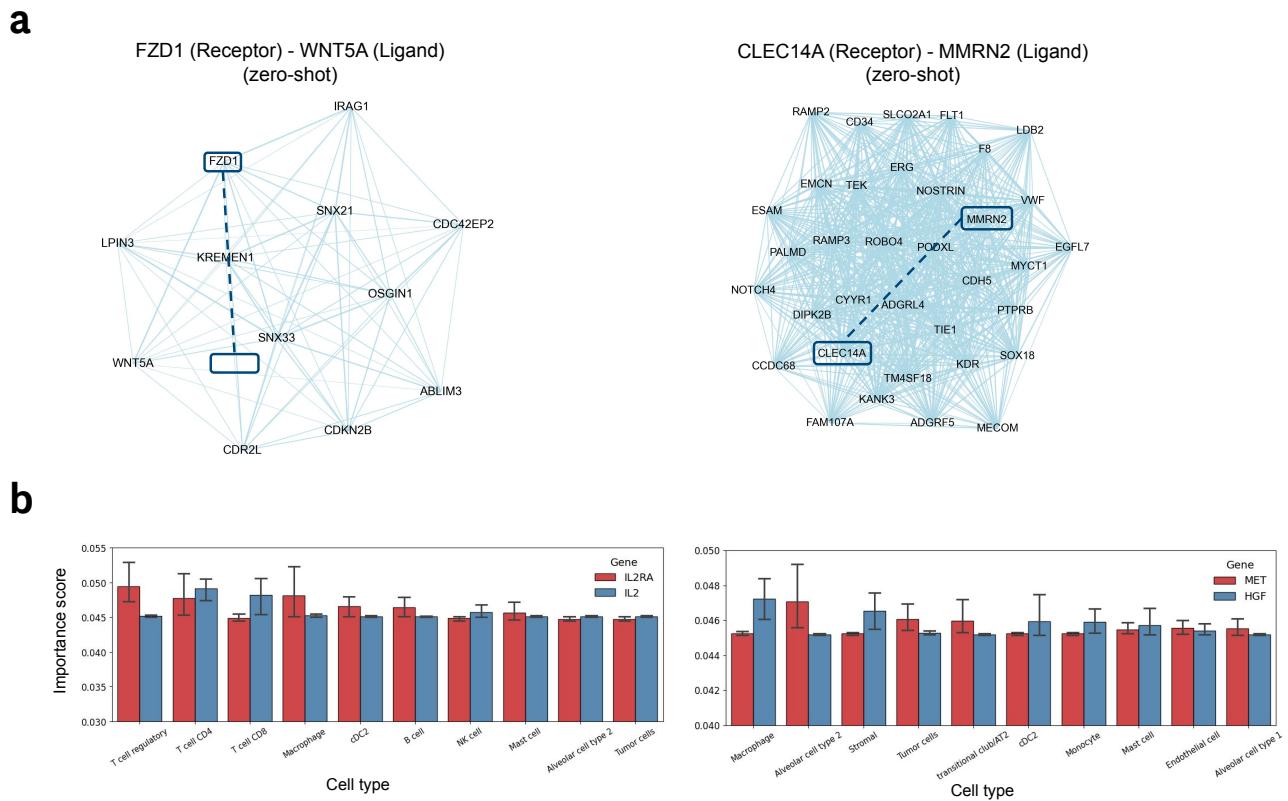


FIGURE S3. Analysis of the zero-shot gene sets with receptor-ligand pairs. **(a)** Examples of the zero-shot gene sets identified by the Multicellular Transformer model at the Leiden resolution of 50 with receptor-ligand pairs. **(b)** Most important cell types for *IL2RA-IL2R* and *MET-HGF* receptor-ligand pairs using integrated gradients (Sundararajan, Taly, and Yan 2017) across validation samples on lung cancer atlas (Salcher et al. 2022). Cell type importance scores have been computed by applying softmax function to aggregated absolute attribution scores across samples. The error bars represent 95% confidence interval. cDC stands for conventional dendritic cells.

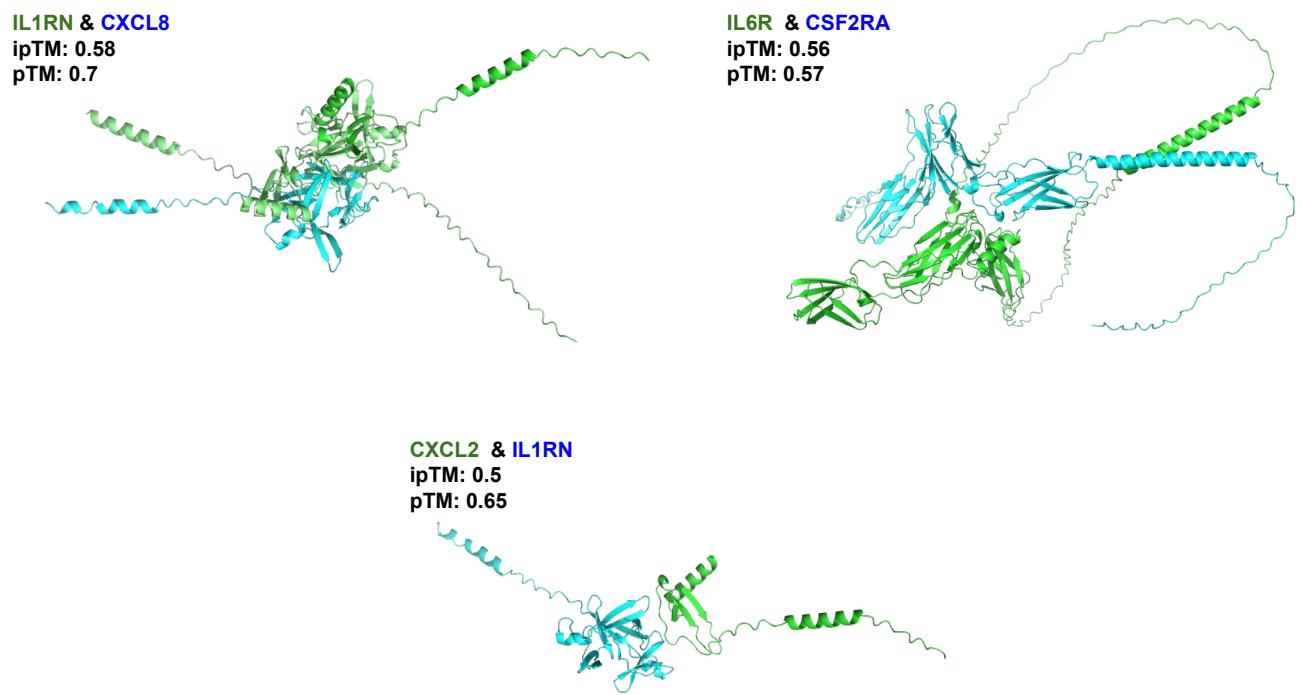


FIGURE S4. Predicted protein complex structures identified with Multicellular Transformer Structures of protein complexes predicted with AlphaFold3 (Abramson et al. 2024) and deemed to interact with Multicellular Transformer. The pTM (predicted Template Modeling) score measure how well different parts of protein fold together, while ipTM (inter-chain pTM) is the adaptation of pTM for assessing the structural confidence of protein-protein interfaces, higher is better.

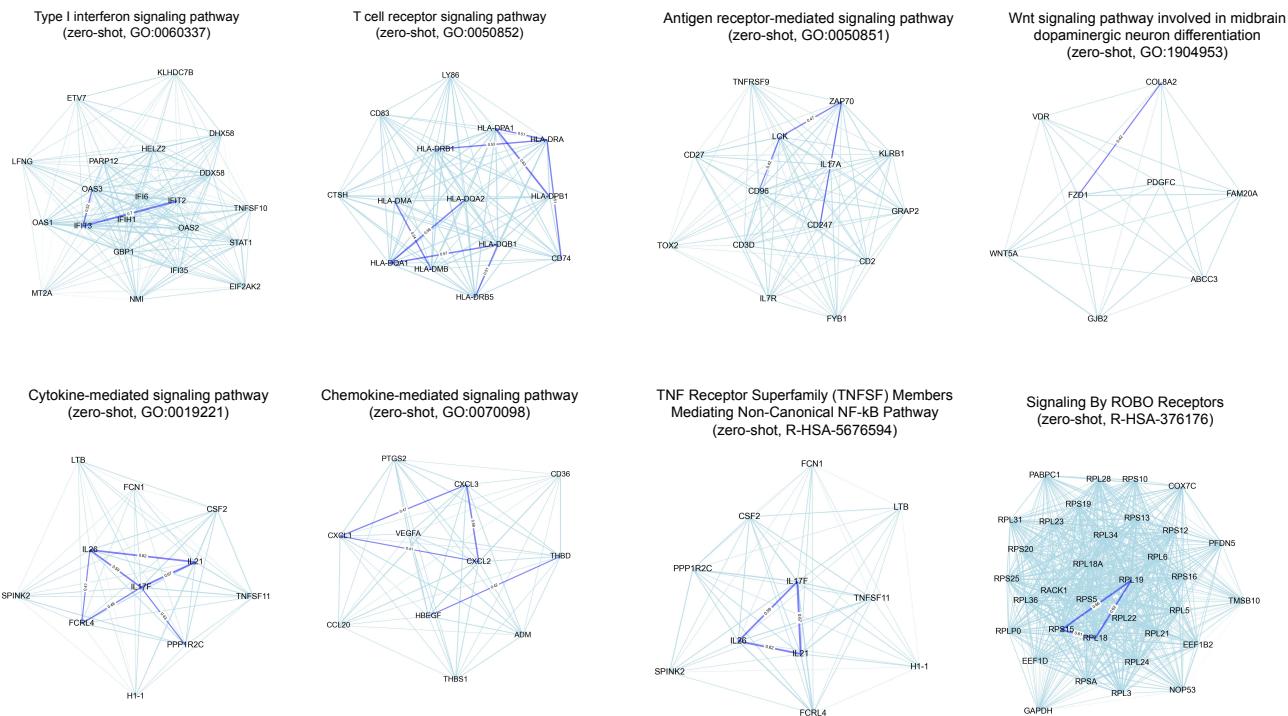


FIGURE S5. Zero-shot gene sets identified with Multicellular Transformer Examples of the zero-shot gene sets identified by the Multicellular Transformer model at the Leiden resolution of 50 together with pathway enriched in the corresponding gene set.

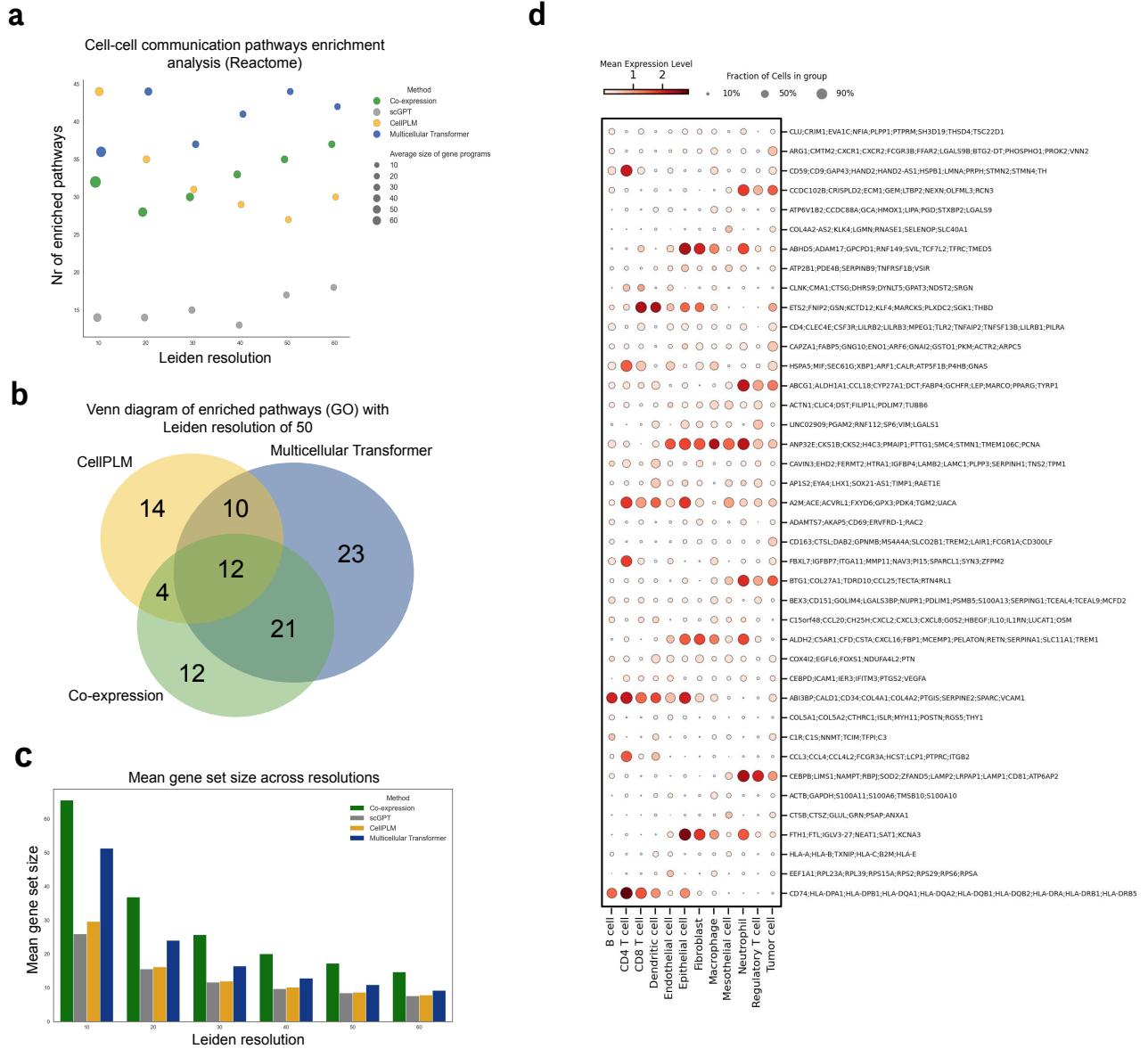


FIGURE S6. Further analysis of the cell-cell communication pathways identified with Multicellular Transformer **(a)** Enrichment of zero-shot gene sets extracted using Multicellular Transformer and alternative methods in cell-cell communication related pathways across Leiden resolutions using Reactome pathways. The genes selected for the analysis were highly variable genes from the tumor microenvironment atlas (Guimarães et al. 2024). The co-expression was computed using the aggregated gene expression at the sample-level using the entire pretraining corpus. **(b)** The Venn diagram comparing the overlap between the zero-shot gene sets enriched in cell-cell communication related pathways using Multicellular Transformer, CellPLM (Wen et al. 2023) and co-expression. **(c)** Mean gene set size across Leiden resolutions and methods. **(d)** Normalized expression of a set of gene sets identified by the Multicellular Transformer at Leiden resolution of 50 across a set of cell types in lung cancer atlas (Salcher et al. 2022).

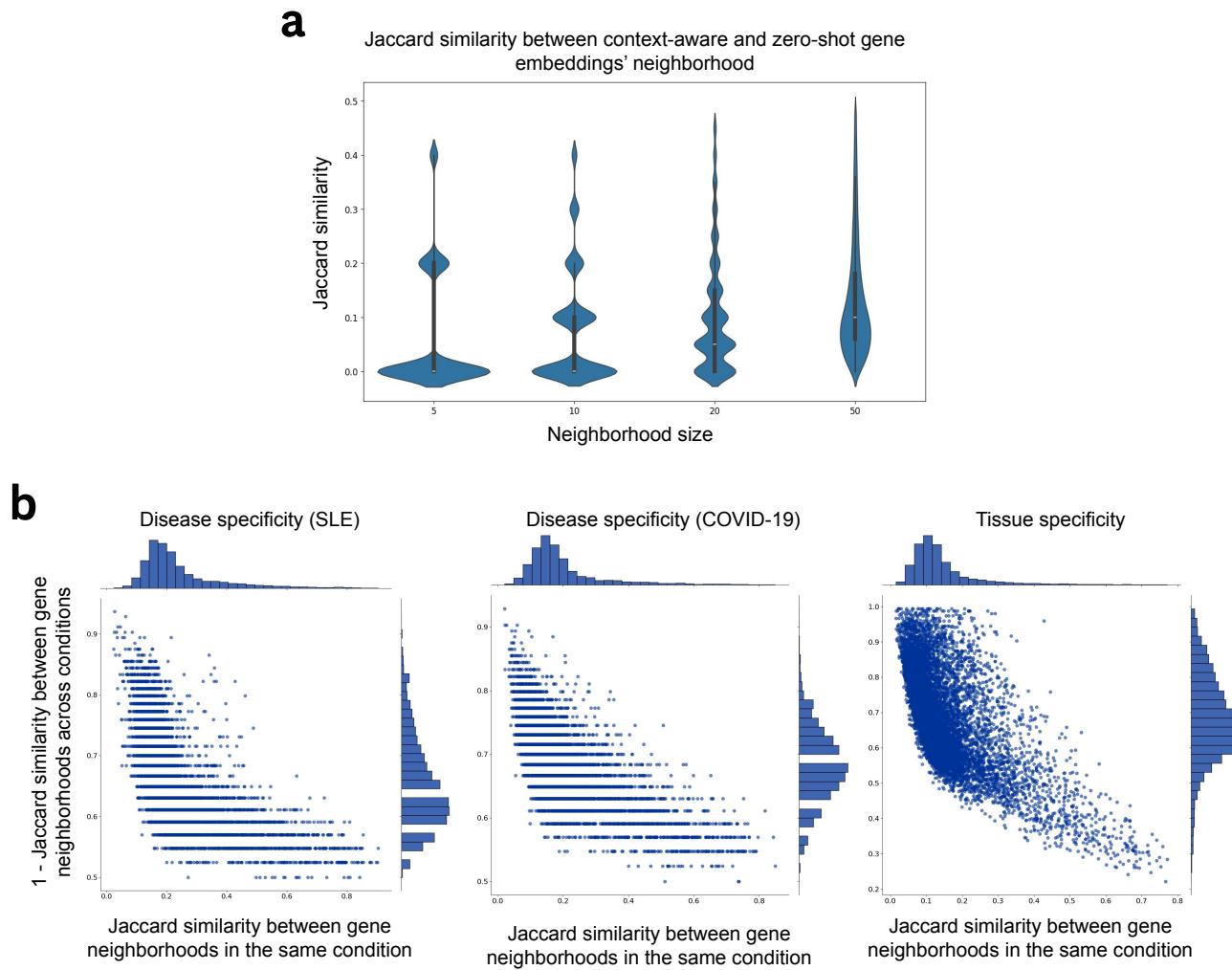


FIGURE S7. Multicellular Transformers identifies condition-specific genes using gene neighborhood across samples. (a) The Jaccard similarity between context-aware and zero-shot gene embeddings' neighborhood at different sizes of the neighborhood across pretraining samples. (b) Pairwise plots showing the mean overlap between gene neighborhoods across healthy vs disease samples in SLE, COVID-19 and tissues. The overlap was calculated by using 20 nearest neighbors for each gene in each sample as measured by cosine similarity between gene embeddings.

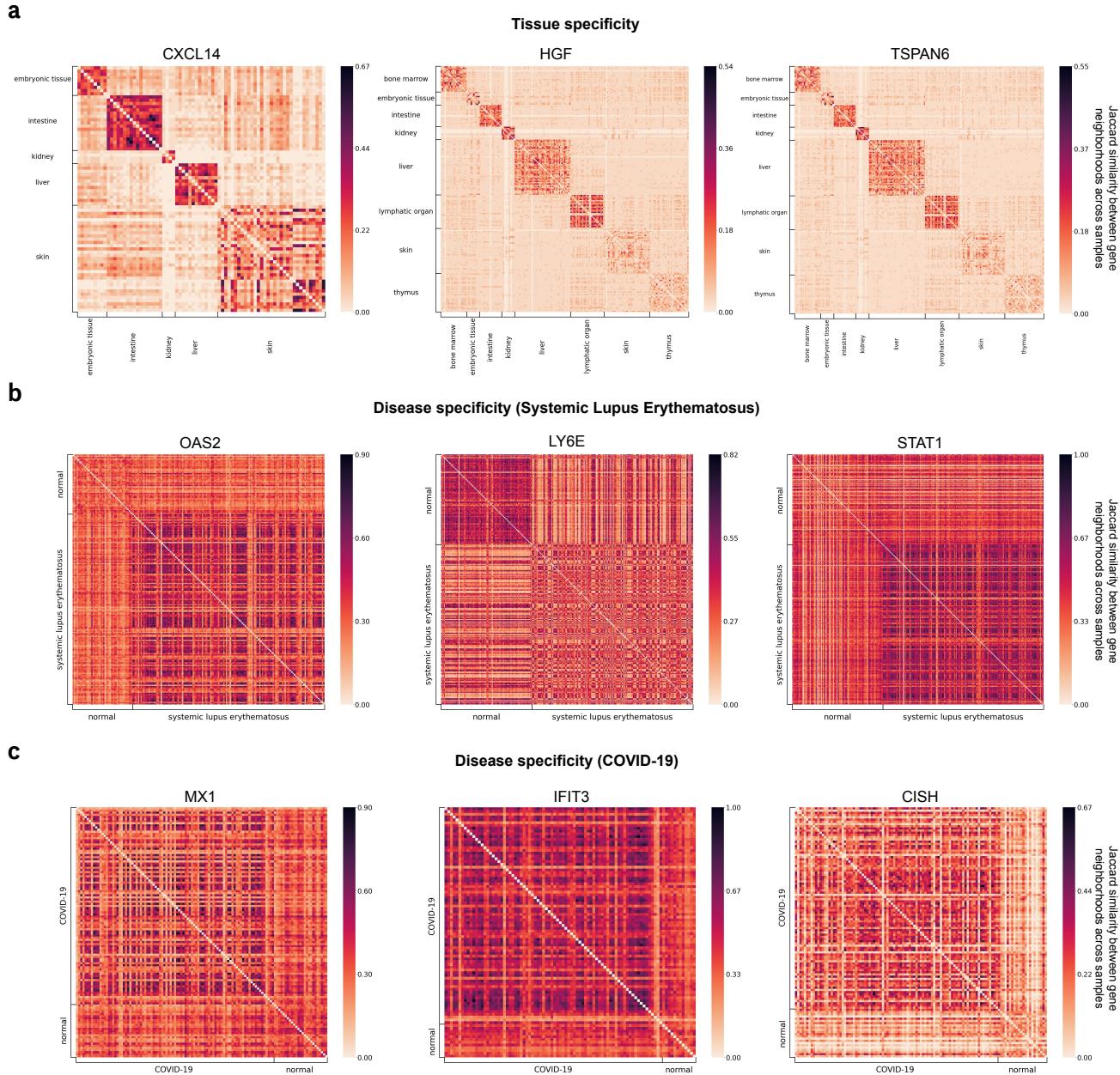


FIGURE S8. Additional analysis of context-aware gene embeddings (a) Genes with high condition-specificity across tissues and diseases (Lupus & COVID-19). The condition-specificity was calculated by dividing the mean overlap between gene neighborhoods in the same condition by the mean overlap between gene neighborhoods across conditions. The datasets used to calculate condition specificity were [Suo et al. \(2022\)](#) for tissue analysis, [Perez et al. \(2022\)](#) for lupus and [Stephenson et al. \(2021\)](#) for COVID-19.

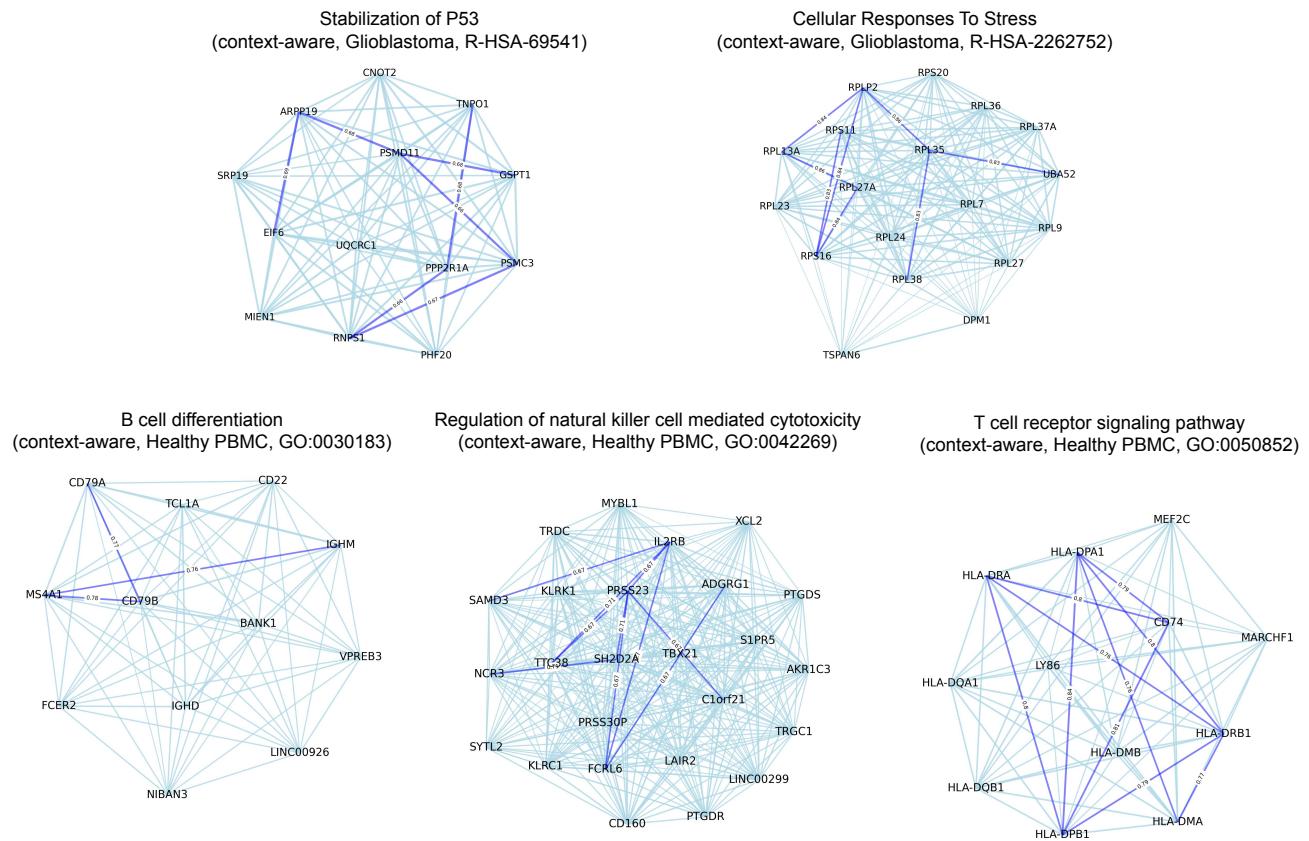


FIGURE S9. Context-aware gene sets identified with Multicellular Transformer gene embeddings. Examples of the context-aware gene sets identified by the Multicellular Transformer model at the Leiden resolution of 50 together with pathway enriched in the corresponding gene set in Glioblastoma (top) and healthy PBMC (bottom) samples. The samples were extracted from Glioblastoma atlas (Ruiz-Moreno et al. 2022) and single-cell eQTL mapping study (Yazar et al. 2022) respectively.

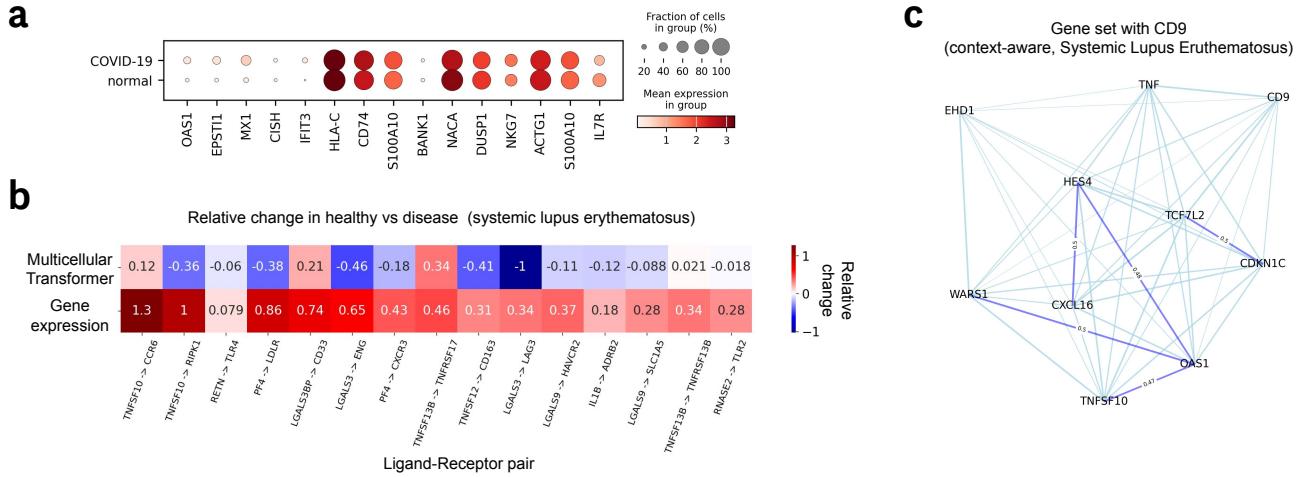


FIGURE S10. Expression of disease-specific genes in COVID-10 identified with Multicellular Transformer & receptor-ligand analysis in systemic lupus erythematosus. (a) The expression of disease-specific genes in COVID-19 (Stephenson et al. 2021) computed using context-aware embeddings from Multicellular Transformer. Many genes have been shown to be associated with the disease but do not exhibit a significant change in expression between healthy and disease. (b) Relative change in cosine similarity (top) vs expression (bottom) in healthy vs disease patients with SLE across a set of known receptor-ligand pairs with highest change according to co-expression in (Perez et al. 2022). (c) Context-aware gene set in SLE (Perez et al. 2022) with CD9 gene. The gene set has been discovered by clustering context-aware gene embeddings from SLE samples at Leiden resolution of 50.

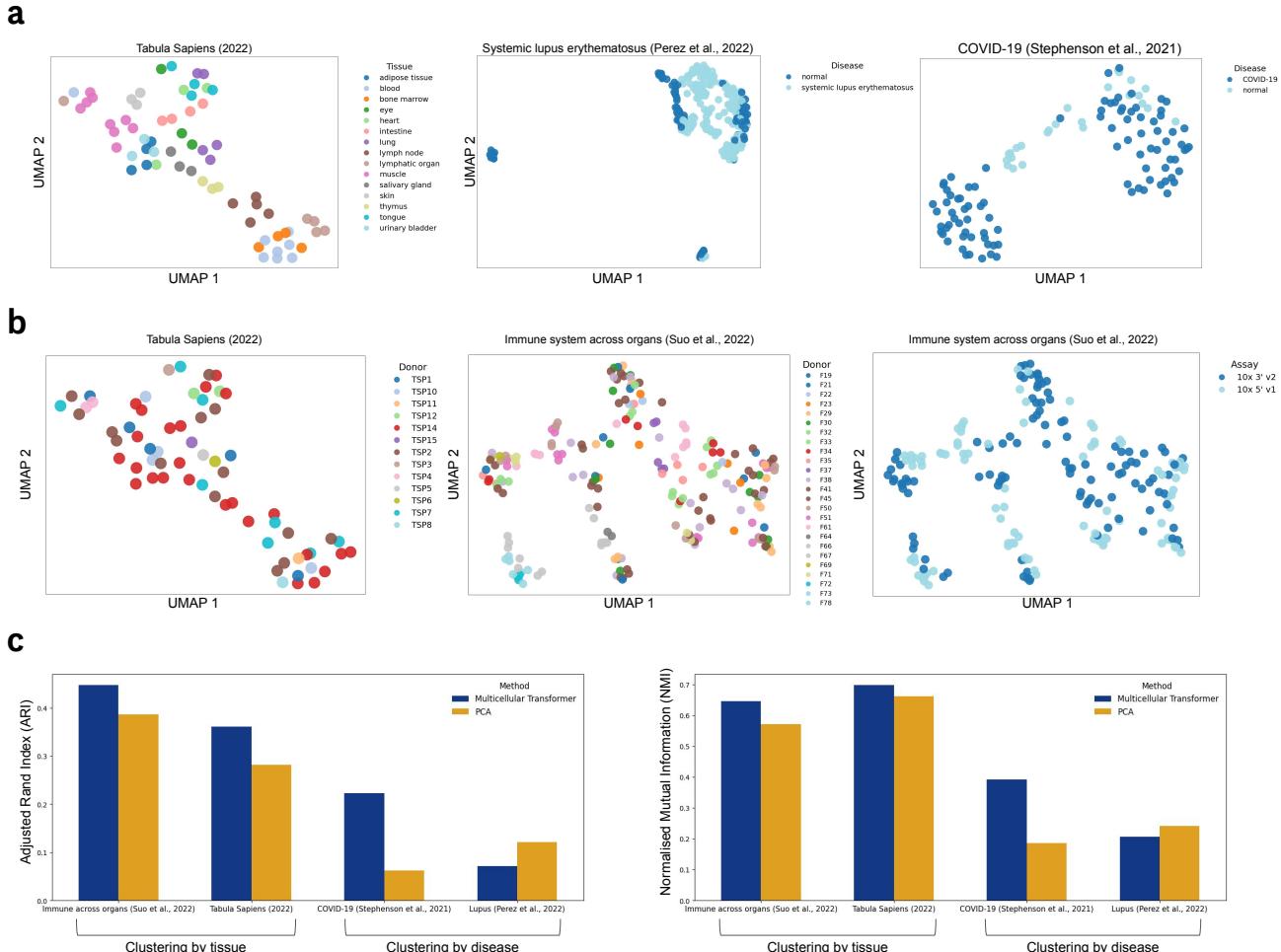


FIGURE S11. Further results on sample-level clustering **(a)** UMAPs of sample-level embeddings across tissues on the Tabula Sapiens (Consortium* et al. 2022) dataset (left) and SLE (Perez et al. 2022) (middle) and COVID-19 (Stephenson et al. 2021). The sample embeddings were extracted by averaging out all of the contextual gene embeddings in the sample. **(b)** UMAPs of sample-level embeddings across donors on the Tabula Sapiens (left) (Consortium* et al. 2022) and across donors (middle) and assays (right) on immune across organs dataset (Suo et al. 2022). The sample embeddings were extracted by averaging out all of the contextual gene embeddings in the sample. **(c)** Performance across clustering metrics of Multicellular Transformer vs PCA baseline. For both Adjusted Rand Index (ARI) and normalized Mutual Information (NMI) higher is better, minimum is 0 and maximum is 1.

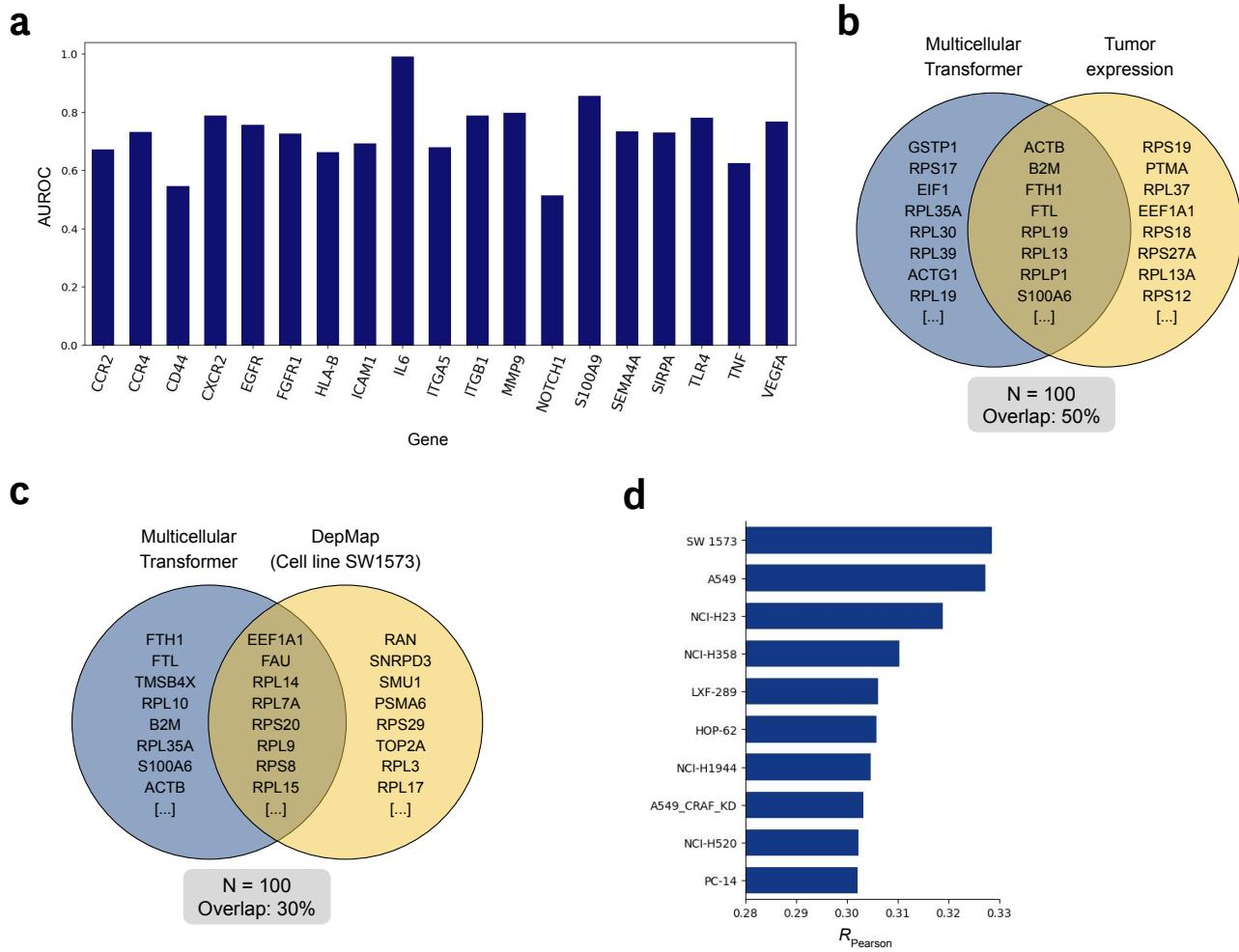


FIGURE S12. *In silico* genetic perturbation effect estimation at cell population level. (a) AUROC per gene on the *in silico* gene knock-out effect on a set of true interacting genes vs control genes. The control genes were randomly selected from a pool of available genes. The knock-out effect is quantified by the maximum log-fold change in gene's logits (Methods). (b) The overlap between genes with highest effect from Multicellular Transformer (i.e. one which leads to highest reduction in the normalized tumor expression) and genes with highest normalized tumor expression. (c) The overlap between genes with highest effect from Multicellular Transformer (i.e. one which leads to highest reduction in the normalized tumor expression) and from viability readouts following CRISPR knock-outs from DepMap (Tsherniak et al. 2017) data across lung cancer cell lines. (d) The lung cancer cell lines with highest pearson correlation between the viability readouts following CRISPR knock-outs from DepMap (Tsherniak et al. 2017) and the predicted normalized effect from Multicellular Transformer.

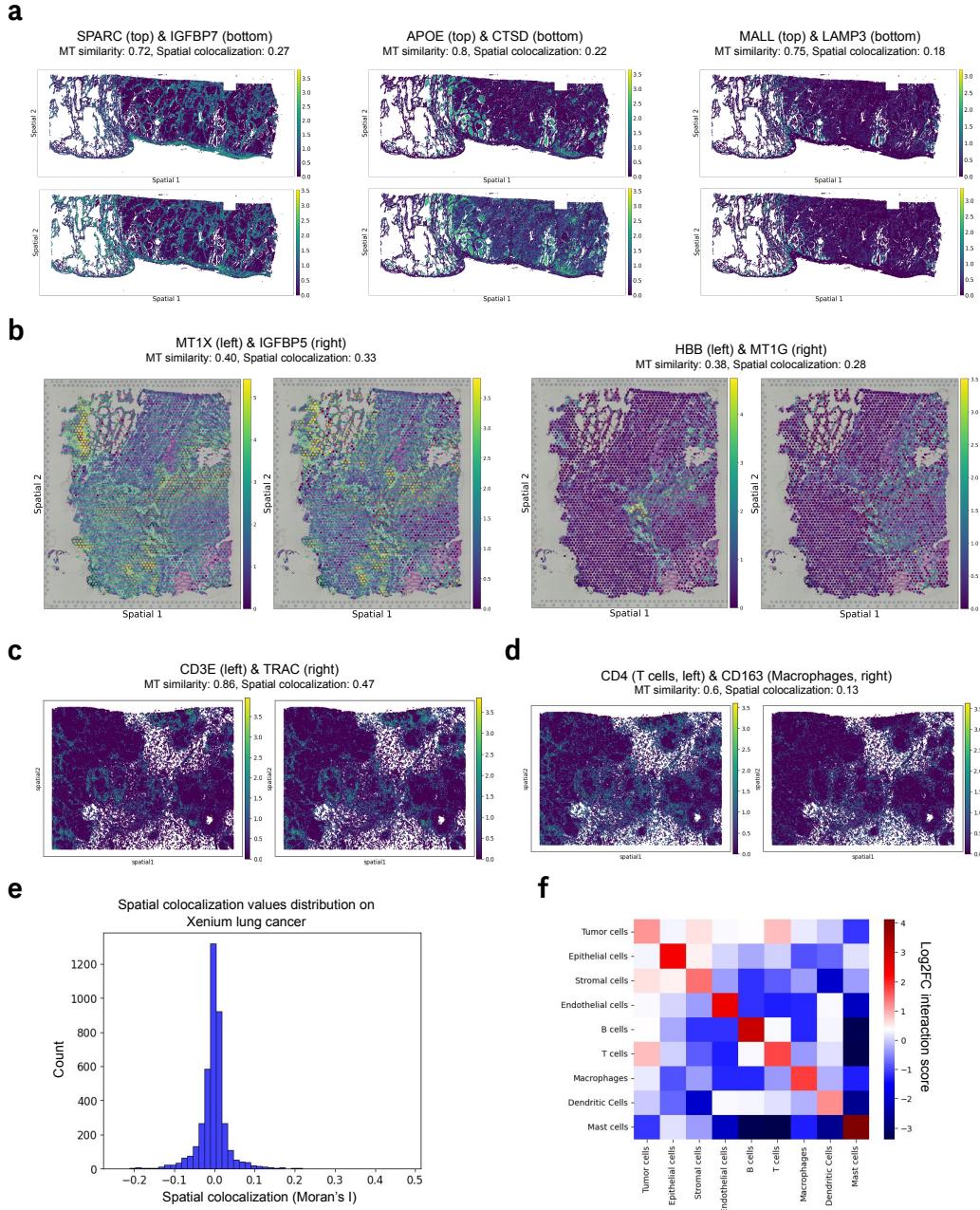


FIGURE S13. Additional results on spatial colocalization patterns across datasets recovered by the Multicellular Transformer **(a)** Spatial colocalization of selected gene pairs in Xenium human lung cancer dataset, recovered by the Multicellular Transformer in a zero-shot fashion by assessing genes with highest cosine similarity in gene embeddings for each gene. The spatial colocalization was measured by computing bivariate Moran's I between two genes. **(b)** Spatial colocalization of selected gene pairs in Visium glioblastoma dataset, recovered by the Multicellular Transformer in a zero-shot fashion by assessing genes with highest cosine similarity in gene embeddings for each gene. The spatial colocalization was measured by computing bivariate Moran's I between two genes. **(c)** Spatial colocalization of CD3E & TRAC genes in Xenium breast cancer dataset, recovered by the Multicellular Transformer fine-tuned on the dataset by assessing genes with highest cosine similarity in the gene embedding space. **(d)** Spatial colocalization of CD4 & CD163 genes in Xenium breast cancer dataset, recovered by the Multicellular Transformer fine-tuned on the dataset by assessing T cell-Macrophage gene pairs with highest cosine similarity in the gene embedding space. **(e)** The distribution of spatial colocalization values as measured by bivariate Moran's I on Xenium lung cancer dataset. **(f)** Interaction matrix between cell types computed using the gene co-expression between marker genes on the Xenium human breast cancer dataset. The interaction score was computed by calculating the \log_2 fold change between the gene pair's co-expression divided by its average co-expression across all genes and taking the average across gene pairs for each cell type-cell type pair.

Study	# Cells	# Samples
AIDA Phase 1 Data Freeze v1: Chinese, Indian, Japanese, Korean, and Malay donors in Japan, Singapore, and South Korea (Kock et al. 2024)	1,058,909	508
COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas (Ren et al. 2021)	1,462,702	284
Single-cell multi-omics analysis of the immune response in COVID-19 (Stephenson et al. 2021)	647,366	143
Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma (Ruiz-Moreno et al. 2022)	1,060,257	191
Human breast cell atlas (Reed et al. 2024)	2,122,065	351
Cells of the adult human heart (Litviňuková et al. 2020)	486,134	88
Cells of the human intestinal tract mapped across space and time (Elmentait et al. 2021)	428,469	159
The integrated Human Lung Cell Atlas (Sikkema et al. 2023)	2,282,120	731
Mapping the developing human immune system across organs (Suo et al. 2022)	908,046	221
Single-cell RNA-seq reveals the cell-type-specific molecular and genetic associations to lupus (Perez et al. 2022)	1,263,676	274
MSK SPECTRUM – Ovarian cancer mutational processes drive site-specific immune evasion (Vázquez-García et al. 2022)	929,690	146
Single-cell eQTL mapping identifies cell type specific genetic control of autoimmune disease (Yazar et al. 2022)	1,248,980	981
Tabula Sapiens (Consortium* et al. 2022)	483,152	129
Total	14,381,566	4,206

TABLE S1. Summary of studies used in pretraining corpus and their respective cell and sample counts.

Cell Type	Marker Genes
Tumor cells	ABCC11, AGR3, DSP, ERBB2, FOXA1, GATA3, S100A14, SERPINA3, TACSTD2, AGR3, CEACAM6, MZB1, SERPINA3, FASN
Epithelial cells	ACTA2, DST, KRT14, KRT5, MYLK, KRT15, KRT23, PTN, SFRP1
Stromal cells	ALDH1A3, GJB2, LUM, POSTN, SFRP4
Endothelial cells	AQP1, EGFL7, PECAM1, VWF
B cells	BANK1, CD79A, MS4A1
T cells	CCL5, CD4, CD8A, CXCR4, CYTIP, IL7R, LTB, TRAC
Macrophages	APOC1, C15orf48, C1QA, C1QC, CD14, CD163, CD68, FGL2, ITGAX, MMP12
Dendritic Cells	CCR7, CD83, IL3RA, LILRA4, PLD4
Mast cells	CPA3, CTSG, KIT

TABLE S2. Marker genes for each cell type used in Xenium breast cancer analysis.