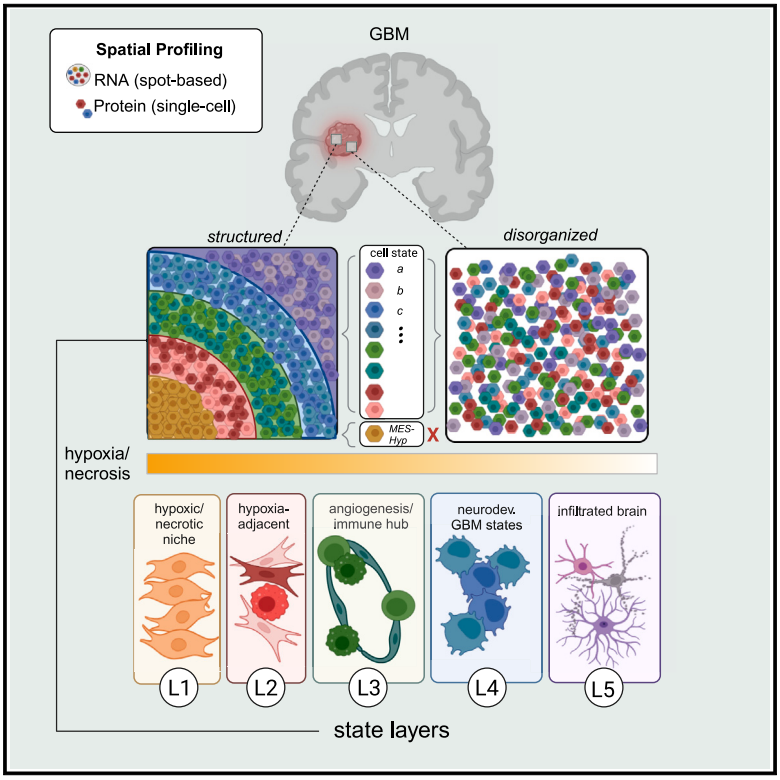
Cell

Integrative spatial analysis reveals a multi-layered organization of glioblastoma

Graphical abstract



Authors

Alissa C. Greenwald, Noam Galili Darnell, Rouven Hoefflin, ..., Marian C. Neidert, Mario L. Suvà, Itay Tirosh

Correspondence

suva.mario@mgh.harvard.edu (M.L.S.), itay.tirosh@weizmann.ac.il (I.T.)

In brief

Spatial mapping shows that glioblastoma consists of both disorganized and structured regions. Structured regions display a five-layered organization that is associated with hypoxia and extends beyond the organization that is visible by histology.

Highlights

- Human glioblastoma consists of disorganized and structured regions
- Organization of cell states extends beyond the structure observed by histopathology
- Organization of cell states is associated with the abundance of hypoxic cancer cells
- Recurring pairs of states define a five-layer organization in the structured regions

Greenwald et al., 2024, Cell *187*, 1–17
May 9, 2024
https://doi.org/10.1016/j.cell.2024.03.029

CellPress

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

Cell

CellPress

Article

Integrative spatial analysis reveals a multi-layered organization of glioblastoma

Alissa C. Greenwald,[1,13] Noam Galili Darnell,[1,13] Rouven Hoefflin,[1,2,13] Dor Simkin,[1] Christopher W. Mount,[3,4] L. Nicolas Gonzalez Castro,[3,4,5,6] Yotam Harnik,[1] Sydney Dumont,[3,4] Dana Hirsch,[7] Masashi Nomura,[3,4] Tom Talpir,[8] Merav Kedmi,[9] Inna Goliand,[9] Gioele Medici,[10] Julie Laffy,[1] Baoguo Li,[11] Vamsi Mangena,[3,4] Hadas Keren-Shaul,[9] Michael Weller,[10] Yoseph Addadi,[9] Marian C. Neidert,[10,12] Mario L. Suvà,[3,4,*] and Itay Tirosh[1,14,*]

[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel
[2]Department of Medicine I, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany
[3]Department of Pathology, Center for Cancer Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
[4]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[5]Center for Neuro-Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[6]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA
[7]Immunohistochemistry Unit, Department of Veterinary Resources, Weizmann Institute of Science, Rehovot, Israel
[8]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel
[9]Department of Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot, Israel
[10]Clinical Neuroscience Center, Department of Neurosurgery, University Hospital Zurich, University of Zurich, Zurich, Switzerland
[11]Department of Systems Immunology, Weizmann Institute of Science, Rehovot, Israel
[12]Department of Neurosurgery, Cantonal Hospital St. Gallen, St. Gallen, Switzerland
[13]These authors contributed equally
[14]Lead contact
*Correspondence: suva.mario@mgh.harvard.edu (M.L.S.), itay.tirosh@weizmann.ac.il (I.T.)
https://doi.org/10.1016/j.cell.2024.03.029

SUMMARY

Glioma contains malignant cells in diverse states. Here, we combine spatial transcriptomics, spatial proteomics, and computational approaches to define glioma cellular states and uncover their organization. We find three prominent modes of organization. First, gliomas are composed of small local environments, each typically enriched with one major cellular state. Second, specific pairs of states preferentially reside in proximity across multiple scales. This pairing of states is consistent across tumors. Third, these pairwise interactions collectively define a global architecture composed of five layers. Hypoxia appears to drive the layers, as it is associated with a long-range organization that includes all cancer cell states. Accordingly, tumor regions distant from any hypoxic/necrotic foci and tumors that lack hypoxia such as low-grade IDH-mutant glioma are less organized. In summary, we provide a conceptual framework for the organization of cellular states in glioma, highlighting hypoxia as a long-range tissue organizer.

INTRODUCTION

In his landmark 1932 paper, Percival Bailey elegantly characterized the spatial architecture of gliomas observing that "the microscopic structure of tumors of the brain is infinitely varied, yet among their kaleidoscopic appearances certain family resemblances may be traced…".[1] Gliomas are typically highly heterogeneous and infiltrative, yet specific spatial patterns are consistently noted. In 1938, Hans Joachim Scherer detailed the recurring spatial structures of gliomas and, in particular, glioblastomas (GBMs): organized secondary structures characteristic of invasion in which glioma cells rely on or mimic the pre-existing normal structures of the brain; amorphous arrangements of cancer cells; proper structures in which the cancer cells form patterns that do not depend on pre-existing normal structures; and a mesenchymal tissue response surrounding areas of necrosis.[2] Accordingly, foci of necrosis (and hypoxia) as well as areas around neuronal structures of the normal brain are associated with local organization, while outside of these regions, glioma tissues are dominated by a lack of histological organization.[3]

Technological developments now enable us to revisit the spatial architecture captured by classical histopathology and describe it quantitatively with a granular, genome-wide analysis of cell types and cellular states. The first step toward this goal has been to define the diverse cellular components within glioma by single-cell RNA sequencing (scRNA-seq). In prior studies that required tissue dissociation, we defined four major states of isocitrate dehydrogenase (IDH)-wildtype GBM cancer cells: neural progenitor-like (NPC), oligodendrocyte progenitor-like (OPC), astrocyte-like (AC), and mesenchymal-like (MES).[4] We conducted similar efforts for IDH-mutant glioma, defining three major

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

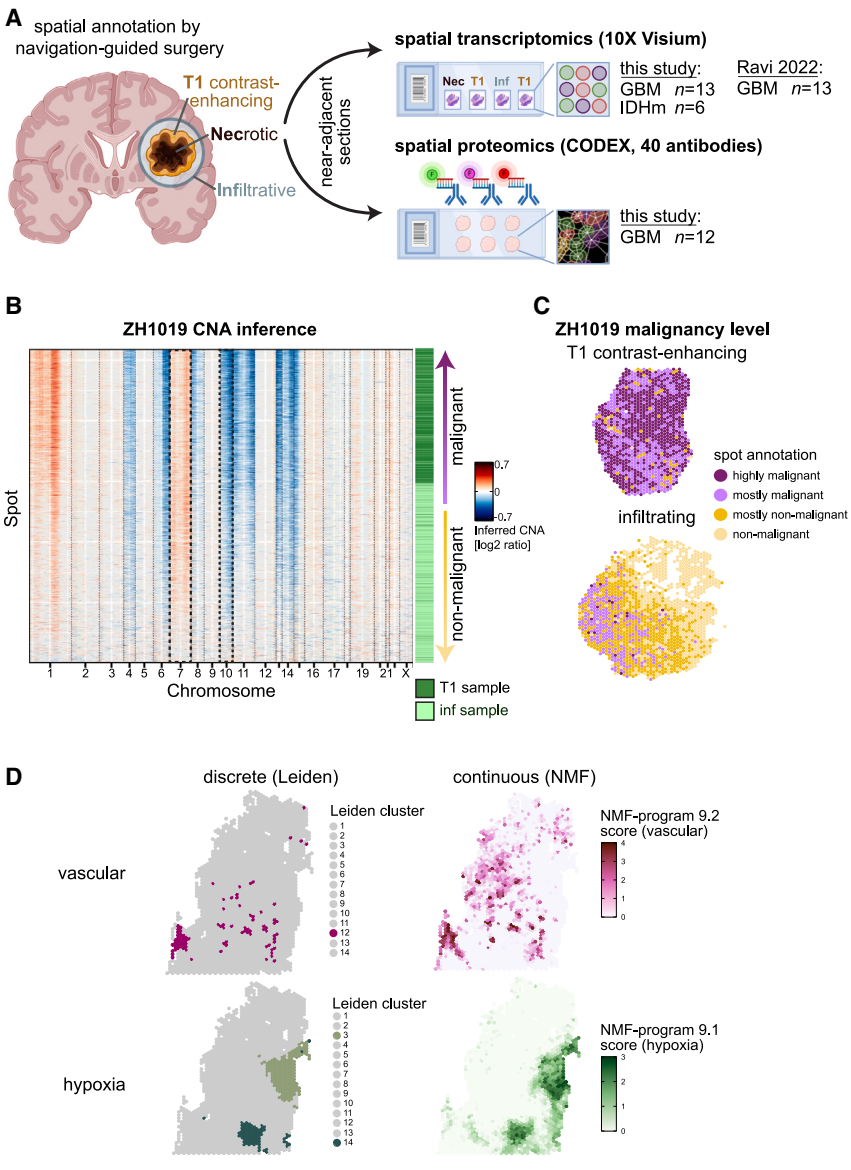


Figure 1. Experimental design and spot classification per sample

(A) Experimental design and patient cohort. Fresh frozen tissue sections from GBM (n = 13) and IDH-mutant gliomas (n = 6) were profiled by 10X Visium. CODEX was performed on 12 near-adjacent GBM tissue sections with a panel of 40 antibodies. Four tumors were spatially annotated by the neurosurgeon during navigation-guided surgery. This GBM Visium cohort was combined with an external GBM Visium cohort[9] for joint analysis. Scheme created with BioRender.com.

(B) Copy-number aberrations (CNAs) were inferred by average relative expression in sliding windows of 150 analyzed genes after sorting genes by their chromosomal location. ZH1019, from which we profiled both an infiltrating sample and a T1-contrast enhancing sample, is shown here as an example. Rows correspond to spots arranged by malignancy level as inferred from CNA; columns correspond to genes arranged by chromosomal position. Annotation bars correspond to the region from which the spot was derived (T1 or infiltrating) and the malignancy level.

(C) Spatial maps of ZH1019 T1 contrast-enhancing and ZH1019 infiltrating samples with spots annotated by malignancy level as described in (B).

(D) Per-sample Leiden (left) and NMF clustering (right) of ZH916 bulk for vascular and hypoxia clusters (all other clusters are shown in gray).

See also Figure S1.

malignant cell states[5,6] as well as the landscape of glioma immune cells[7,8] across glioma subtypes. The next step requires *in situ* spatial mapping of these defined cellular components within the glioma ecosystem. Multiple methods emerged recently for spatial transcriptomics (e.g., Visium) and proteomics (e.g., CODEX), enabling such mapping.

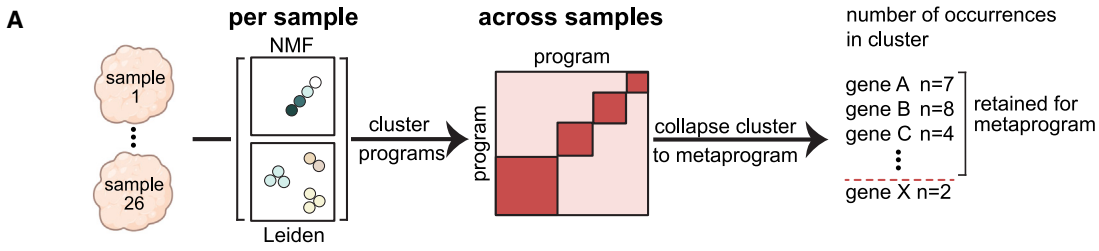
The first study of GBM by Visium addressed this challenge by defining spatial neighborhoods, but without resolving them to individual cell states and their organization.[9] Ravi et al. observed spatial segregation of neurodevelopmental neighborhoods from hypoxia-associated mesenchymal neighborhoods and inflammation-associated neighborhoods, suggesting that the local tumor microenvironment (TME) is a major determinant of cancer cell states.[9] Additional studies identified specific spatial interactions, including induction of the MES-like state by macrophages,[8] and synaptic crosstalk between OPC-like GBM cancer cells and neurons via glutamatergic signaling and AMPARs.[10] However, many open questions remain concerning how the cell states and cell types relate to each other and to other elements of the microenvironment. Given their remarkable spatial diversity, are there consistent patterns or rules of organization across gliomas? To what degree does spatial location dictate the diversity of cellular states? To address these questions and others, we profiled 19 glioma samples by 10X Visium and 12 by CODEX, integrated them with published GBM Visium data,[9] and developed a quantitative framework to systematically describe recurring features of the spatial organization of glioma.

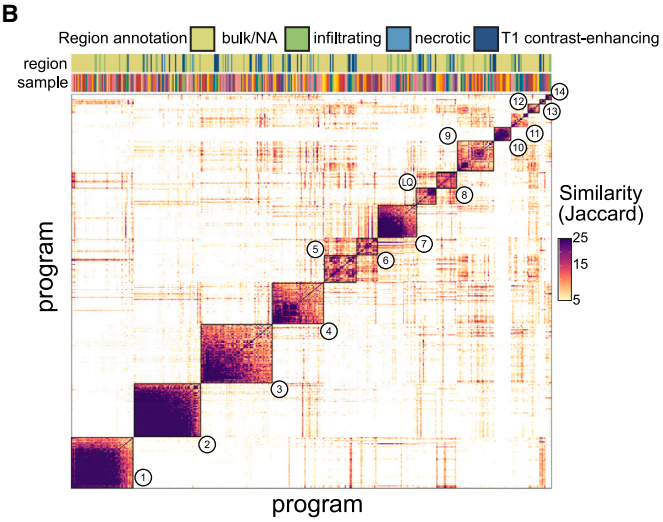
RESULTS

Spatial transcriptomics of glioma

We used the 10X Visium spatial transcriptomics platform to spatially profile gliomas by RNA-seq (Figure 1A).[11,12] The diameter of a capture spot is 55 μm, such that each spot contains a mixture of 1–35 cells, with a median of 8 cells in GBM based on image analysis (Figure S1A). In four cases, we analyzed multiple frozen tissue blocks from the same tumor, each isolated from a different region and annotated as *necrotic*, *infiltrating*, or *T1 contrast-enhancing* by navigation-guided tumor resection.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

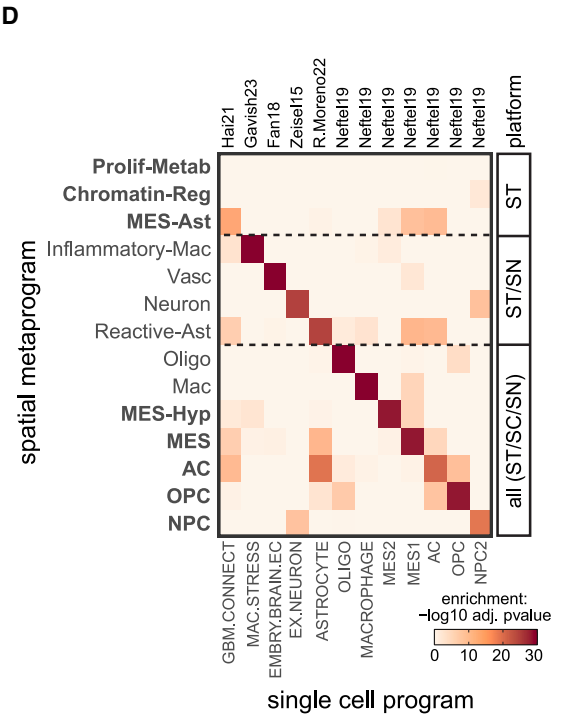


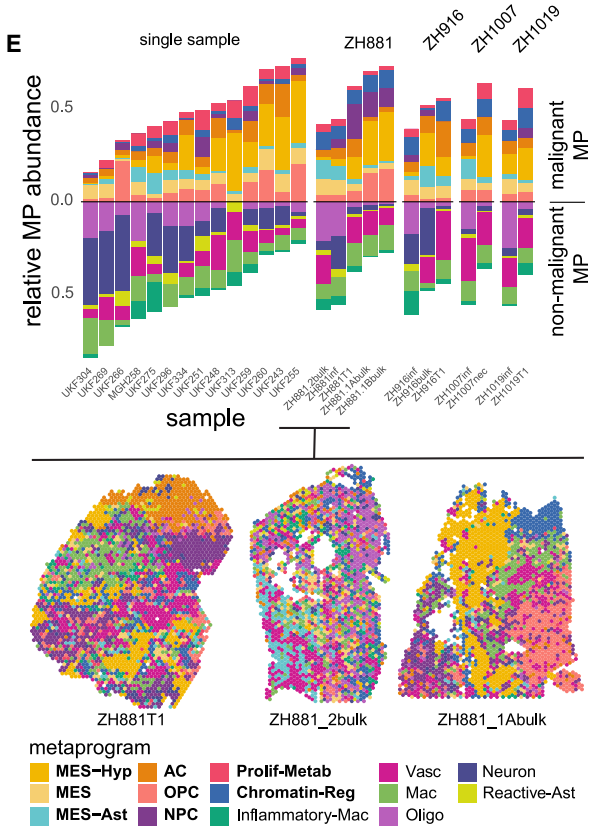


C

clust	metaprogram	selected genes
1	Neuron	*SNAP25, SYT1, SLC17A7*
2	Vasc	*VWF, PDGFRB, ESAM*
3	**MES-Hyp (malig.)**	*VEGFA, NDRG1, PGK1*
4	Mac	*CD14, CD74, FCGR3A*
5	**AC (malig.)**	*S100B, AQP1, TTYH1*
6	**OPC (malig.)**	*SOX10, BCAN, BCAS1*
7	Oligo	*PLP1, MAG, CLDND1*
8*	**Chromatin-Reg (malig.)**	*ATRX, NFIX, CCNI, KMT2E, SOX4*
LQ	low quality (excluded)	*MALAT1, MT-C03, MT-ATP8*
9	**MES (malig.)**	*EFEMP1, CHI3L1, CD44*
10*	**Prolif-Metab (malig.)**	*CNTD1, SLC16A1, CTNNB1, SOX9*
11*	**MES-Ast (malig.)**	*GAP43, KCNF1, HOPX, FABP7*
12	Reactive-Ast	*AQP4, GJA1, GSN, HLA-DRA*
13	**NPC (malig.)**	*INA, NELL2, DCX, STMN2*
14	Inflammatory-Mac	*CCL2, IL1B, CCL3, STAT1*

*unique to spatial





(legend on next page)

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

We profiled 13 GBM sections and 6 IDH-mutant glioma sections (three 1p/19q co-deleted oligodendroglioma and three astrocytoma). These samples were integrated with 13 GBM sections from an external Visium dataset,[9] retaining a total of 70,618 spots after quality control (Table S1, STAR Methods).

We used two initial approaches to classify and annotate spots. First, we performed copy-number aberration (CNA) inference by averaging the expression of genes in each chromosomal region to classify spots as mostly malignant, mostly non-malignant, or mixed (STAR Methods).[5,13] As expected, inferred CNAs included the hallmarks of GBM (chromosome 7 gain and chromosome 10 deletion) and oligodendroglioma (chromosome 1p/19q co-deletion) and were significantly associated with cancer-rich regions (e.g., tumor core vs. infiltrating regions) (Figures 1B, 1C, S1B, and S1C). Second, we clustered spots by Leiden clustering both per sample and jointly per tumor in cases with multiple sections and annotated the clusters by gene set enrichment analysis (Figures 1D and S1D).

While the per-sample Leiden clustering discretely groups together spots of similar composition within samples, we next aimed to better capture the continuous nature of cellular states. To this end, we performed non-negative matrix factorization (NMF) per sample and derived robust expression programs that were consistently detected across multiple parameters (Figures 1D and S1E, STAR Methods).[14,15]

Recurrent patterns of expression heterogeneity across gliomas

In order to define expression programs that reflect core patterns of expression heterogeneity, we compared the 492 gene expression programs identified in individual GBM samples either by Leiden clustering or by NMF. We found considerable similarities that allowed us to define 14 clusters of programs based on their gene identity overlap, each covering programs derived from multiple GBM samples (Figures 2A and 2B, STAR Methods). For each group of programs, we defined a consensus program of 50 genes, termed a metaprogram (MP), that reflects a recurrent pattern of heterogeneity in GBM (Figures 2C and S1E, malignant MPs are in bold throughout figures). Similarly, we defined six MPs in IDH-mutant glioma (Figures S2A and S2B; Table S2, STAR Methods).

While expression profiles of spots usually reflect combinations of states, the MPs are highly similar to single-cell states (Figures 2D and S2C). The ability of our approach to highlight individual states is further supported by simulations (Figures S2D and S2E, STAR Methods) and is linked to two features —the use of NMF to derive our programs, and the definition of MPs as a consensus across programs that are derived from many samples. As a toy example, consider multiple spots, each containing exactly 8 cells, with a certain cell state X encompassing (1, 2,…, and 8) cells of those spots, respectively. Clustering the spots and defining the profile of each cluster would lead to a cluster with enrichment of X cells but with an expression profile that also contains some influence of other cell states. However, NMF is designed to detect the main profiles that coherently vary across the spots and hence tends to detect only X as one of the factors. If X correlates with a second cell state, Y, then NMF may identify combinations of X and Y. However, only if X and Y are consistently highly correlated across all samples then we would expect the MP to reflect their combination, while if this correlation is lower in some samples, then the consensus MP would highlight shared NMF genes (reflecting X) over sample-specific genes (that include Y). Our MP approach differs from earlier work that focused on spatially defined programs, which often contain a mixture of cell types (Figure S2F)[9] and from deconvolution of spot-based spatial data with paired sc/snRNA-seq,[19–23] which is well suited for capturing sample-specific cell states but is sensitive to platform-dependent differences such as dissociation biases against some cell states and cell types.

We identified 14 GBM spatial MPs, including eight malignant and six non-malignant programs, each reflecting a cancer cell state or non-malignant cell type (Figures 2C and 2D; Table S2). Non-malignant MPs included Mac (macrophage/microglia) and Inflammatory-Mac (inflammatory macrophage/neutrophil), Oligo (oligodendrocyte), Vasc (endothelial cells and pericytes), Neuron, and Reactive-Ast (reactive astrocyte). The latter included classical astrocytic markers (e.g., *AGT* and *GJA1*) and additional markers suggesting a reactive astrocytic state (e.g., metallothioneins). Of the eight malignant MPs, five directly map to the single-cell GBM states: MES-hypoxia (MES2), MES (MES1), NPC-like, OPC-like, and AC-like (Figures 2D and S2G). As expected, the neurodevelopmental-related malignant MPs (NPC-like, OPC-like, and AC-like) had high gene overlap with signatures of the respective non-malignant cell-type signatures, as also seen for the respective MPs derived from scRNA-seq (Figure S2H).

The three additional malignant spatial MPs include: (1) an astrocytic-like mesenchymal MP (MES-Ast) with enrichment of genes associated with glioma tumor microtubes (e.g., *GAP43, KCNF1*, and *PTN*) (Figure S2I);[24–26] (2) proliferation

Figure 2. Deriving and annotating spatial metaprograms

(A) Scheme of metaprogram (MP) generation approach created with BioRender.com. Each sample is clustered individually by Leiden and NMF. All Leiden and all NMF programs across all samples are clustered by their gene overlap. Each cluster is collapsed to a consensus MP by selecting for the most recurrent genes across programs within the cluster.

(B) Similarity matrix based on gene identity overlap (quantified by Jaccard index), for all programs derived from NMF and Leiden clusters. Programs are annotated by sample and region identity. Cluster numbers correspond to the table in (C).

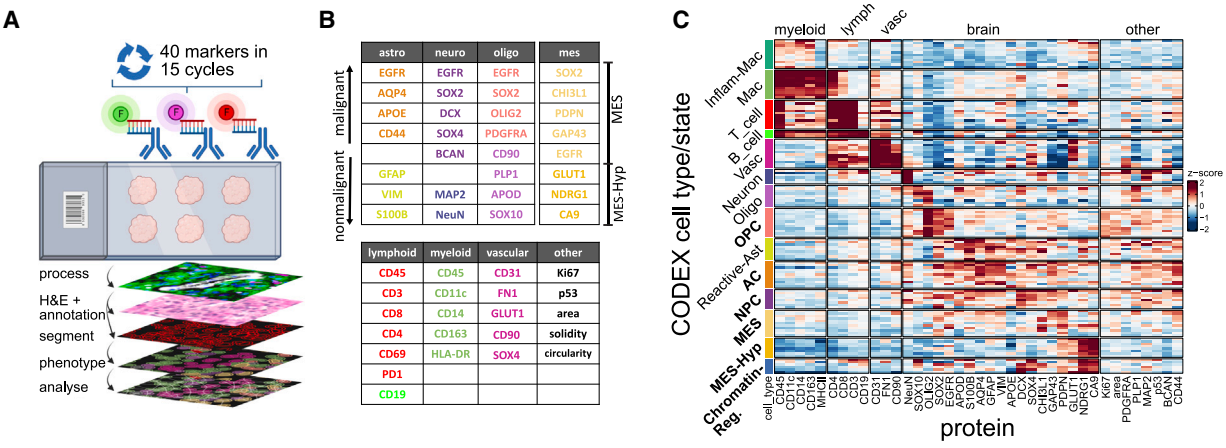
(C) Table of metaprogram names and selected genes corresponding to clusters numbered in (B). Malignant metaprograms are depicted in bold.

(D) Enrichments of spatial metaprograms (rows) with gene sets (columns) previously defined from studies indicated at the top.[4,9,15–18] Enrichment calculated by hypergeometric test ($-\log_{10}$ FDR-adjusted p values).

(E) For each sample (represented by one bar), the proportions of spots assigned to each MP are shown. Four tumors with multiple sections are at the right and are grouped by tumor. Below are spatial MP maps of several ZH881 tissue sections, demonstrating differences in composition and spatial organization between sections isolated from the same tumor.

See also Figures S2 and S3.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029



(legend on next page)

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

and metabolism (Prolif-Metab), enriched with proliferation-related (e.g., *CTNNB1, CNTD1,* and *TP53*) and metabolism (e.g., *SLC16A1* [MCT1], *GGCX*, and *PHGK1*) genes; and (3) chromatin regulation (Chromatin-Reg), enriched with chromatin and transcriptional regulators (e.g., *ATRX, KMT2E, BRD4*, and *SOX4*), as well as with NPC-related genes (Figure S2J). Re-analysis of GBM scRNA-seq data supports these MPs as representing rare cellular states with partial similarity to previously defined states (Figures S2K, S2L, and S3A) and further shows that MES-Ast represents a unique state and not the simple combination of colocalized MES-like and AC-like cancer cells (Figure S3B).

Next, we annotated all spots by their highest-scoring spatial MP (STAR Methods). The overall composition of each GBM sample was highly variable (Figures 2E and S3C), with MES-Hyp and Neuron frequencies varying the most between samples and MES-like frequency varying the least. Even samples isolated from different regions of the same tumor were highly variable in their cell-type composition (Figures 2E and S1D), highlighting the degree of sampling bias when a single tissue section is considered representative of the composition of an entire tumor. Assignment of spots by MPs also correlated with their assignment by other signatures associated with invasiveness and connectivity[10,24–30](Figure S2I). Spots of the three mesenchymal MPs (MES-like, MES-Hyp, and MES-Ast) and Reactive-Ast were all associated with high connectivity, while invasiveness was significantly enriched with neurodevelopmental MPs (OPC-like, NPC-like, and AC-like) and with cell-cycle signal (Figure S3D), consistent with recent work.[26,27,30] Spots classified as ''Connected'' or ''Invasive'' were spatially distinct in most samples (Figure S3E).

Spatial proteomics as a complementary approach to map gliomas

While Visium provides comprehensive data by covering most genes, it suffers from low spatial resolution, which is not well suited for capturing rare or spatially scattered cell types and cell states. Therefore, we used CODEX as a complementary approach to validate our findings on a true single-cell protein level. Our 40-marker antibody panel was designed based on single-cell GBM MPs along with canonical markers to cover almost all relevant cell types and cell states (Figures 3A and 3B; Table S3). After quality control, we retained 428,395 cells from 12 samples (STAR Methods). Protein expression profiles of single cells were clustered using PhenoGraph,[31] and cluster assignment was performed based on the criteria described in Table S4. All major differentiated non-malignant cell types were identified, including astrocytes, oligodendrocytes, neurons, vascular cells (endothelial cells and pericytes), T cells, B cells, and macrophages/microglia (Figures 3C, 3D, S4A, and S4B). We also identified the major malignant cell states including MES-like, MES-Hyp, Chromatin-Reg, OPC-like, NPC-like, and AC-like (Figure 3E).

To enable direct comparison of CODEX and Visium data, we spatially aligned both datasets using STalign,[32] then transformed the CODEX data from single cells to pseudospots comparable to Visium, and assigned these to MPs. In 80% of the samples, MP assignments were significantly correlated between CODEX and Visium data from near-adjacent sections ($p < 0.05$, t test), serving as a cross-validation of our MP assignments (Figures 3F and S4C, STAR Methods).

In Visium data, several cell types and states were not detected by unsupervised analysis since they tend to represent a minor component of the spots in which they reside. For example, we did not identify a cell-cycle MP, raising the possibility that, in most cases, a cycling cell is surrounded by many non-cycling cells that dilute the cell-cycle signal (Figure S4D). Similarly, T cells and B cells rarely dominate a spot, and their low mRNA content further limits their signal such that we could not identify T cell or B cell MPs by unsupervised analysis. Accordingly, CODEX was better suited for identifying these cell types at the single-cell level (Figure 3D). T cells were detected in all samples but were lowly abundant in most tissues (0.4% median per sample), whereas B cells were absent in all but three samples (Figure S4B). In line with our previous findings,[4] OPC-like was the most cycling malignant state (median 10.8% Ki67+), whereas Vasc was the most cycling non-malignant state (median 4.6% Ki67+), highlighting abundant microvascular proliferation (MVP), a hallmark of GBM (Figure S4E). Supervised cell-cycle analysis of the Visium data further supported these results and additionally showed that among the malignant states, Prolif-Metab contained the largest percentage of cycling spots (median 10.6%) followed by OPC-like (median 9.4%) (Figure S4F, STAR Methods).

Using single-cell CODEX data as a ground truth allowed us to examine the full composition of cells within Visium-sized pseudospots. We found that in 64% of pseudospots, one MP is

Figure 3. Spatial profiling of gliomas by CODEX

(A) Schematic workflow of CODEX experiment, image processing, and computational analysis. Created with BioRender.com.

(B) Protein markers profiled by CODEX.

(C) Relative protein expression per cell type (or state) by *Z* score. Columns represent proteins or morphological features. Each row represents the indicated cell type/state in one sample. Number of rows differ between cell types/states because some samples lack specific cell types/states.

(D) CODEX staining of representative examples: top row shows images with the indicated markers. Bottom row shows the corresponding nuclear segmentation masks with cytoplasmic expansion by 3 μm. Masks are colored by cell state. Scale bar, 20 μm.

(E) Voronoi diagram showing cell-type/state abundances across all samples. Dashed lines separate malignant from non-malignant and immune cells.

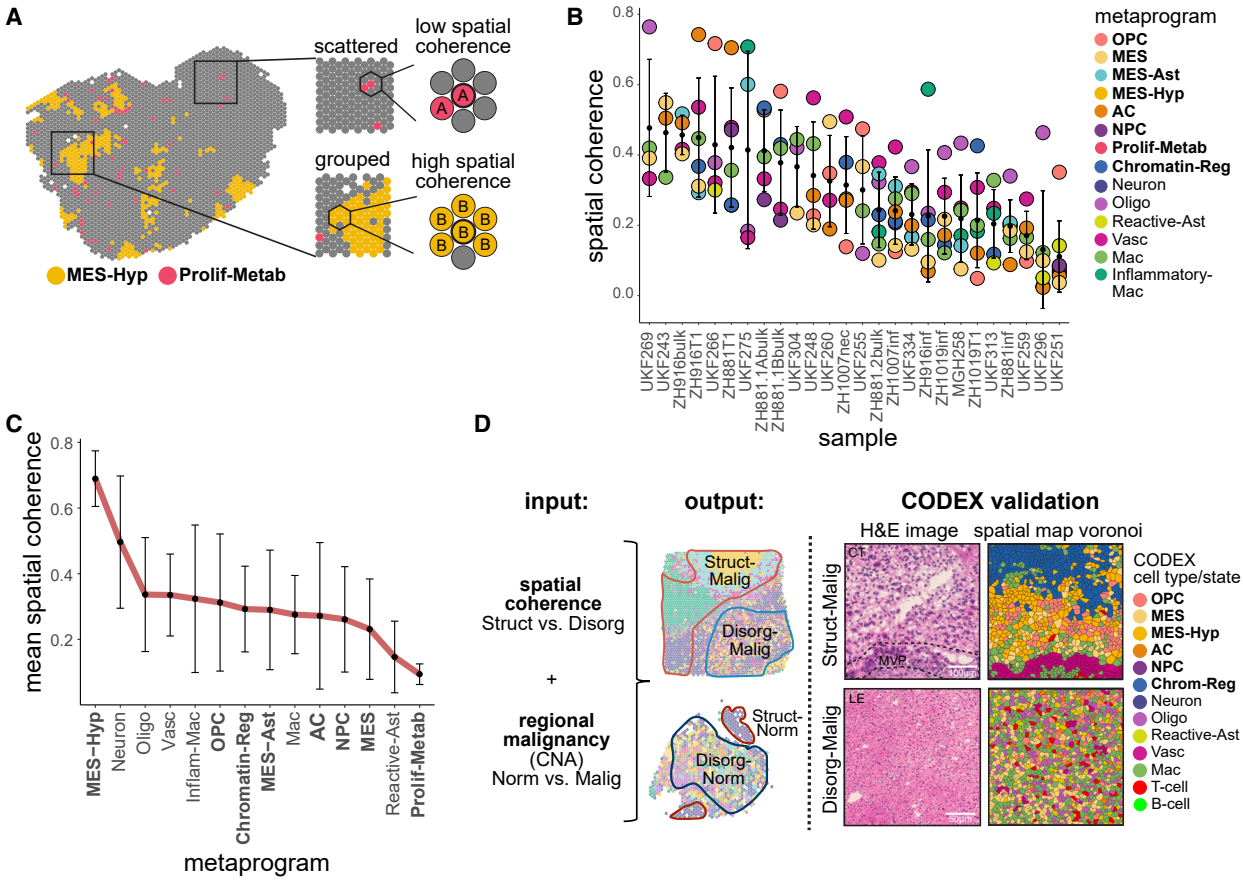
(F) Spatial maps of cell types/states annotations for (i) CODEX single-cell data, (ii) CODEX pseudospots, and (iii) near-adjacent section of ZH1019 infiltrating following alignment between Visium and CODEX with STalign (STAR Methods), median Pearson correlation between CODEX and Visium = 0.664, p value = 0.0105 by t test). Visium spots annotated by their highest scoring MP and CODEX pseudospots are annotated by the most abundant state within each spot.

(G) Cumulative distribution plot showing the frequency of pseudospots in which the highest cell type/state, the second highest, or both cover a higher proportion of the cells than the value specified on the x axis.

(H) Exemplary CODEX image overlaid with pseudospot grid (white) and nuclear segmentation masks colored by cell type/state.

See also Figure S4.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029



(legend on next page)

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

sufficient to account for most cells in that pseudospot, and in 99% of pseudospots, one or two MPs are sufficient to account for most cells (Figures 3G, 3H, and S4G). Thus, most spots are dominated by just one state, supporting our approach of annotating Visium spots by their dominant MP (Figure 2E). Moreover, considering the full composition of spots, as opposed to focusing on the dominant state, did not influence the downstream analysis and conclusions (Figure S4H, STAR Methods).

CODEX data also enabled us to assess the cell density of each state. Non-malignant brain cell types had lower density (median 5 cells/spot) than malignant cells, immune cells, and endothelial cells (median of 9 cells/spot). Of the malignant states, Chromatin-Reg and MES-like were the densest, while NPC-like, OPC-like, and AC-like were the least dense (Figure S4I). We also calculated average cell densities for classical GBM histology features. As expected, *infiltrating* regions had the lowest density (median 6 cells/spot), whereas *MVP* (16 cells/spot) and *pseudo-palisading* (median 18 cells/spot) had the highest cell density, though in the case of *pseudo-palisading*, this increased density was not associated with increased proliferation (Figure S4J).

GBM tumors contain co-occurring structured and disorganized regions

After annotating each spot by its highest-scoring spatial MP (Figures S3C and S4A), we wondered to what degree each state is enriched in neighboring vs. in scattered spots in each sample, which we term "spatial coherence" (Figure 4A, STAR Methods). Spatial coherence varies between samples, such that in some samples (termed "structured"), most states tend to have high coherence, while in other samples (termed "disorganized"), most states tend to have low coherence (Figure 4B). While most states varied together in spatial coherence based on sample and location, a few states were either consistently grouped (high spatial coherence) or scattered (low spatial coherence) (Figure 4C). These exceptions included MES-Hyp and Neuron, which had consistently high spatial coherence across samples, and Prolif-Metab, which had consistently low spatial coherence. Apart from these exceptions, spatial coherence varied more across samples than across states, indicating that it is more region specific than state specific (two-way ANOVA: sample effect $p = 0.000003$; state effect $p = 0.23$).

To further explore this distinction, we devised a computational approach to classify each spot as belonging to a *structured*, *disorganized*, or *intermediate* region based on the coherence of its local environment (Figure 4D, STAR Methods). Structured regions (48% of total spots) were found both at the core of tumors (enriched with MES-Hyp) and in infiltrated areas of the normal brain (enriched with Neuron) (Figures S5A and S5B). Therefore, we further subdivided structured and disorganized regions based on their malignancy level (CNA signal), defining four classes: Struct-Malig ($n = 20$ regions, 39% of spots), Disorg-Malig ($n = 10$ regions, 22% spots), Struct-Norm ($n = 6$ regions, 8% spots), and Disorg-Norm ($n = 4$ regions, 6% spots) (Figures 4D, 4E, and S5C). The remaining spots were annotated as intermediate, corresponding to the boundaries between organizational zones. All GBM tumors with multiple sections were composed of both structured and disorganized compartments, suggesting that the co-occurrence of these patterns is a recurring spatial feature of GBM (Figures 4E and S5D). Struct-Norm regions reflected the organization of the non-malignant brain parenchyma and had the highest frequency of Neuron, while Disorg-Norm was composed of normal brain regions with a high degree of cancer infiltration and had the highest frequency of Oligo (Figure S5B).

Spatial organization of cancer cell states is associated with MES-Hyp abundance

As noted above, MES-Hyp was enriched in Struct-Malig regions. Samples enriched with MES-Hyp had high mean spatial coherence even after removing the MES-Hyp spots from the calculation of spatial coherence ($p = 0.0013$ by t test, Figure S5E). Therefore, the increased organization associated with hypoxia extends beyond the hypoxic cancer cell state itself, and all malignant cell states in proximity to MES-Hyp spots were more organized than those same states in regions lacking MES-Hyp spots (Figure 4F, STAR Methods). Accordingly, structured regions around hypoxia were significantly more abundant than the two histological annotations associated with hypoxia-related structure—*pseudopalisading and necrosis* (PAN) and MVP

Figure 4. Spatial distribution of cell states and organization zones

(A) Scheme of ZH881_T1 illustrating spatial coherence score calculation depicting a high spatial coherence score for a grouped cell state and a low spatial coherence score for a scattered cell state.

(B) MP spatial coherence by sample. Standard deviation is shown in error bars.

(C) MP mean spatial coherence across all samples; standard deviation is shown in error bars. Two-way ANOVA to compare variance across samples ($p = 0.000003$) vs. across states ($p = 0.23$).

(D)Left: scheme describing the assignment to organization zones. Spatial coherence and spot malignancy level (inferred from CNA) were used to assign spots to structured-malignant (Struct-Malig), structured-normal (Struct-Norm), disorganized malignant (Disorg-Malig), and disorganized-normal (Disorg-Norm) in Visium data. Right: structured and disorganized regions are also found in CODEX data. H&E of the Struct-Malig zone shows microvascular proliferation (MVP) but no pseudopalisades or necrosis in a region that is structured at the CODEX cell state level. Upper scale bar, 100 μm; lower scale bar 50μm.

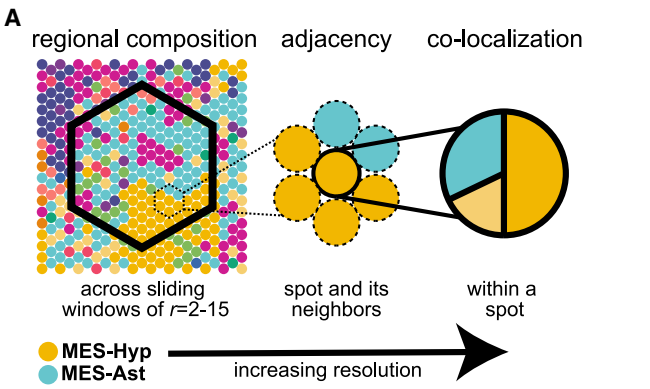
(E) Left: heatmap of the relative abundance of the different organizational patterns (rows) in each sample (columns). Annotation bars show per-sample MES-Hyp abundance (yellow) and assignment to glioma type (green). Right: spatial maps of two samples from the same tumor profiled by Visium showing multiple organization zones.

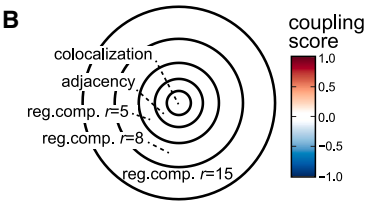
(F) Boxplot of spatial coherence score of malignant MPs across regions (windows) with high abundance (>10%) of MES-Hyp vs. low abundance of MES-Hyp. p values are shown on the plot (by Bonferroni-adjusted t test). Boxes indicate the median and the 1st and 3rd quartiles. The upper and lower whiskers extend to the maximal and minimal values.

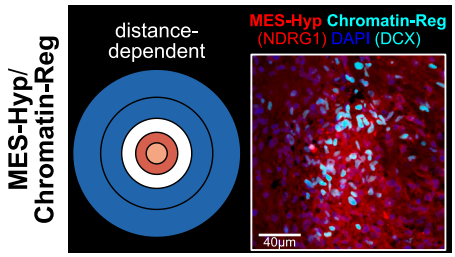
(G) Fraction of structured regions based on H&E staining (MVP and PAN) vs. fraction of structured regions based on CODEX spatial coherence ($p = 0.0019$, by Wilcoxon rank-sum test), boxes as defined in (F).

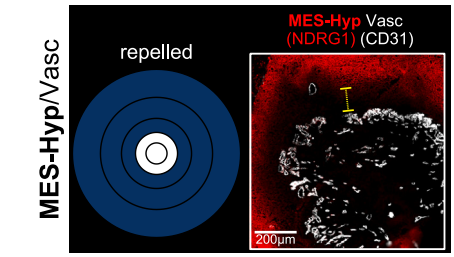
See also Figures S5 and S6.

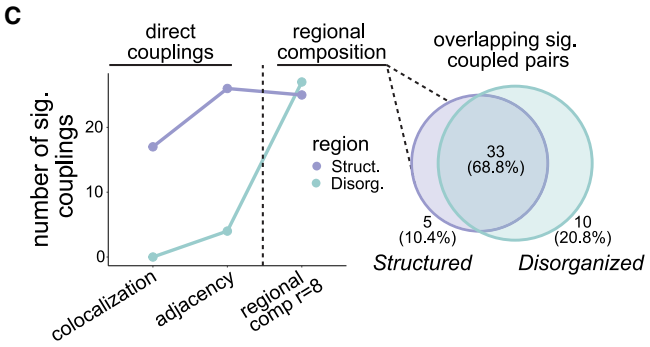
Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

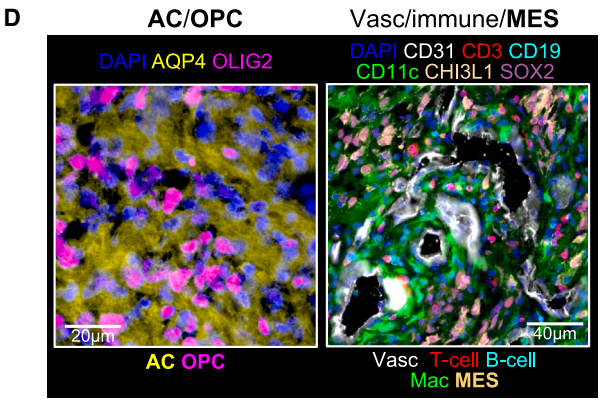


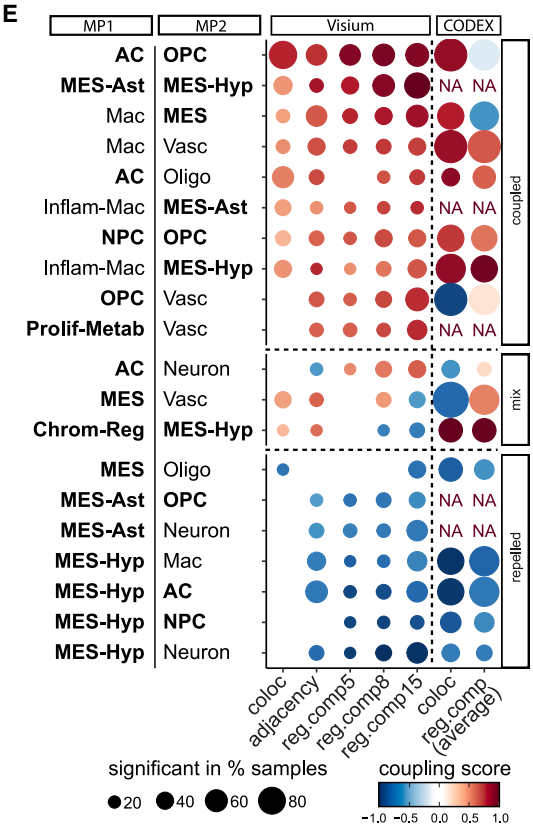


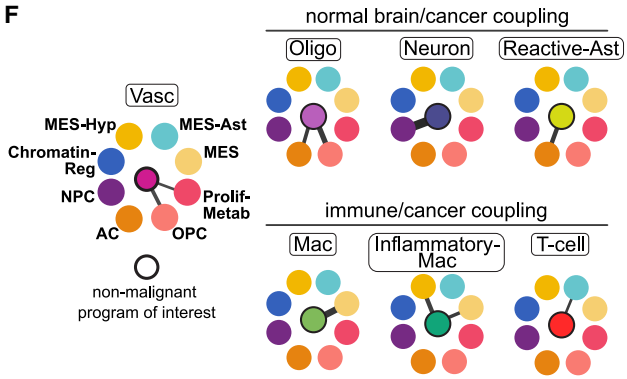












(legend on next page)

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

(Figure 4G, $p = 0.0019$ by Wilcoxon rank-sum test). While Neuron spots were also highly spatially coherent, this effect of extended structure on adjacent cell states was not observed for Neuron-containing regions ($p = 0.35$ by t test, Figure S5E). Thus, hypoxia may provide an extrinsic force that confers a structured continuum of expression states over large tissue regions[33] resulting in a known local structure that is visible by standard histology and a previously unappreciated long-range structure that is detected by molecular profiling of cellular states. Non-hypoxic regions in both GBM and IDH-mutant gliomas were largely disorganized (Disorg-Malig) (Figures 4E and S5F), consistent with the overall disorganization of gliomas seen on histology.

We hypothesized that if hypoxia is the driver of spatial organization, then subjecting disorganized GBM organoids to hypoxia may induce the organization of cancer cell states. To this end, we subjected a patient-derived organoid model[34] to hypoxia (2% O_2) for 24 h ($n = 6$) or 48 h ($n = 4$) with respective controls from the same model at normoxia and profiled them by CODEX (Figure S5G). While exogenous hypoxia globally upregulated classical hypoxia-related markers (i.e., GLUT1, CA9, and NDRG1), it did not induce the MES-Hyp state, suggesting that a full state transition might require long-term exposure to hypoxia and/or necrosis, immune cells, and vasculature (Figures S5H–S5J). Additionally, while the organoid model contained the core malignant GBM cellular states, it did not faithfully recapitulate the spatial organization of human tumor samples, even in control organoids containing MES-Hyp (Figures S5K and S5L). Therefore, we lacked an appropriate model to address the role of hypoxia on the spatial organization of cancer cell states *ex vivo*.

While MES-Hyp and Neuron reflected the greatest difference between structured and disorganized regions, we observed additional differences in regional composition. Vasc spots were more abundant in Struct-Malig regions ($p = 0.03$), while Prolif-Metab was more abundant in disorganized regions ($p = 0.003$) (Figure S6A). Vasc could be further subclustered into two populations with differing spatial coherence—an angiogenic (Vasc-Ang) program with high coherence and enrichment in structured regions, and an immunomodulatory (Vasc-IMEC) program with lower coherence and enrichment in disorganized and infiltrative regions (Table S2; Figures S6B–S6D).

Notably, we also identified structured and disorganized regions in the CODEX data, with high consistency between near-adjacent Visium and CODEX samples ($R = 0.73$, $p = 0.011$; Figures 4D,S6E, S6F, and S4A-iv). Moreover, the CODEX data demonstrated that the difference in spatial coherence between structured and disorganized regions is not restricted to neighboring pseudospots but is also mirrored within pseudospots (i.e., single cells and their neighbors). We defined a measure of pseudospot homogeneity and found significantly higher homogeneity in structured regions than in disorganized regions for 13 of 14 MPs (Figure S6G). Importantly, the differences in spatial coherence between *structured* and *disorganized* regions remained when controlling for pseudospot homogeneity (Figure S6H).

Pairwise state-state spatial associations recur across samples and scales

To quantify spatial relationships between states, we devised three complementary measures highlighting state coupling at varying levels of resolution (STAR Methods). First, the *regional composition* of the two states, defined as the correlation between their abundance across hexagonal windows of a predefined radius (denoted by *r*) (Figure 5A). Second, *adjacency* between two states, defined as the enrichment of one state in the immediate neighboring spots of the other state. Third, the *colocalization* of two states within the same spot (or pseudospot in CODEX). We observed overall consistency of state relationships over increasingly sized areas, while also identifying specific scale-dependent shifts (*distance-dependent coupling*) when considering all measures together (67% consistency across measures in structured regions, 57% consistency across measures in disorganized regions) (Figure 5B; Data S1).

Structured and disorganized regions had a similar number of significant spatial associations by regional composition, and these associations were largely consistent, while the colocalization and adjacency measures mostly identified interactions that were specific to structured regions (Figure 5C). Therefore,

Figure 5. Spatial associations between states across scales

(A) Scheme depicting the three measures of spatial relationships between MPs across scales (from low to high resolution): (i) regional composition (across sliding windows of $r = 2$–15), (ii) adjacency, and (iii) colocalization (STAR Methods). The example shown highlights the coupling between MES-Ast and MES-Hyp across scales of resolution.

(B) Nested circle plots depicting spatial relationship strength across scales of resolution. Top: schematic of nested circle plot. Bottom: example of associations with nested circle plots (Visium) adjacent to CODEX images, scale bar indicated in the images. Yellow line shows the distance from hypoxia (*NDRG1*) to blood vessels (CD31) ~160 μm. Note: MES-Hyp/Vasc association is NA (white) at the colocalization and adjacency levels due to too low abundance of vascular spots within hypoxic regions to perform the calculation.

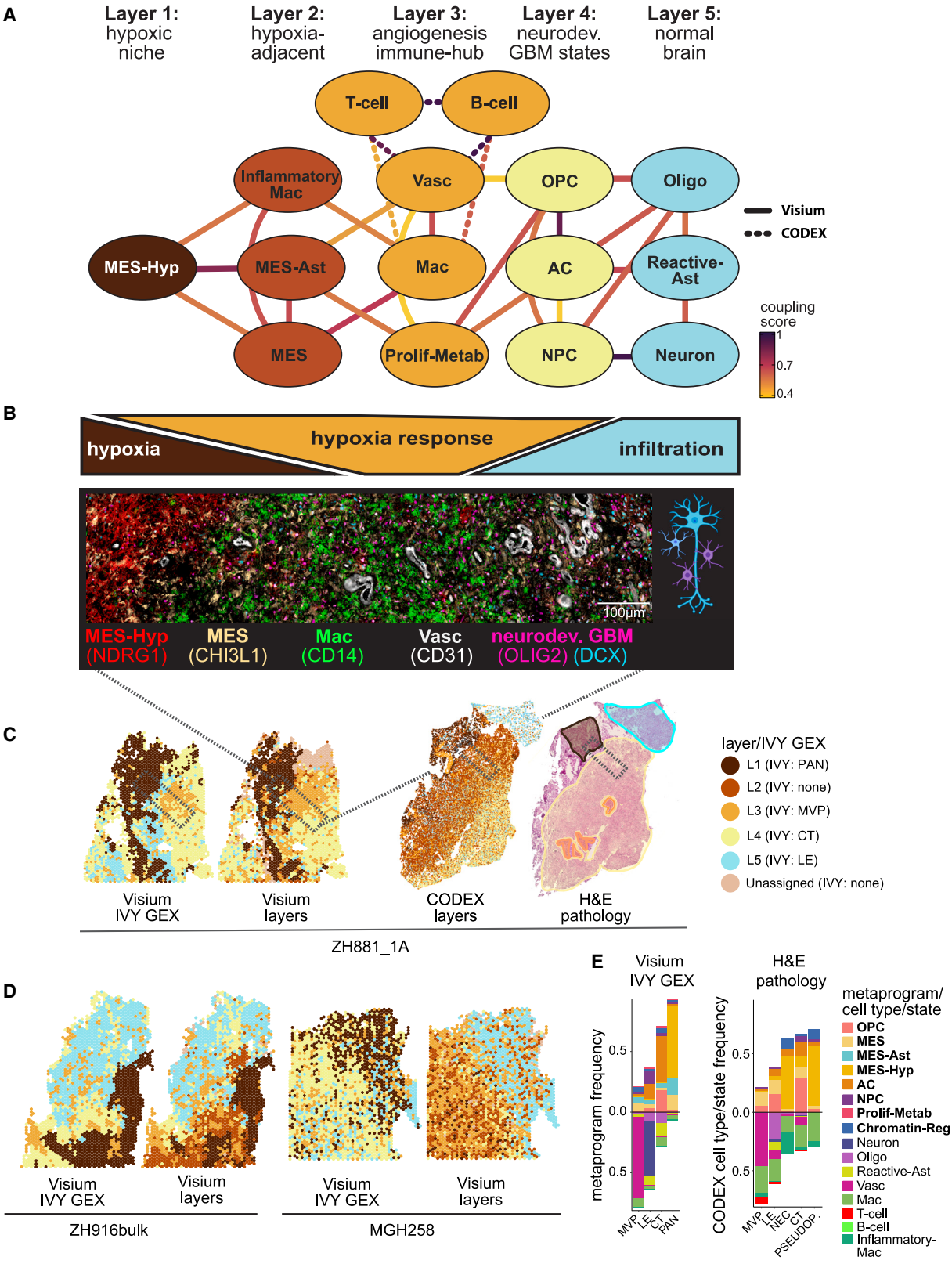
(C) Left: line plot showing the number of significant interactions across different analyses in structured vs. disorganized regions. Right: Venn diagram of the number of significant interactions across the regional composition analysis (see criteria for "consensus interactions" in STAR Methods).

(D) CODEX images highlighting selected pairs of coupled cell types from (C). Left: AC/OPC coupling (scale bar, 20 μm). Right: spatial relationships in the vascular niche (scale bar, 40 μm). While Vasc and immune cells are immediately adjacent, Vasc/MES coupling occurs over larger areas. (E) Summary heatmap of state-state associations across scales in structured samples covering most consensus interactions and other pairs of interest (see also Data S1). Each column represents the summary of a different spatial relationship measure (Visium data) across all GBM samples, with the final two columns corresponding to CODEX colocalization and regional composition, respectively. Dots are colored by mean scaled relationship strength (coupling score) (STAR Methods), and dot size corresponds to the fraction of samples in which the relationship is significant by Fisher's exact test.

(F) Top: network graphs representing couplings between normal brain cell types or Vasc with the malignant states; bottom: couplings of immune cells with malignant states. In each graph, the central node is a non-malignant cell type, and the surrounding nodes represent cancer states. Edge strength represents the mean scaled relationship strength across all measures (Visium). Only consensus interactions are shown.

See also Figure S7.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029



(legend on next page)

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

regional composition tends to be maintained even in the absence of direct spatial patterning defined by neighboring cells or spots.

To identify the most robust interactions that drive the organization of structured GBM regions, we defined a consensus set of state-state interactions that are supported by multiple measures and across multiple samples (STAR Methods). We found ~10-fold more consensus interactions in structured GBM regions than in disorganized GBM regions (21/91 vs. 2/91 state-pairs) (Figure 5C). The only consensus interactions in disorganized GBM regions were MES-like/Mac and NPC-like/Neuron. The consensus interactions were validated by CODEX (Figures 5D and S7A). Most consensus interactions are briefly summarized (Figure 5E) and described in more detail in Data S1. Malignant states were largely divided by the consensus interactions into two groups, each with many within-group interactions: (1) neurodevelopmental states (OPC-like, NPC-like, and AC-like) and (2) a set of mesenchymal and hypoxia-associated states (MES-Hyp, MES-Ast, MES-like, and Chromatin-Reg), consistent with other studies that describe a spatially segregated GBM microenvironment[9,35] (Figures 5E and S7A). The neurodevelopmental malignant states not only interacted among themselves but also with the non-malignant states corresponding to the same lineage—AC-like interacted with Reactive-Ast, NPC-like interacted with Neuron, and OPC-like interacted with Oligo (Figure 5F). OPC-like also interacted strongly with Vasc, reminiscent of OPC-endothelial migration[36] and cross-talk[37] in development. Likewise, Prolif-Metab was strongly associated with Vasc, in line with a previously described metabolic vessel-associated GBM state.[38]

All three mesenchymal states had strong interactions with immune cells,[8] but each interacted with different subsets of immune cells (Figure 5F). Of all the malignant states, MES-Ast spots were the most enriched with T cells (Figure S7B). The core MES state had an intermediate association with T cells, was spatially associated with both macrophage states and, interestingly, upregulated major histocompatibility complex (MHC)-II genes (Figures S7C–S7E). In contrast, MES-Hyp spots were most depleted of T cells and were strongly spatially associated with Inflammatory-Mac. These results suggest that MES-Ast is associated with immune activity, while MES-Hyp is associated with immunosuppression, consistent with earlier studies.[9,39–42]

Immune cells (Mac, T cells, and B cells) also strongly interacted with Vasc, highlighting their dependence on trafficking (especially for bone marrow-derived macrophages) (Figures 5E and S7A). While Vasc was coupled to immune cells, it was decoupled from MES-Hyp, as expected (Figure 5B). This decoupling was observed up to ~160 μm, suggesting that this is the effective length scale of vascular perfusion and that beyond this distance from blood vessels, there is a hypoxia response (Figure S7F).

A layered model of GBM spatial organization for structured regions

To examine whether the individual interactions described above may combine to form a higher-order organization, we generated a network graph in which nodes represent states and edges represent consensus interactions (Figures 6A and 6B). This graph reveals a five-layered organization in which edges exist only within the same layer or between adjacent layers. This organization appears to be dominated by a hypoxia gradient, with a first layer consisting of MES-Hyp (*L1: core hypoxia/necrosis*), followed by a second layer consisting of MES-Ast, MES-like, and Inflammatory-Mac (*L2: hypoxia-associated*). The third layer includes immune and angiogenesis-related cell types and states that may help to resolve the hypoxia, including Vasc, Mac, and Prolif-Metab (*L3: angiogenic response/immune hub*). Following *L3*, hypoxia is presumably resolved, enabling the presence of neurodevelopmental malignant states that may be more oxygen dependent—AC-like, OPC-like, and NPC-like (*L4: malignant neurodevelopmental states*). Finally, non-malignant brain cell types (Reactive-Ast, Oligo, and Neuron) reflect the transition to the infiltrated brain parenchyma (*L5: brain parenchyma*). Mapping of spots to the five layers (instead of mapping to MPs) preserves the distinction between structured and disorganized regions (Figures 6C, 6D, S8A, and S8B). CODEX data generated a similar graph with the addition of B cells and T cells to *L3* (*angiogenic response/immune hub*) (Figure S8C).

To better contextualize our layers model, we compared the assignment of our samples to the five layers with their histopathological annotations by a neuro-pathologist and by the IVY Gap transcriptional signatures derived from histopathological features (Figures 6E and S8D).[43] We found high consistency between our model and classical annotations, with *L1* resembling *PAN, L3* resembling *MVP*, *L4* resembling *cellular tumor*, and *L5* resembling *leading edge*. Notably, the hypoxia-adjacent layer (*L2*) and small areas of local hypoxia are not distinguished by classical histopathology or by the IVY Gap transcriptional

Figure 6. A layered model of GBM spatial organization

(A) Top: network graph with nodes representing cell types/states and edges representing recurrent interactions (mean scaled coupling score >0.35 across samples and significantly coupled in >20% of samples for at least two measures of spatial relationships). Edge color represents the mean coupling score across levels of resolution, for structured regions of Visium data. Edges with dashed lines represent connections with cell types coming from CODEX (T cells and B cells). Bottom: scheme showing gradients of hypoxia, hypoxia response, and infiltration in alignment with the layers.

(B) CODEX image showing the indicated cell types and markers, representing layers 1–4. Layer 5 was added as scheme depicting normal brain. Created with BioRender.com.

(C) Spatial maps for sample ZH881_1A: (i) Visium sample annotated by IVY Gap histological feature-associated transcriptional programs and state layers and (ii) CODEX sample annotated by state layers and neuropathologist annotations of H&E staining from the CODEX sample.

(D) Spatial maps comparison for Visium samples annotated by IVY Gap-histological-feature-associated transcriptional programs (left) vs. annotated by our layers (right).

(E) Stacked bar plot showing MP composition per IVY Gap histological feature transcriptional program annotation (left) and CODEX MP composition per histological annotations (right). Bars below 0 represent non-malignant MPs.

See also Figure S8.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

programs. Likewise, comparing transcriptional signatures derived from the five layers with transcriptional signatures corresponding to micro-dissected histopathological features showed overall agreement apart from *L2*, which did not map directly to a histopathological feature (Figure S8E; Table S2). Therefore, by unsupervised analysis of gene expression, we largely converged to the major histological features of glioma, yet with increased accuracy and with the addition of a new layer (*L2*) of hypoxia-adjacent states. Moreover, we can now redefine the classical histopathological features at a detailed resolution of cellular states (Figures 6E and S8E).

DISCUSSION

We combined Visium with CODEX and computational approaches to define the organization of gliomas. The limited resolution of Visium led previous studies to define spatial neighborhoods, often representing mixtures of cell states and cell types.[9] Resolving individual cell states and cell types from spatial transcriptomics remains challenging. Previous analysis relied on deconvolution of spot profiles with matched single-cell or single-nuclei RNA-seq.[20,23,44–47] Here, we present an alternative approach by deriving MPs that approximate single-cell profiles directly from spatial transcriptomics data. Moreover, we identified several MPs in the spatial transcriptomics data that were not previously defined from single-cell or single-nuclei data. These programs could be retrospectively identified by supervised analysis of single-cell data (Figures S2K and S2L), suggesting that spatial clustering of rare or subtle cellular states may have amplified their signals and facilitated their identification by Visium. These states may also have been depleted during tissue dissociation in scRNA-seq (e.g., tumor microtubes of MES-Ast or a bias against hypoxia/necrosis-associated states better captured *in situ*).

Conversely, signals for cellular states with limited spatial clustering are diluted by Visium and may not be detected by unsupervised analysis, as is the case for cell cycle, and for lowly abundant cell types such as T cells (see Figure S8F for a comparison of GBM cell states and cell types captured by different transcriptomics platforms). Our MP approach focuses on identifying the most robust and recurrent patterns of gene expression heterogeneity; therefore, by design, it will not capture rare cell states and cell types. To compensate for some of these limitations, we also performed CODEX on near-adjacent tissue sections, allowing us to detect lowly abundant cell types and uncover the composition of Visium spots at single-cell resolution.

Overall, we found three modes of spatial organization. First, cells tend to be surrounded by other cells in the same state (*state-specific clustering*), forming local environments that are highly enriched with an individual state. This mode is evident by the clustering of similar cells within each spot, as shown directly by CODEX, as well as by the clustering of spots assigned to the same MP. This clustering suggests that spatial location plays a central role in regulating cell state.

Second, many pairs of states are consistently associated across multiple scales in structured regions —within spots, in adjacent spots, and within small tissue regions (*state-state association*). Some of these interactions may mimic those of normal brain cell types in development. For example, malignant OPC-like is coupled to Vasc, consistent with migration of normal OPCs along the brain endothelium[36] and NPC-like is coupled with neurons in infiltrative areas, mimicking the migration of normal NPC toward developing neurons.[48,49] The exact mechanisms driving state-state associations require further studies and may reflect not only physical connections but also recruitment through secreted factors, dynamic transitions from one state to the other, synergistic growth or survival of the interacting cells, or a common dependence on microenvironmental components.

Third, the associations between states in structured regions aggregate to form a higher-order organization with five layers, such that states in each layer are associated only with the same layer or adjacent layers (*state layers*). The presence of higher-order organization is reminiscent of normal tissue structure and may facilitate large-scale coordination between cancer cell states that could present a therapeutic opportunity. Four layers are largely consistent with histological features, allowing us to redefine the classical GBM histology by specific cell states and to add the layer of hypoxia-associated states (*L2*).

Hypoxia appears to be a central driver of this organization, and each layer can be interpreted by its relation to hypoxia. Hypoxia is intimately linked to both necrosis and aberrant vasculature, both of which could also contribute to the emergence of organization, consistent with the previously observed association between blood vessel proximity and GBM cell state through metabolic zonation.[38] Importantly, the effect of hypoxia on spatial organization extends beyond the features visible by histopathology. Therefore, while the known structure of gliomas is typically restricted to small areas of hypoxia/necrosis or brain parenchyma, the structure at the level of cell states extends further. Additionally, some of the hypoxic regions defined by spatial profiling were not visible by histopathology yet still associated with the structured organization of cell states. This hypoxia-centric model is consistent with recent modeling studies, which suggested that an external force such as hypoxia acting over a large area can result in a spatial gradient that generates a continuum of expression states.[33,50] Likewise, hypoxia gradients were shown to generate a continuum of macrophage states generating predictable spatial patterns in mouse breast cancer models.[51]

Although tissue organization may be intuitively linked to normal physiology while chaos is reminiscent of aggressive tumor phenotypes, we note that hypoxia and necrosis, and hence cancer cell state organization, are hallmarks of high-grade glioma. Such organization may imply that certain layers are less accessible to drugs or to immune cells and thereby more resistant to particular therapies. In contrast, IDH-mutant gliomas tend to lack hypoxia and hence are typically disorganized. In the case of an IDH-mutant glioma that does have a high degree of hypoxia, we also observe a high degree of organization, consistent with the model of hypoxia as driving spatial organization (Figures S3C and S5F).

In addition to its association with global organization, hypoxia is linked to specific states in its vicinity, including the core hypoxic state of MES-Hyp, but also Chromatin-Reg and three additional states that constitute the hypoxia-associated layer. We speculate that most (or all) of these MPs represent

hypoxia-associated versions of other MPs observed further away from hypoxia. Pairs of related malignant states that may differ due to hypoxia proximity include MES-Hyp vs. MES-like, MES-Ast vs. AC-like, and Chromatin-Reg vs. NPC-like. The pairing of similar states also extends to myeloid cells, with Inflammatory-Mac reflecting a hypoxia-associated version of the macrophage program, characterized by upregulation of inflammatory cytokines, downregulation of MHC class II, and an association with CD8+ T cells.

In summary, we provide an extensive spatial description of glioma that demonstrates stereotypical spatial organization at multiple scales, with a prominent role of hypoxia as an organizer and with relative disorganization of regions that lack hypoxia. This adds a spatial dimension to our growing understanding of the glioma ecosystem and may aid in the development of future treatments.

Limitations of the study

Spot-based spatial transcriptomics suffers from low resolution, making it challenging to resolve individual cell states and cell types. Our computational approach partially resolves this, but very similar cell states and cell types (i.e., malignant NPC-like and non-malignant neurons) cannot be fully distinguished when they are colocalized within spots, and other states that are spatially scattered or lowly abundant are not efficiently detected by unsupervised analysis. Due to these limitations, we also profiled samples by CODEX and integrated both data types in much of our analysis. We believe that the integration of lower resolution, high complexity data (i.e., many genes) with higher resolution, low complexity data (i.e., a smaller number of preselected proteins) maximizes the potential of both approaches to reveal biological insights.

Our results suggest a major role of hypoxia in driving overall tissue structure and a specific 5-layer organization. Yet, we were unable to directly prove causality, due to the use of MES-Hyp as a proxy for hypoxia and the lack of an appropriate experimental model. Initial experiments with GBM organoids suggested that these models may not faithfully recapitulate human tumor organization and that subjecting them to hypoxia did not recapitulate the induction of MES-Hyp seen in human samples, potentially due to the absence of necrosis, immune cells, vasculature, and a true hypoxia gradient (Figures S5H–S5J).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Sample preparation
  - H&E staining and imaging
  - 10X Visium cDNA synthesis and library generation
  - Sequencing
  - CODEX antibodies and imaging
  - GBM hypoxia organoid experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Spatial transcriptomics (Visium) analysis
  - Gene-set enrichment analysis
  - CODEX analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.03.029.

ACKNOWLEDGMENTS

This work was supported by the European Research Council Consolidator Grant 101044318 (to I.T.) and a Broad Institute-Israel Science Foundation Collaborative Project Award (to I.T. and M.L.S.). I.T. is the incumbent of the Dr. Celia Zwillenberg-Fridman and Dr. Lutz Zwillenberg Career Development Chair and is supported by the Zuckerman STEM Leadership Program, the Mexican Friends New Generation, the Benoziyo Endowment Fund, and the Israel Cancer Research Fund. The study was also supported by N.I.H. R37CA245523 and R01CA258763 (both to M.L.S.). R.H. is funded by the Walter Benjamin Programme from the German Research Foundation. N.G.D. is supported by the Israeli Council for Higher Education (CHE) via the Weizmann Data Science Research Center and by a research grant from the Estate of Tully and Michele Plesser. V.M. is supported by a NIH T32 Cancer Neuroscience Training Grant (T32CA272386). The images in this paper were acquired at the Advanced Optical Imaging Unit, de Picciotto-Lesser Cell Observatory Unit at the Moross Integrated Cancer Center Life Science Core Facilities, Weizmann Institute of Science. The authors would like to thank Dr. Ofra Golani for helpful discussions regarding image analysis and cell segmentation and Rony Chanoch-Myers and Rotem Tal for helpful feedback on the manuscript. The authors would like to thank Abcam for the generous donation of carrier-free antibodies for CODEX experiments.

AUTHOR CONTRIBUTIONS

A.C.G., N.G.D., R.H., and I.T. conceived the project, designed the study, interpreted results, and wrote the manuscript. C.W.M., M.C.N., and M.L.S. provided guidance and feedback and reviewed the manuscript. A.C.G., N.G.D., R.H., D.S., T.T., and J.L. performed computational analyses. N.G.D. developed computational methods for spatial analysis. L.N.G.C., G.M., M.C.N., M.W., and M.L.S. provided glioma samples for spatial transcriptomics and spatial proteomics. A.C.G. performed spatial transcriptomics. R.H. performed spatial proteomics. R.H., C.W.M., Y.H., S.D., and V.M. performed GBM organoid experiments. C.W.M. annotated histology sections. M.N. performed snRNA-seq experiments. D.H. and B.L. sectioned tissues and provided histology expertise. M.K. and H.K.S. provided genomics expertise and Visium support. I.G. and Y.A. provided imaging and microscopy expertise and CODEX support. I.T. and M.L.S. supervised the study.

DECLARATION OF INTERESTS

I.T. is an advisory board member of Immunitas Therapeutics. M.L.S. is an equity holder, scientific co-founder, and advisory board member of Immunitas Therapeutics. Abcam provided carrier-free antibodies for CODEX experiments (to R.H.).

Received: June 16, 2023
Revised: January 11, 2024
Accepted: March 21, 2024
Published: April 22, 2024

REFERENCES

1. Bailey, P. (1932). HISTOLOGIC DIAGNOSIS OF TUMORS OF THE BRAIN. Arch. Neur.Psych. *27*, 1290. https://doi.org/10.1001/archneurpsyc.1932.02230180019003.

2. Scherer, H.J. (1938). Structural Development in Gliomas. Am. J. Cancer *34*, 333–351. https://doi.org/10.1158/ajc.1938.333.

3. Kleihues, P., and Ohgaki, H. (1999). Primary and secondary glioblastomas: from concept to clinical diagnosis. Neuro. Oncol *1*, 44–51. https://doi.org/10.1093/neuonc/1.1.44.

4. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835–849.e21. https://doi.org/10.1016/j.cell.2019.06.024.

5. Tirosh, I., Venteicher, A.S., Hebert, C., Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C., Mount, C., Filbin, M.G., et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature *539*, 309–313. https://doi.org/10.1038/nature20123.

6. Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., Hovestadt, V., Escalante, L.E., Shaw, M.L., Rodman, C., et al. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. Science *355*, eaai8478. https://doi.org/10.1126/science.aai8478.

7. Mathewson, N.D., Ashenberg, O., Tirosh, I., Gritsch, S., Perez, E.M., Marx, S., Jerby-Arnon, L., Chanoch-Myers, R., Hara, T., Richman, A.R., et al. (2021). Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis. Cell *184*, 1281–1298.e26. https://doi.org/10.1016/j.cell.2021.01.022.

8. Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell *39*, 779–792.e11. https://doi.org/10.1016/j.ccell.2021.05.002.

9. Ravi, V.M., Will, P., Kueckelhaus, J., Sun, N., Joseph, K., Salié, H., Vollmer, L., Kuliesiute, U., Von Ehr, J., Benotmane, J.K., et al. (2022). Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. Cancer Cell *40*, 639–655.e13. https://doi.org/10.1016/j.ccell.2022.05.009.

10. Venkatesh, H.S., Morishita, W., Geraghty, A.C., Silverbush, D., Gillespie, S.M., Arzt, M., Tam, L.T., Espenel, C., Ponnuswami, A., Ni, L., et al. (2019). Electrical and synaptic integration of glioma into neural circuits. Nature *573*, 539–545. https://doi.org/10.1038/s41586-019-1563-y.

11. Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82. https://doi.org/10.1126/science.aaf2403.

12. Maniatis, S., Äijö, T., Vickovic, S., Braine, C., Kang, K., Mollbrink, A., Fagegaltier, D., Andrusivová, Ž., Saarenpää, S., Saiz-Castro, G., et al. (2019). Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. Science *364*, 89–93. https://doi.org/10.1126/science.aav9776.

13. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044.

14. Kinker, G.S., Greenwald, A.C., Tal, R., Orlova, Z., Cuoco, M.S., McFarland, J.M., Warren, A., Rodman, C., Roth, J.A., Bender, S.A., et al. (2020). Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. Nat. Genet. *52*, 1208–1218. https://doi.org/10.1038/s41588-020-00726-6.

15. Gavish, A., Tyler, M., Greenwald, A.C., Hoefflin, R., Simkin, D., Tschernichovsky, R., Galili Darnell, N., Somech, E., Barbolin, C., Antman, T., et al. (2023). Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. Nature *618*, 598–606. https://doi.org/10.1038/s41586-023-06130-4.

16. Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142. https://doi.org/10.1126/science.aaa1934.

17. Ruiz-Moreno, C., Salas, S.M., Samuelsson, E., Brandner, S., Kranendonk, M.E.G., Nilsson, M., and Stunnenberg, H.G. (2022). Harmonized single-cell landscape, intercellular crosstalk and tumor architecture of glioblastoma. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.27.505439.

18. Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., Yong, J., Sun, L., Wang, X., Zhao, Y., et al. (2018). Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. Cell Res. *28*, 730–745. https://doi.org/10.1038/s41422-018-0053-3.

19. Li, H., Zhou, J., Li, Z., Chen, S., Liao, X., Zhang, B., Zhang, R., Wang, Y., Sun, S., and Gao, X. (2023). A comprehensive benchmarking with practical guidelines for cellular deconvolution of spatial transcriptomics. Nat. Commun. *14*, 1548. https://doi.org/10.1038/s41467-023-37168-7.

20. Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., et al. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. Nat. Biotechnol. *40*, 661–671. https://doi.org/10.1038/s41587-021-01139-4.

21. Kanemaru, K., Cranley, J., Muraro, D., Miranda, A.M.A., Ho, S.Y., Wilbrey-Clark, A., Patrick Pett, J., Polanski, K., Richardson, L., Litvinukova, M., et al. (2023). Spatially resolved multiomics of human cardiac niches. Nature *619*, 801–810. https://doi.org/10.1038/s41586-023-06311-1.

22. Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Irizarry, R.A. (2022). Robust decomposition of cell type mixtures in spatial transcriptomics. Nat. Biotechnol. *40*, 517–526. https://doi.org/10.1038/s41587-021-00830-w.

23. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I., and Heyn, H. (2021). SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. Nucleic Acids Res. *49*, e50. https://doi.org/10.1093/nar/gkab043.

24. Venkataramani, V., Tanev, D.I., Strahle, C., Studier-Fischer, A., Fankhauser, L., Kessler, T., Körber, C., Kardorff, M., Ratliff, M., Xie, R., et al. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. Nature *573*, 532–538. https://doi.org/10.1038/s41586-019-1564-x.

25. Osswald, M., Jung, E., Sahm, F., Solecki, G., Venkataramani, V., Blaes, J., Weil, S., Horstmann, H., Wiestler, B., Syed, M., et al. (2015). Brain tumour cells interconnect to a functional and resistant network. Nature *528*, 93–98. https://doi.org/10.1038/nature16071.

26. Hai, L., Hoffmann, D.C., Wagener, R.J., Azorin, D.D., Hausmann, D., Xie, R., Huppertz, M.C., Hiblot, J., Sievers, P., Heuer, S., et al. (2024). A clinically applicable connectivity signature for glioblastoma includes the tumor network driver CHI3L1. Nat. Commun. *15*, 968. https://doi.org/10.1038/s41467-024-45067-8.

27. Ratliff, M., Karimian-Jazi, K., Hoffmann, D.C., Rauschenbach, L., Simon, M., Hai, L., Mandelbaum, H., Schubert, M.C., Kessler, T., Uhlig, S., et al. (2023). Individual glioblastoma cells harbor both proliferative and invasive capabilities during tumor progression. Neuro-Oncology *25*, 2150–2162. https://doi.org/10.1093/neuonc/noad109.

28. Hausmann, D., Hoffmann, D.C., Venkataramani, V., Jung, E., Horschitz, S., Tetzlaff, S.K., Jabali, A., Hai, L., Kessler, T., Azoŕin, D.D., et al. (2023). Autonomous rhythmic activity in glioma networks drives brain tumour growth. Nature *613*, 179–186. https://doi.org/10.1038/s41586-022-05520-4.

29. Heuer, S., and Winkler, F. (2023). Glioblastoma revisited: from neuronal-like invasion to pacemaking. Trends Cancer *9*, 887–896. https://doi.org/10.1016/j.trecan.2023.07.009.

30. Venkataramani, V., Yang, Y., Schubert, M.C., Reyhan, E., Tetzlaff, S.K., Wißmann, N., Botz, M., Soyka, S.J., Beretta, C.A., Pramatarov, R.L., et al. (2022). Glioblastoma hijacks neuronal mechanisms for brain invasion. Cell *185*, 2899–2917.e31. https://doi.org/10.1016/j.cell.2022.06.054.

31. Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, E.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015).

Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184–197. https://doi.org/10.1016/j.cell.2015.05.047.

32. Clifton, K., Anant, M., Aihara, G., Atta, L., Aimiuwu, O.K., Kebschull, J.M., Miller, M.I., Tward, D., and Fan, J. (2023). STalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. Nat. Commun. *14*, 8123. https://doi.org/10.1038/s41467-023-43915-7.
33. Adler, M., Korem Kohanim, Y.K., Tendler, A., Mayo, A., and Alon, U. (2019). Continuum of Gene-Expression Profiles Provides Spatial Division of Labor within a Differentiated Cell Type. Cell Syst. *8*, 43–52.e5. https://doi.org/10.1016/j.cels.2018.12.008.
34. Jacob, F., Salinas, R.D., Zhang, D.Y., Nguyen, P.T.T., Schnoll, J.G., Wong, S.Z.H., Thokala, R., Sheikh, S., Saxena, D., Prokop, S., et al. (2020). A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. Cell *180*, 188–204.e22. https://doi.org/10.1016/j.cell.2019.11.036.
35. Mossi Albiach, A., Janusauskas, J., Kapustová, I., Kvedaraite, E., Codeluppi, S., Munting, J.B., Borm, L.E., Kjaer Jacobsen, J., Shamikh, A., Persson, O., and Linnarsson, S. (2023). Glioblastoma is spatially organized by neurodevelopmental programs and a glial-like wound healing response. Preprint at bioRxiv. https://doi.org/10.1101/2023.09.01.555882.
36. Tsai, H.-H., Niu, J., Munji, R., Davalos, D., Chang, J., Zhang, H., Tien, A.-C., Kuo, C.J., Chan, J.R., Daneman, R., and Fancy, S.P. (2016). Oligodendrocyte precursors migrate along vasculature in the developing nervous system. Science *351*, 379–384. https://doi.org/10.1126/science.aad3839.
37. Arai, K., and Lo, E.H. (2009). An oligovascular niche: cerebral endothelial cells promote the survival and proliferation of oligodendrocyte precursor cells. J. Neurosci. *29*, 4351–4355. https://doi.org/10.1523/JNEUROSCI.0035-09.2009.
38. Kumar, S., Sharife, H., Kreisel, T., Mogilevsky, M., Bar-Lev, L., Grunewald, M., Aizenshtein, E., Karni, R., Paldor, I., Shlomi, T., and Keshet, E. (2019). Intra-Tumoral Metabolic Zonation and Resultant Phenotypic Diversification Are Dictated by Blood Vessel Proximity. Cell Metab. *30*, 201–211.e6. https://doi.org/10.1016/j.cmet.2019.04.003.
39. Sattiraju, A., Kang, S., Giotti, B., Chen, Z., Marallano, V.J., Brusco, C., Ramakrishnan, A., Shen, L., Tsankov, A.M., Hambardzumyan, D., et al. (2023). Hypoxic niches attract and sequester tumor-associated macrophages and cytotoxic T cells and reprogram them for immunosuppression. Immunity *56*, 1825–1843.e6. https://doi.org/10.1016/j.immuni.2023.06.017.
40. Gabrusiewicz, K., Li, X., Wei, J., Hashimoto, Y., Marisetty, A.L., Ott, M., Wang, F., Hawke, D., Yu, J., Healy, L.M., et al. (2018). Glioblastoma stem cell-derived exosomes induce M2 macrophages and PD-L1 expression on human monocytes. OncoImmunology *7*, e1412909. https://doi.org/10.1080/2162402X.2017.1412909.
41. Miller, T.E., El Farran, C.A.E., Couturier, C.P., Chen, Z., D'Antonio, J.P., Verga, J., Villanueva, M.A., Castro, L.N.G., Tong, Y.E., Saadi, T.A., et al. (2023). Programs, Origins, and Niches of Immunomodulatory Myeloid Cells in Gliomas. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.24.563466.
42. Ravi, V.M., Neidert, N., Will, P., Joseph, K., Maier, J.P., Kückelhaus, J., Vollmer, L., Goeldner, J.M., Behringer, S.P., Scherer, F., et al. (2022). T-cell dysfunction in the glioblastoma microenvironment is mediated by myeloid cells releasing interleukin-10. Nat. Commun. *13*, 925. https://doi.org/10.1038/s41467-022-28523-1.
43. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.G., Smith, K.A., Lankerovich, M., Bertagnolli, D., Bickley, K., et al. (2018). An anatomic transcriptional atlas of human glioblastoma. Science *360*, 660–663. https://doi.org/10.1126/science.aaf2666.
44. Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. *24*, 425–436. https://doi.org/10.1038/s41593-020-00787-0.
45. Olaniru, O.E., Kadolsky, U., Kannambath, S., Vaikkinen, H., Fung, K., Dhami, P., and Persaud, S.J. (2023). Single-cell transcriptomic and spatial landscapes of the developing human pancreas. Cell Metab. *35*, 184–199.e5. https://doi.org/10.1016/j.cmet.2022.11.009.
46. Madissoon, E., Oliver, A.J., Kleshchevnikov, V., Wilbrey-Clark, A., Polanski, K., Richoz, N., Ribeiro Orsi, A., Mamanova, L., Bolt, L., Elmentaite, R., et al. (2023). A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. Nat. Genet. *55*, 66–77. https://doi.org/10.1038/s41588-022-01243-4.
47. Lopez, R., Li, B., Keren-Shaul, H., Boyeau, P., Kedmi, M., Pilzer, D., Jelinski, A., Yofe, I., David, E., Wagner, A., et al. (2022). DestVI identifies continuums of cell types in spatial transcriptomics data. Nat. Biotechnol. *40*, 1360–1369. https://doi.org/10.1038/s41587-022-01272-8.
48. Cuddapah, V.A., Robel, S., Watkins, S., and Sontheimer, H. (2014). A neurocentric perspective on glioma invasion. Nat. Rev. Neurosci. *15*, 455–465. https://doi.org/10.1038/nrn3765.
49. Nadarajah, B., and Parnavelas, J.G. (2002). Modes of neuronal migration in the developing cerebral cortex. Nat. Rev. Neurosci. *3*, 423–432. https://doi.org/10.1038/nrn845.
50. Adler, M., Moriel, N., Goeva, A., Avraham-Davidi, I., Mages, S., Adams, T.S., Kaminski, N., Macosko, E.Z., Regev, A., Medzhitov, R., and Nitzan, M. (2023). Emergence of division of labor in tissues through cell interactions and spatial cues. Cell Rep. *42*, 112412. https://doi.org/10.1016/j.celrep.2023.112412.
51. Carmona-Fontaine, C., Deforet, M., Akkari, L., Thompson, C.B., Joyce, J.A., and Xavier, J.B. (2017). Metabolic origins of spatial organization in the tumor microenvironment. Proc. Natl. Acad. Sci. USA *114*, 2934–2939. https://doi.org/10.1073/pnas.1700600114.
52. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
53. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.
54. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. *7*, 16878. https://doi.org/10.1038/s41598-017-17204-5.
55. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell Detection with Star-convex Polygons. In Proceedings of Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference (Springer), pp. 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.
56. Eling, N., Hoch, T., Zanotelli, V., Fischer, J., and Schulz, D. imcRtools. ([object Object]). https://doi.org/10.18129/B9.BIOC.IMCRTOOLS.
57. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367. https://doi.org/10.1186/1471-2105-11-367.
58. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.
59. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
60. Righelli, D., Weber, L.M., Crowell, H.L., Pardo, B., Collado-Torres, L., Ghazanfar, S., Lun, A.T.L., Hicks, S.C., and Risso, D. (2022).

SpatialExperiment: infrastructure for spatially-resolved transcriptomics data in R using Bioconductor. Bioinformatics *38*, 3128–3131. https://doi.org/10.1093/bioinformatics/btac299.

61. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.
62. Black, S., Phillips, D., Hickey, J.W., Kennedy-Darling, J., Venkataraaman, V.G., Samusik, N., Goltsev, Y., Schürch, C.M., and Nolan, G.P. (2021). CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat. Protoc. *16*, 3802–3835. https://doi.org/10.1038/s41596-021-00556-8.
63. Jacob, F., Ming, G.-L., and Song, H. (2020). Generation and biobanking of patient-derived glioblastoma organoids and their application in CAR T cell testing. Nat. Protoc. *15*, 4000–4033. https://doi.org/10.1038/s41596-020-0402-9.
64. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.
65. Venables, W.N., Ripley, B.D., and Venables, W.N. (2002). Modern applied statistics with S, Fourth Edition (Springer).
66. Chanoch-Myers, R., Wider, A., Suva, M.L., and Tirosh, I. (2022). Elucidating the diversity of malignant mesenchymal states in glioblastoma by integrative analysis. Genome Med. *14*, 106. https://doi.org/10.1186/s13073-022-01109-8.
67. Tabula; Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367–372. https://doi.org/10.1038/s41586-018-0590-4.
68. Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D.H., and Kriegstein, A.R. (2019). Single-cell genomics identifies cell type-specific molecular changes in autism. Science *364*, 685–689. https://doi.org/10.1126/science.aav8130.
69. Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176. https://doi.org/10.1038/nature05453.
70. Camp, J.G., Sekine, K., Gerber, T., Loeffler-Wirth, H., Binder, H., Gac, M., Kanton, S., Kageyama, J., Damm, G., Seehofer, D., et al. (2017). Multilineage communication regulates human liver bud development from pluripotency. Nature *546*, 533–538. https://doi.org/10.1038/nature22796.
71. Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. (2020). A human cell atlas of fetal gene expression. Science *370*, eaba7721. https://doi.org/10.1126/science.aba7721.
72. Durante, M.A., Kurtenbach, S., Sargi, Z.B., Harbour, J.W., Choi, R., Kurtenbach, S., Goss, G.M., Matsunami, H., and Goldstein, B.J. (2020). Single-cell analysis of olfactory neurogenesis and differentiation in adult humans. Nat. Neurosci. *23*, 323–326. https://doi.org/10.1038/s41593-020-0587-9.
73. Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318–1323. https://doi.org/10.1126/science.aap8809.
74. Taube, J.M., Akturk, G., Angelo, M., Engle, E.L., Gnjatic, S., Greenbaum, S., Greenwald, N.F., Hedvat, C.V., Hollmann, T.J., Juco, J., et al. (2020). The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. J. Immunother. Cancer *8*, e000155. https://doi.org/10.1136/jitc-2019-000155.
75. Hickey, J.W., Tan, Y., Nolan, G.P., and Goltsev, Y. (2021). Strategies for Accurate Cell Type Identification in CODEX Multiplexed Imaging Data. Front. Immunol. *12*, 727626. https://doi.org/10.3389/fimmu.2021.727626.
76. Hickey, J.W., Becker, W.R., Nevins, S.A., Horning, A., Perez, A.E., Zhu, C., Zhu, B., Wei, B., Chiu, R., Chen, D.C., et al. (2023). Organization of the human intestine at single-cell resolution. Nature *619*, 572–584. https://doi.org/10.1038/s41586-023-05915-x.
77. Glaunès, J., Qiu, A., Miller, M.I., and Younes, L. (2008). Large Deformation Diffeomorphic Metric Curve Mapping. Int. J. Comput. Vis. *80*, 317–336. https://doi.org/10.1007/s11263-008-0141-9.
78. Wang, L., Jung, J., Babikir, H., Shamardani, K., Jain, S., Feng, X., Gupta, N., Rosi, S., Chang, S., Raleigh, D., et al. (2022). A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nat. Cancer *3*, 1534–1552. https://doi.org/10.1038/s43018-022-00475-x.
79. Al-Dalahmah, O., Argenziano, M.G., Kannan, A., Mahajan, A., Furnari, J., Paryani, F., Boyett, D., Save, A., Humala, N., Khan, F., et al. (2023). Reconvolving the compositional landscape of primary and recurrent glioblastoma reveals prognostic and targetable tissue states. Nat. Commun. *14*, 2586. https://doi.org/10.1038/s41467-023-38186-1.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
See Table S3 for list of all primary antibodies used in CODEX panel	N/A	N/A
Cy™2 AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	Cat #711-225-152; RRID: AB_2340612
Cy™3 AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	Cat #711-165-152; RRID: AB_2307443
Cy™3 AffiniPure Donkey Anti-Rat IgG (H+L)	Jackson ImmunoResearch	Cat #712-165-153; RRID: AB_2340667
Cy™3 AffiniPure Donkey Anti-Mouse IgG (H+L)	Jackson ImmunoResearch	Cat #715-165-150: RRID: AB_2340813
Alexa Fluor® 680 AffiniPure Donkey Anti-Goat IgG (H+L)	Jackson ImmunoResearch	Cat #705-625-147; RRID: AB_2340440
Biological samples		
Human glioblastoma samples	University Hospital Zurich	N/A
Human glioblastoma and IDH mutant glioma samples	Massachusetts General Hospital/Brigham and Women's Hospital	N/A
Human glioblastoma organoid model BWH911	Dana Farber Cancer Institute	N/A
Chemicals, peptides, and recombinant proteins		
OCT Compound	Scigen	Cat #4586
Aqueous Eosin Y solution	Millipore Sigma	Cat #HT110216
Mayer's Hematoxylin	Millipore Sigma	Cat #MHS16
Bluing Buffer	Dako	Cat #CS70230-2
SPRI Select Reagent	Beckman Coulter	Cat #B23318
DMEM/F-12	ThermoFisher Scientific	Cat #11320033
MEM non-essential amino acids	ThermoFisher Scientific	Cat #11140050
Neurobasal medium	ThermoFisher Scientific	Cat # 21103049
Critical commercial assays		
Quick RNA Microprep Kit	Zymo	Cat #ZR-R1051
Tapestation RNA Screen Tape	Agilent	Cat #5067-5576
Tapestation High Sensitivity D5000 Screen Tape	Agilent	Cat #5067-5592
Tapestation High Sensitivity D1000 Screen Tape	Agilent	Cat #5067-5584
Visium Spatial Tissue Optimization Kit	10X Genomics	Cat # PN-1000193
Visium Accessory Kit	10X Genomics	Cat #PN-1000194
Dual Index Kit TT Set A	10X Genomics	Cat #PN-1000215
Novaseq SP 100 Cycles Sequencing Kit	Illumina	Cat #20028401
Visium Spatial Gene Expression Slide and Reagent Kit	10X Genomics	Cat # PN-1000184
FAST SYBR qPCR Master Mix	KAPA Biosystems	Cat # KK4600
NEBNext Library Quanit Kit for Illumina	New England Biolabs	Cat # E7630L
Conjugation Kit	Akoya Biosciences	Cat # 7000009
Staining Kit	Akoya Biosciences	Cat # 7000008
CODEX buffer	Akoya Biosciences	Cat # 7000001
Nuclear stain	Akoya Biosciences	Cat # 7000003
Assay reagents	Akoya Biosciences	Cat # 7000002
Software and algorithms		
10x Genomics SpaceRanger 1.0 and 1.1	10X Genomics	N/A

(*Continued on next page*)

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
10x Genomics Loupe Browser 5.0.1	10X Genomics	N/A
Seurat 4.3.0	Hao et al.[52]	https://github.com/cran/Seurat
Scalop	J. Laffy, Tirosh Lab	https://github.com/jlaffy/scalop
Cytoscape 3.9.1	Shannon et al.[53]	https://cytoscape.org
QuPath	Bankhead et al.[54]	https://qupath.github.io/
StarDist	Schmidt et al.[55]	https://github.com/stardist/stardist
imcRtools	Eiling et al.[56]	https://github.com/BodenmillerGroup/imcRtools
R 4.1.1	N/A	https://cran.r-project.org/bin/windows/base/old/4.1.1/
NMF	Gaujoux, Seoighe et al.[57]	https://cran.r-project.org/ web/packages/NMF/index.html
MASS	Venables WN, Ripley BD. 2002	https://www.stats.ox.ac.uk/pub/MASS4/
inferCNA	Patel et al.[58]	https://github.com/jlaffy/infercna
MSigDB	Subramanian, Tamayo et al.[59]	https://www.gsea-msigdb.org/gsea/msigdb
SpatialExperiment	Righelli et al.[60]	https://github.com/drighelli/SpatialExperiment
Fusion 1.0.5	Akoya Biosciences	
PhenoGraph	Levine JH et al.[31]	https://dpeerlab.github.io/dpeerlab-website/phenograph.html
STalign	Clifton et al.[32]	https://jef.works/STalign/
WeightedTreemaps	Jahn et al.	https://github.com/m-jahn/WeightedTreemaps
Deposited data		
Glioblastoma and IDH-mutant glioma 10X Visium spatial transcriptomics	This study	[https://doi.org/10.5281/zenodo.8105466]; GEO: GSE237183
Glioblastoma CODEX imaging data	This study	[https://doi.org/10.5281/zenodo.8105466]
Glioblastoma 10X Visium spatial transcriptomics	Ravi et al.[9,42]	[https://doi.org/10.5061/dryad.h70rxwdmj]
Glioblastoma single-cell RNA sequencing (Smart Seq2)	Neftel et al.[4]	GEO: GSE131928
CODEX antibody validations	This study	[https://doi.org/10.5281/zenodo.8105466]
Ivy Glioblastoma Atlas Project Anatomic Structures RNA-Seq	Puchalski et al.[43]	https://glioblastoma.alleninstitute.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Itay Tirosh (itay.tirosh@weizmann.ac.il).

Materials availability

- This study did not generate new unique reagents.

Data and code availability

- Spatial transcriptomics data have been deposited at GEO (raw data) and Zenodo (processed data) and are publicly available as of the date of publication under the accession number GSE237183 (GEO) and [https://doi.org/10.5281/zenodo.8105466] (Zenodo). Imaging (CODEX) data and antibody validations have been deposited at Zenodo under the [https://doi.org/10.5281/zenodo.8105466]. This paper analyzes existing, publicly available data. The accession numbers for these datasets are listed in the key resources table.
- Original code as well as the input data used to generate the main analyses of the paper are publicly available at: http://github.com/tiroshlab/Spatial_Glioma.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Tumor samples used for Visium spatial transcriptomics and CODEX were obtained from patients undergoing tumor resection at University Hospital Zurich, Zurich, Switzerland (ZH samples), Massachusetts General Hospital, Boston, MA (MGH samples), and Brigham and Women's Hospital, Boston, MA (BWH samples) carried out in accordance with approved guidelines and with patient written consent under ethics approval KEK-ZH-Nr. 2015-0163, University Hospital Zurich, IRB #10-417, Dana Farber Cancer Institute, and IRB #1360-1, Weizmann Institute of Science. The clinical characteristics of the patient cohort are detailed in Table S1. Tumors ZH1007, ZH1019, ZH881, ZH916, and ZH1041 were spatially annotated by the surgeon during navigated-guided surgery. In these cases, multiple samples were collected from different regions of the same tumor annotated as *necrotic*, *T1 contrast-enhancing*, *infiltrating*, or *bulk*. Equal numbers of samples from males and females were used in this study (n=17 of each).

METHOD DETAILS

Sample preparation

Samples were either flash frozen by liquid nitrogen and embedded in cold OCT on dry ice (Scigen OCT Compound, #4586) when already frozen or embedded in OCT at the time of freezing. The RNA quality of each sample was evaluated by Tapestation (Tapestation RNA Screen Tape, Agilent) after isolating RNA (Zymo Quick RNA MicroPrep Kit, #ZR-R1051) from multiple tissue sections. Samples with RIN values >7 were profiled by spatial transcriptomics. Fresh frozen samples were sectioned at 10um thickness with a cryostat onto 10X Visium Spatial Transcriptomics slides (Visium Spatial Gene Expression Slide and Reagent Kit, PN-1000184) with spatially barcoded capture areas according to the manufacturer's instructions. Tissues sectioned onto Visium slides were profiled either on the same day as sectioning or within 1 week following storage at -80C.

H&E staining and imaging

Tissue sections on Visium slides were first fixed in methanol (Millipore Sigma #34860) followed by an aqueous eosin-based H&E protocol according to manufacturer's instructions (10X Visium Methanol Fixation, H&E Staining, and Imaging Protocol CG000160). Brightfield imaging was performed using a wide-field Leica DMIi8 inverted microscope (Leica-microsystems CMS GmbH Germany) equipped with a DFC310FX color camera. Images were acquired with a 10x/0.25 dry objective and stitched by Leica Application Suite X software. Image post-processing was performed using Fiji version 2.3.1.

10X Visium cDNA synthesis and library generation

Following imaging of H&E staining, permeabilization was carried out on the Visium slide to capture mRNA released from the tissue. The optimal permeabilization time (9 minutes) was determined by a permeabilization time course experiment (i.e., Visium tissue optimization experiment). cDNA synthesis and library generation were performed with the Visium Spatial Gene Expression Slide and Reagent Kit (10X Genomics). Briefly, reverse transcription was performed by a template switch oligo protocol to generate a second strand and cDNA synthesis was carried out according to qPCR results (KAPA FAST SYBR qPCR master mix, KAPA Biosystems). The amplified cDNA was fragmented, end repaired, ligated with index adaptors and size selected with cleanups between each step using the SPRIselect Reagent kit (Beckman Coulter). Quality control and quantification of the resulting dual-indexed barcoded libraries was performed with Agilent TapeStation and by qPCR (NEBNext Library Quant Kit for Illumina, New England Biolabs).

Sequencing

QC of final libraries was performed by Tapestation and paired-end dual indexed final libraries were diluted to 1.8nM, pooled, and denatured prior to sequencing on Novaseq (Illumina) using the Novaseq SP 100 cycles sequencing kit (Illumina) with 1% PhiX and the following sequencing parameters: Read 1 – 28 cycles, Read 2 – 90 cycles, Index 1 – 10 cycles, Index 2 – 10 cycles.

CODEX antibodies and imaging

The focus of our antibody panel was to cover the major cancer cell states as defined by scRNA-seq by Neftel et al.[4] We therefore took the top 50 genes of the AC, OPC, NPC2, MES1 and MES2 gene signatures and created an antibody priority list based on the following criteria: (i) moderate to high RNA/protein correlation as predicted by the Human Protein Atlas[61] (ii) expected protein localization: nuclear > membrane > cytosol, (iii) carrier-free antibody formulation available. All antibodies not commercially available from Akoya and hence requiring custom-conjugation with DNA-barcodes[62] were first validated with conventional immunofluorescence. More than 100 antibodies were rigorously tested to assemble the 40-plex panel. A list of all antibodies can be found in Table S3. Of note: CD19 staining was not performed on 3 samples (MGH258, ZH881_INF and ZH881_T1). Control tissues included healthy duodenum for immune markers, carcinosarcoma for p53, and adequate GBM samples containing the respective cell types and states as identified by Visium. Fresh frozen sections were fixed with 4% PFA for 20 min and blocked with 4% BSA and 0.25% TritonX-100 for 30 min. Primary antibodies were incubated overnight at 4 °C and secondary antibodies incubated for 2h at room temperature. All antibodies (including pre-conjugated antibodies from Akoya) were additionally validated on the respective GBM control samples using multiple CODEX runs to adjust antibody concentration and exposure times to optimize signal to noise ratio. CODEX runs were

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

performed using PhenoCycler-Fusion and imaged with PhenoImager (both Akoya Biosciences) according to manufacturer instructions. After the CODEX run, hematoxylin and eosin staining was performed inside the flow cell using a syringe.

GBM hypoxia organoid experiments

GBM organoids were generated and maintained as previously described.[34] BWH911 organoids were initially derived from an IDH wild-type GBM tumor specimen. Organoids were maintained under continuous mechanical agitation in serum-free media consisting of a 1:1 mix of DMEM:F12 and Neurobasal media supplemented with MEM-NEAAs as previously described.[63] For hypoxic treatment experiments, organoids were maintained under normoxic or 2% oxygen conditions for up to 48 hours under otherwise standard culture conditions, and were subsequently embedded fresh in OCT for CODEX studies.

QUANTIFICATION AND STATISTICAL ANALYSIS

Spatial transcriptomics (Visium) analysis

Alignment, processing, and QC of Visium data

Alignment of FASTQ files with human reference genome GRCh38, UMI counting, and spot barcode filtering were performed using SpaceRanger (versions 1.0 and 1.1, 10x Genomics). Alignment between positionally barcoded Visium spots and tissue images to obtain spatial coordinates necessary to generate spatial maps was performed using Loupe Browser (version 5.0.1, 10x Genomics). Expression levels were quantified as $E_{i,j} = \log_2 (1 + \mathrm{CPM}_{i,j}/10)$, in which counts per million $(\mathrm{CPM})_{i,j}$ refers to $10^6 \times \mathrm{UMI}_{i,j}/\mathrm{sum}(\mathrm{UMI}_{1\ldots n,j})$, for gene i in sample j, with n being the total number of analyzed genes. The average number of UMIs detected per spot was less than 100,000; thus, CPM values were divided by 10 to avoid inflating the differences between detected ($E_{i,j} > 0$) and undetected ($E_{i,j} = 0$) genes as previously described.[64]

For each spot, the number of counts was used as a proxy for sample quality. Spots with fewer than 1000 counts and/or expressing more than 20% mitochondrial genes, another proxy for low quality, were filtered out. The top 7,000 most highly expressed genes were retained and centering was performed per sample in order to define relative expression values by subtracting the average expression of each gene i across all k spots: $\mathrm{Er}_{i,j} = E_{i,j} - \mathrm{average}(E_{i,1\ldots k})$, where Er represents relative expression values.

Thirteen GBM Visium samples profiled by Ravi et al.[9] (Table S1) were included in our analysis. We excluded samples from the Ravi Visium GBM dataset that were annotated as normal cortex and samples in which >50% of spots were excluded after QC filtering (<1000 UMI counts/spot and/or >20% mitochondrial genes/spot). Additionally, two normal Visium cortex samples (UKF256_C and UKF265_C) from Ravi et al.[9,42] were used as a normal brain reference for CNA analysis. No batch integration was required since our analysis is not performed on combined data, but rather is done per sample and only the results of various analysis (i.e., NMF/Leiden-derived sets of genes, spatial coherence scores, colocalization or of other metrics) are combined across samples.

Per sample clustering

We applied two clustering approaches to individual samples in order to capture discrete patterns of variation (Leiden clustering) and continuous patterns of variation (NMF). For each sample, following PCA, Leiden clustering was performed on the SNN graph (implemented with Seurat version 4.3.0). Gene programs were defined per Leiden cluster by differential expression analysis based on the top 50 most differentially expressed genes by the Wilcoxon Rank Sum Test with a p value of <0.005. For tumors in which we had profiled multiple tissue sections from different regions, we also performed per tumor Leiden clustering in which the expression matrices from the individual tissue sections were merged and jointly clustered.

We performed NMF on each sample separately to capture continuous patterns of gene expression variation. Negative values in each centered expression matrix were transformed to zero. To minimize the influence of selection of an individual k parameter, we ran NMF with multiple k values ranging from 2-11, generating 65 NMF programs per sample. Each NMF program was summarized by the top 50 genes based on NMF score.

Generating consensus metaprograms

To integrate across samples, we generated metaprograms (MPs) – consensus gene signatures corresponding to a cell state or cell type. To identify robust gene programs across samples, we jointly clustered both the per sample NMF and Leiden gene programs collected from all GBM samples by their overlap (intersection/union, i.e., Jaccard index).

Given the high number of NMF programs, programs included in the clustering were limited to those that were both robust within a sample (i.e., 30/50 gene overlap between at least two programs across more than one value of k), non-redundant (i.e., for a group of similar programs, defined as at least 20% overlap with another program within a sample, only one is retained), and similar to programs generated from other samples (i.e., robust across samples, defined as at least 20% overlap with a program from another sample).

In total, 492 gene programs from individual samples derived from NMF and Leiden clustering were used to generate metaprograms. The clustering process used to define metaprograms was carried out as described in Gavish et al.[15] as follows: each robust program was compared to all other robust programs to assess the degree of gene overlap between programs. Considering overlap instances of at least 12 genes, the programs with the maximal number of overlaps was selected as a potential founder of a new cluster. If the number of overlapping programs (>12 genes) exceeded four instances, the program with the highest gene overlap to the founder program was added, and thus a cluster was formed. The MP for the cluster was initially defined by the genes that appeared in

both programs such that each metaprogram reflects the most robust genes within each cluster of programs. By this approach, 13 clusters corresponding to 13 MPs were derived from all per sample GBM programs.

In two cases where we manually observed clear subclusters, we further split a single MP cluster to two clusters by hierarchical clustering: cluster 5 and cluster 7 (one subcluster of cluster 7 was excluded due to strong enrichment of mitochondrial and ribosomal genes suggesting low data quality). The MP generation process described above was performed separately for the IDH-mutant samples resulting in 7 IDH-mutant MPs (see spot scoring and assignment of spots to MPs for further details on IDH-mutant MPs). The Vasc MP cluster was also further subclustered by hierarchical clustering to identify cell subtype signatures (Vasc-Ang and Vasc-IMEC). Consensus gene programs were derived from the gene program subclusters selecting for the top 50 most recurring genes per subcluster.

Validation of MP generation approach by simulation

In order to illustrate and validate how our NMF/metaprogram approach generates single-cell equivalent signatures from mixed composition Visium spots, we performed a simulation with synthetic spots. The simulation was composed of 10 samples with 200 synthetic spots per sample. Each spot contained eight synthetic cells sampled from a pool of 1000 cells per spot. To generate synthetic cells, we first defined the cell state fractions in the sample by scaling four randomly selected fractions to sum to 1. Next, we randomly assigned a cell state to each cell by the corresponding cell state fraction. Each cell was assigned a random program score between 1-5 to its assigned cell state and a random score between (-1) to 1 for all other cell states. All scores were then divided by a noise parameter (set to a constant 1.2). In order to simulate correlations between similar programs that occur in real data, in some cases, we generated a correlation between two of the programs by adding the score of one of them to the second multiplied by a coupling parameter.

Our simulation data contained four different cell states; however, by design, the gene-sets represented only three of them, with each cell state represented by 50 genes. Given the programs scores, we defined the gene values in the following way: gene(i) = program(i)_score + rand_genes (i), where program(i)_score is the corresponding cell score to each of the programs and rand_genes(i) is a sampled value out of a distribution generated from GBM single-cell data[4] by vectorizing and shuffling all the gene values after log transformation and per sample centering.

To generate synthetic spots, we sampled eight cells out of the pool of 1000 generated cells. In order to simulate correlations between spot compositions that occur in real data, in some cases, we created a correlation between the composition distribution of two cell types in the following way: First, we generated four data vectors with a normal distribution, two of which were correlated using a correlation level pre-defined by a correlation degree parameter, using mvrnorm from the R package MASS.[65] Next, we performed scaling with a softmax function on each corresponding four data points to generate a percent-like vector. We multiplied this vector by eight and rounded down. The output was then used to determine how many cells to sample from each cell type. The remining cells (out of eight) were sampled randomly from all cell types.

Finally, we ran NMF per sample with k=3 and extracted three clusters. We then defined final clusters derived from a Jaccard matrix comprised of all NMF-generated clusters and compared the consensus MP signatures derived from the clusters to the original program genes. For each original cell type program (org_p) and newly define signature (sig) we calculate the Jaccard similarity score. We define the simulation success rate as the average of the maximum score for each original program. We generated 25 different scenarios by varying each parameter across five values, including gene correlations between pairs of programs and spot composition correlations in order to simulate the correlations that occur in real data. We ran 50 simulations per scenario, yielding expression correlations between -0.27 to 0.93, and composition correlations between -0.33 to 0.49 (averaged across all simulations per scenario).

Gene-set enrichment analysis

MPs typically represented a single cell state or cell type. We assessed the enrichment of MP signatures with Gene Ontology terms (MSigDB modules H, C2, C5, C8) as well as published signatures for glioma cell states,[4,17,26,66] non-malignant brain cell types,[16,18,67–73] and pan-cancer cell states and cell types[15] by hypergeometric test (FDR adjusted p<0.01 was considered significant).

Analysis of new spatial malignant states in single-cell data

We re-examined GBM scRNA-seq data[4] to look for evidence of the new malignant states that we identified by spatial analysis. We used two approaches to identify the new states in single-cell data and determine their frequency:

scRNA-Seq data was scored for the new spatial malignant states. For this analysis, genes from the MES-Ast gene program that overlap with the MES or AC programs were removed from the MES-Ast program and genes from the Chromatin-Reg gene program that overlap with the NPC programs were removed from the Chromatin-Reg program. The actual scoring distribution was then compared to a control score distribution derived by shuffling the single-cell data and cells with a score >99% percentile of the shuffled (control) distribution were classified as positive.

Single cells from scRNA-seq data were scored for all spatial malignant MPs including those representing new states and cells were classified by their highest score with a minimum difference of $\geq$ 0.2 between the highest scoring MP and the second highest-scoring MP.

MP correlations between platforms and final MP signatures

In order to directly compare the core GBM malignant cell states (MES1, MES2/MES-Hyp, OPC-like, NPC-like, and AC-like) across platforms (ST=spatial transcriptomics, SC=single-cell RNA-Seq), correlations were calculated between each spatial MP gene and

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

the MP score of the analogous SC[4] MP in the SC dataset. In order to derive truly single-cell equivalent MPs, a filtering step was added in order to remove genes that were more correlated with another single-cell MP than the MP it currently belonged to according to the following criteria: (i)The gene has a correlation below 0.3 for the spatial MP it currently belongs to (ii) The gene has a higher correlation with a single-cell program different from the MP it belongs to (iii) The difference between the correlation of the best-matching single-cell program and the MP it belongs to is >0.2. Lastly, mitochondrial and ribosomal genes were removed from the final MP gene signatures.

Spot scoring and assignment of spots to MPs

Spots were scored for MPs as previously described using the scalop R package (https://github.com/jlaffy/scalop).[4,8] Given a set of genes (*Gj*) reflecting an MP corresponding to a cell state or cell type, we calculate for each spot *i*, a score, *SCj(i),* quantifying the relative expression of *Gj* in spot *i*, as the average relative expression (*Er*) of the genes in *Gj*, compared to the average relative expression of a control gene set (*Gj cont*): *SCj (i)* = average[*Er*(*Gj,i)*] − average[*Er*(*Gj cont,i*)]. The control gene-set is defined by first binning all analyzed genes into 30 bins of aggregate expression levels (*Ea*) and then, for each gene in the gene set *Gj*, randomly selecting 100 genes from the same expression bin. In this way, the control gene set has a comparable distribution of expression levels to that of *Gj*, and the control gene set is 100-fold larger, such that its average expression is analogous to averaging over 100 randomly selected gene sets of the same size as the considered gene set. Spots were assigned to the MP for which it scored most highly. For the analysis in which single cells from Neftel et al.[4] were scored for the spatial MPs, a stricter approach was used for assignment of MPs to cells. In this case, in order to be confidently annotated with a given MP, cells needed to have a minimum score of 1 for the highest-scoring MP and a difference of at least 0.1 between the highest and second-highest MP scores.

Due to the relatively low number of IDH-mutant samples (n=6) and corresponding MPs generated from them when they are considered separately (n=7), when assigning IDH-mutant spots to MPs we integrated the IDH-mutant MPs with GBM MPs according to the following procedure. First, we scored IDH-mutant spots for the GBM MPs with the following thresholds: if an IDH-mutant spot scores highest for a GBM MP, in order for it to be assigned to a GBM MP, it must have a minimum score of at least 1.5 for the MP and there must be a difference of at least 0.2 between the highest and second-highest MP score. If there is no equivalent IDH-mutant MP, and at least 5% of spots could be confidently assigned to the GBM MP, then that GBM MP was added to the list of IDH-mutant MPs for scoring and assignment of IDH-mutant spots. By these criteria, the following GBM MPs were added to the IDH-mutant MP list: Prolif-Metab, Reactive-Ast, Inflammatory-Mac, MES-Hyp, and Chromatin-Reg.

Spot homogeneity (confidence) score

We derived a homogeneity score for each spot by subtracting the second highest MP score of the spot from the highest MP score. Most spots are dominated by one or two MPs such that considering the two highest scoring MPs within each spot largely captures its composition, and the difference between those scores reflects the confidence of the assignment to the MP with the highest score.

Validation of spot assignment approach

Multiple analyses were performed to assess the validity and downstream implications of classifying spots by their dominant MP. In order to compare MP sample composition as calculated by the dominant MP annotation approach to MP sample composition as calculated from cell state/cell type fractions, we performed deconvolution to infer spot composition. Deconvolution was performed by scoring spots for MP signatures and then performing a softmax function on all spot scores $\geq$ 0.1 (where a score of $\geq$ 0.1 was considered as 'presence' of a given cell state/type within a spot) to generate inferred fractions per spot summing to 1. These values were then used to calculate overall MP abundance compared to MP abundance as calculated by the dominant MP spot classification approach. We also performed a simulation with synthetic spots to evaluate the accuracy of our spot annotations using the dominant MP approach across spot homogeneity score (i.e., confidence score) filtering thresholds (by quantiles). We generated synthetic spots by sampling single cells from the Neftel dataset.[4] Each synthetic spot contains the average scores for single-cell MPs of 8 cells. Each spot is then assigned to the highest scoring MP. We also calculate a homogeneity score and the true composition of each spot. A correct assignment is defined as an assignment matching to the most dominant cell type in a spot.

CNA inference

CNAs were estimated as described previously, sorting genes by their chromosomal location and calculating a moving average of gene expression with a sliding window.[5,58] We used an increased size sliding window of 150 genes as ST data is noisier than single cell data, and two normal Visium cortex samples (UKF256_C and UKF265_C) from Ravi et al.[9,42,9] were used as a normal brain reference to define a baseline for normal karyotype. We then scored each spot for ''CNA signal'', defined as the mean of the absolute of CNA values across regions with high CNAs, and ''CNA correlation'' which refers to the correlation between the CNA profile of each spot and the average CNA profile of all spots of the same sample. Following this we reclassified spots using the two measurements, and spots within the query sample with low scores were integrated with the external reference. Finally, we performed another iteration after which we scored again each spot for CNA signal and correlation. To infer spots and MP malignancy level we used the CNA correlation score plus a scaled CNA signal score: $CNAcor + \left(CNAsig * \frac{\max(CNAcor)}{\max(CNAsig)}\right)$. We then performed a min-max scaling for the final score.

Spatial coherence score

The spatial coherence score of a specific MP in a sample is defined as the scaled average number of immediate neighbors of the same MP across all MP spots. To scale the observed spatial coherence score we shuffle the spots positions 100 times and perform a similar calculation over the shuffled samples, averaging across the results to generate an expected minimum value. We also define

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

a maximum expected value by modeling the average number of immediate neighbors of the same MP where all the MP spots are organized in a regular hexagon pattern. The final score is derived as follows: $\frac{observed_{Score} - expected_{min}}{expected_{max} - expected_{min}}$

We validated spatial coherence using Moran's I as an alternative measure. We calculated Moran's I, per MP per sample, based on a scoring-based spot deconvolution (described in *"Validation of spot annotation approach"*). To directly compare it to our spatial coherence calculation, we defined only adjacent spots as neighbors. As a control, we also calculated Moran's I per MP per sample after shuffling the spatial location of spots.

Defining organizational zones

In order to assign each spot as *structured* or *disorganized* we first generate subsamples of three different sizes (radii 5, 8, 11) surrounding each spot. Next, we calculated the spot's spatial coherence score in each subsample by averaging the spatial coherence of all MPs in the subsample. Each spot is defined as *structured* if it scores higher than the 40% quantile (across all spots from all samples) for each of the three different window sizes, and *disorganized* if it scores lower than the 40% quantile for each of the three different window sizes. Otherwise, the spot is defined as intermediate. To smooth out the resulting regions, we set a sliding window with $r = 4$ surrounding each spot and make a final region assignment based on the assignment of the majority of spots, to either *structured, disorganized* or intermediate, within the window. After the initial division to *structured* and *disorganized* regions, we performed a second division by malignancy level of the spots. We inferred the spots malignancy level based on their CNA as described above. Finally, we scaled the score to range from 0 to 1. Spots that were classified as *structured* and had a malignancy score $\geq$ 0.4 were classified as *Structured-Malignant* and those with a malignancy score <0.4 were classified as *Structured-Normal*. Spots that were classified as disorganized and had a malignancy score $\geq$0.5 were classified as *Disorganized-Malignant* and those with a malignancy score <0.5 were classify as *Disorganized-Normal*.

Validating structured vs. disorganized regions

We used the spot homogeneity score to validate the division to structure and disorganized regions in two ways: First, we compared the mean homogeneity score of each MP between spots that were annotated as structured and those labeled as disorganized. Second, we compared the fraction of adjacent spots of the same MP in structured vs. disorganized spots across different spot homogeneity filtering thresholds to show that difference in organization remains across varying spot homogeneity levels.

MP spatial coherence scores and proximity to MES-Hyp

Using a sliding window of r=11, we classified regions as MES-Hyp high if they had at least 10% MES-Hyp spots within the window. We then calculated the mean spatial coherence per malignant MP in MES-Hyp high vs. MES-Hyp low windows per sample and averaged across samples.

Spot colocalization

We first performed a scoring-based deconvolution (per sample) on all spots by scoring spots for MPs as described in 'Spot scoring and assignment of spots to MPs'. For each spot, all MP scores above 0.1 were considered as presence of the MP in a spot and deconvolution matrices were binarized such that for a given spot, presence of MP=1 and absence of MP=0. The following analysis was performed separately on spots in structured and disorganized regions per sample. First, for MP *a* and MP *b*, we compute the proportion of spots containing *a* or *b* (denoted as Ta or *Tb*). Second, we compute *N*, the number of times *a,b* co-occur within the same spot. Finally, we compute a colocalization value defined as *N/ Ta+Tb*. After calculating the mean colocalization of each MP pair per sample (the observed colocalization value), an expected colocalization value was computed by performing the measure on 500 shuffled deconvolution matrices per sample. A *p* value was then computed by Fisher's exact test. The per sample effect size was defined as *expected/observed* per MP. Colocalization of each MP pair per sample was defined as significant if the effect size was $\geq$ 1.3 and the *p* value was $\leq$0.01. Finally, the proportion of samples for which a MP pair was considered significant was calculated to consider robustness across samples. For downstream analysis, the mean colocalization value per pair and the proportion of samples for which it was significantly colocalized were used.

Spot adjacency

Spot adjacency was calculated per sample. The observed asymmetrical adjacency between MP *A* and MP *B* within a given sample is defined as: $\frac{\#immediate\_neighbors_{B\ of\ A}}{\#immediate\_neighbors_{allMP\ of\ A}}$.

In order to regress out the spatial coherence effect of MP *B* on its adjacency, we normalize this score by the *relative adjacency capacity* of MP *B,* i.e., the number of MP *B* spots with non-MP *B* neighbors out of all MP adjacency capacities (except for MP *A*): $\frac{MP\ \#spots\ with\ non\ MP\ neighbors_{B}}{\sum_{all\ MP\ but\ A} MP\ \#spots\ with\ non\ MP\ neighbors}$

Next, we shuffle the spots positions 10,000 times and calculate the expected adjacency scores to create a null distribution. The *p* value is computed by the number of times the *observed* score was higher/lower than the shuffled scores (the smaller of the two).

Regional composition

Regional composition was computed per sample. For each radius (defined as the number of hexagons circling a spot) ranging from 1 to 15, we defined a window around each spot. We then calculated the abundance of all MPs in the window and finally, calculated the correlations of abundance between pairs across all the windows of the same size, omitting windows that included only one MP. We then repeated the process over 500 shuffled permutations of the spot positions to generate a random distribution, based on which we calculated significance for the correlation by Fisher's exact test. In order to test the impact of varying spot homogeneity on spatial relationships, we calculated the regional composition for all MP pairs per sample filtering across spot homogeneity score thresholds

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

(set by quantiles: >15%, >30%, >45%). We then calculated the correlation between the original unfiltered regional composition and regional composition at each filtering threshold across multiple window sizes (1-7), per MP pair per sample.

Defining consensus interactions

To be able to compare scores of pairs between analyses, we scaled the scores of all analyses by a min-max scaling so that all scores range between 1 to -1. For each analysis after scaling, a MP pair that exhibited a mean scaled score above 0.35 across samples and was significantly connected (by Fisher's exact test) in at least 20% of samples was considered 'strongly connected'. A MP pair was defined as consistently strongly connected across analyses (i.e., consensus interaction) if it was 'strongly connected' within at least 2 out of 5 analyses (colocalization, adjacency, regional composition with window sizes r= 5,8,15) including at least one of the measures of direct coupling: *colocalization* or *adjacency*. By this definition, 21 structured pairs were consistently strongly connected compared to only 2 pairs in disorganized regions.

Generation of layers network graph

We generated a graph of robustly coupled pairs in structured regions across analyses with Cytoscape (version 3.9.1) by the following criteria: pairs that are defined as "strongly connected" in any 2/5 measures of spatial relationships (see 'defining consensus interactions'). By this definition, we include an additional 4 pairs beyond the 21 pairs initially defined as consensus interactions. All MPs except for the Chromatin-Reg MP could be confidently assigned to layers. Chromatin-Reg is unassigned in the layers model due to its internal inconsistency in coupling between measures in the Visium analysis and inconsistency between the Visium and CODEX analysis.

MP frequency in infiltrative vs. non-infiltrative environments

We defined a spot to be in an infiltrative environment if 30% of the spots surrounding it, within a window of a radius of 4 spots, are annotated as normal brain (i.e., Neuron, Oligo, or Reactive-Ast). Next, we measured the overall frequency of each MP in the two environments (infiltrative and non-infiltrative) across all the samples.

T-cell inference

To infer the presence of T-cells within spots, we lowered the UMI count QC filter to 150 to include low complexity spots due to the association between T-cells and low RNA content/low complexity. Spots that include at least two counts for at least two canonical T-cell markers (CD2, CD3E, CD3G, CD3D, FOXP3, CTLA4, CD28) were identified as putative T-cell containing spots. Putative T-cell-containing spots with at least one CD8A or CD8B count and more CD8A/CD8B counts than CD4 counts were further classified as CD8 T-cell-containing and spots containing at least two CD4 counts and more CD4 counts than CD8A or CD8B counts were further classified as CD4 T-cell containing spots with the caveat that macrophages can also express CD4.

Cell cycle inference

Using the cell cycle metaprogram gene signatures obtained from Neftel et al.[4] we scored each spot per sample for "G1/S" and "G2/M" programs and maintain the higher of the two as the spot cell cycle signal. We set a threshold of 0.85 for both GBM and IDH-mut samples and in both cases, this corresponds to 90% to 95% quantiles, above which a spot was classified as cycling. Next, we calculated the frequency of spots classified as cycling within each MP. The results remained overall consistent within each group at different thresholds.

Comparison to IVY Gap transcriptional programs

To compare between the transcriptional programs derived from laser micro-dissected histopathological features as part of the IVY Gap Atlas (also referred to as ("IVY Gap programs")[43] and the state layers, we derived layers signatures (Table S2). We performed DEG analysis between layers per sample (including only structured samples) to generate layers signatures. Final signatures included genes that received an adjusted p value < 0.05 and a $\log_2$(fold change) > 1 in at least 5 different samples. We then calculated Jaccard similarity between the layers programs and IVY Gap programs.

Associations between MPs, organization, and functional states

Using published GBM functional signatures for Connectivity and Invasion,[26,30] we classified spots as "Connected", "Invading", or "Unannotated" as follows: Spots with an absolute score ≥ 0.75 for "Connected" or "Invading" and ≥ 0.2 difference between the two scores were annotated as "Connected" or "Invading"; all other spots were unannotated. Enrichment values between "Connected" and "Invading" spots and cell cycle, spatial MPs, state layers, and organizational zone were calculated per sample by hypergeometric test and then averaged across samples. An effect size was calculated by the proportion of samples for which a given enrichment was significant. In order to analyze the overall spatial relationship between "Connected" and "Invading" spots, we calculated the mean spatial proximity of connected and invading spots, across sliding windows using the regional composition calculation (see "regional composition").

CODEX analysis

Segmentation of CODEX images

Nuclear segmentation was performed with the StarDist[55] plugin within QuPath[54] applying the pre-trained model (dsb2018_heavy_augment.pb) to the DAPI channel. Areas with staining or tissue artifacts (i.e. tissue folds, freezing artifacts, outer border of the tissue) were excluded manually from segmentation and further analysis. Nuclei were expanded by 3 μm to conservatively estimate cell boundaries. Mean fluorescence intensity for every marker and cell compartment, together with morphology quantifications (area, solidity, circularity) were exported and further analyzed in R.

Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

Cell phenotyping of CODEX data

Cell phenotyping in highly multiplexed imaging data is challenging because of technical artifacts (i.e. autofluorescence, segmentation artifacts) that differ from other single cell technologies.[74] We followed a workflow as suggested by Hickey et al. who benchmarked multiple normalization and clustering methods specifically for CODEX data[75,76]: (i) Log_2 transformation and z-score normalization per sample (ii) Over-clustering of the data using PhenoGraph (PMID: 26095251) with k= 10, such that even rare cell types (i.e., B-cells, neurons) would be assigned to a distinct cluster (iii) Calculating average protein expression per cluster (iv) Cluster assignment following the criteria described in Table S4. Clusters with the same assignments were merged. For phenotyping, we excluded markers with a staining pattern that was not limited to the cell boundaries and thus resulted in spillover artifacts (PLP1, CD44, GAP43, CD90, BCAN, FN1, GLUT1, CA9). T-cells were defined by CD3 expression only and in a second step sub-typed using functional T-cell markers (CD4, CD8, CD279, CD69). Ki67 expression was used to define proliferating cells (z-score > 2.1) after cell identities were defined. The calling for malignant vs. non-malignant cells was based on SOX2 and EGFR staining (Table S4).

Clean clusters were defined as depicted in Table S4 and named accordingly. Mixed or ambiguous clusters (i.e., clusters that fulfilled > 1 criteria of Table S4) were further sub-clustered. T-cell clusters were always sub-clustered to exclude CD3 spillover artifacts that were specifically high in this cell population. In cases where the sub-clustering did not derive clean clusters, cells were excluded from further analysis (n=96,327 of 500,738 cells, 19,2%). This included the following categories: (i) artifact (i.e., autofluorescence, dust, tissue fold) as identified by unreasonable protein co-expression and visual inspection in QuPath. (ii) low expression as identified by top marker in cluster z-score < 0.5 (iii) low solidity and circularity as defined by z-score < -2 reflecting irregular and complex nuclear shapes corresponding to segmentation artifacts (iv) small cluster as defined by cluster size < 150 cells (low abundant cell types like T- and B-cells were excluded from this criteria) (v) spillover as defined by unreasonable protein co-expression and visual inspection in QuPath (i.e., T-cell/Vasc, Mac/Vasc, Mac/MES-Hyp). (vi) unknown cell phenotype as defined by the remaining clusters containing n=27.865 cells from 8 samples. These "unknown" clusters were heterogeneous and not redundant across samples. Hence, we could not annotate them to a unique proteomically defined cell type/state.

Creation of pseudospots

We created and overlaid a virtual tissue microarray (TMA) grid corresponding to the Visium spot properties (diameter=55μm, centroid distance=100 μm) to every sample in QuPath. Cells were associated to pseudospots if their nuclear segmentation mask centroid was within the pseudospot boundaries. The identity of a pseudospot was defined by the most abundant cell type/state within each pseudospot.

Spatial alignment of CODEX to Visium data

To spatially align CODEX to Visium data we used "STalign".[32] In short, STalign uses Large Deformation Diffeomorphic Metric Mapping (LDDMM[77]) to align CODEX to Visium datasets using image varifolds. A rasterized density image is created from the centroid positions of the CODEX data and serves as source image. In the first step, the source image was aligned to the target H&E Visium image based on manually assigned landmark points using affine transformations. Visually, this resulted in a good alignment in half of the samples (n=5). The remaining five samples were subjected to a second step in which the source (CODEX density image) and target (Visium H&E) images were aligned to minimize their dissimilarities using a LDDMM framework. This diffeomorphic solution generates a smooth, continuous, invertible transformation which permits mapping back and forth from the rasterized image (CODEX) original cell positions (centroids) while respecting the biological constraints such that cell neighbor relationships stay relatively the same. Next, we generated aligned CODEX pseudospots using the Visium spot coordinates and assigning single cell centroids from CODEX data to a pseudospot if the distance is < 27.5μm. To compare similarity between aligned CODEX and Visium samples we included only spots that contained information from both platforms (overlapping spots) and included only MPs that can be detected in both datasets (Prolif-Metab and MES-Ast excluded). To compare the similarity between the CODEX and Visium spatial maps, we segmented each sample into tiles (n=9). We used the tiles to calculate per sample Pearson correlations for each MP individually. We then took the per sample median of the MPs correlations and assessed statistical significance ($p<0.05$) by using a t-test comparing against 0.

Defining organizational zones

To define spatial patterns (*structured/disorganized*) in the CODEX samples we used CODEX pseudospots and performed a similar process to Visium with window size of r=5 and 8 and a similar threshold of higher than 40% quantile across all spots from all samples for each window size in order to define structured regions and lower than the 40% quantile to define disorganized regions.

Minimum distance to vascular cell

We created a distance matrix for each samples using the x- and y-coordinates of the segmentation masks centroids and calculated the minimum distance from every cell type to the closest vascular cell.

Colocalization

We created a neighborhood matrix for each sample using the imcRtools package in R. In the neighborhood matrix, rows represent cells within a sample and columns represent the count of every cell type in that neighborhood. The neighborhood for every cell was defined by a radius of 27.5μm (coloc1: corresponding to the dimensions of 1 Visium spot) or 127.5μm (coloc3: corresponding to the dimension of 3 Visium spots). We then calculated the mean interaction count between every possible cell type pair. Example: For cell type *A*, we calculated the sum of counts that cell type *B* is in the neighborhood of cell type *A* and divided this value by the count of cell type *A* in a given sample. To focus on co-localizations between different cell types (rather than emphasizing self-pairs), we excluded neighborhoods that were 80% dominated by a single cell type/cell state. This filtering was important to detect neighboring cell types

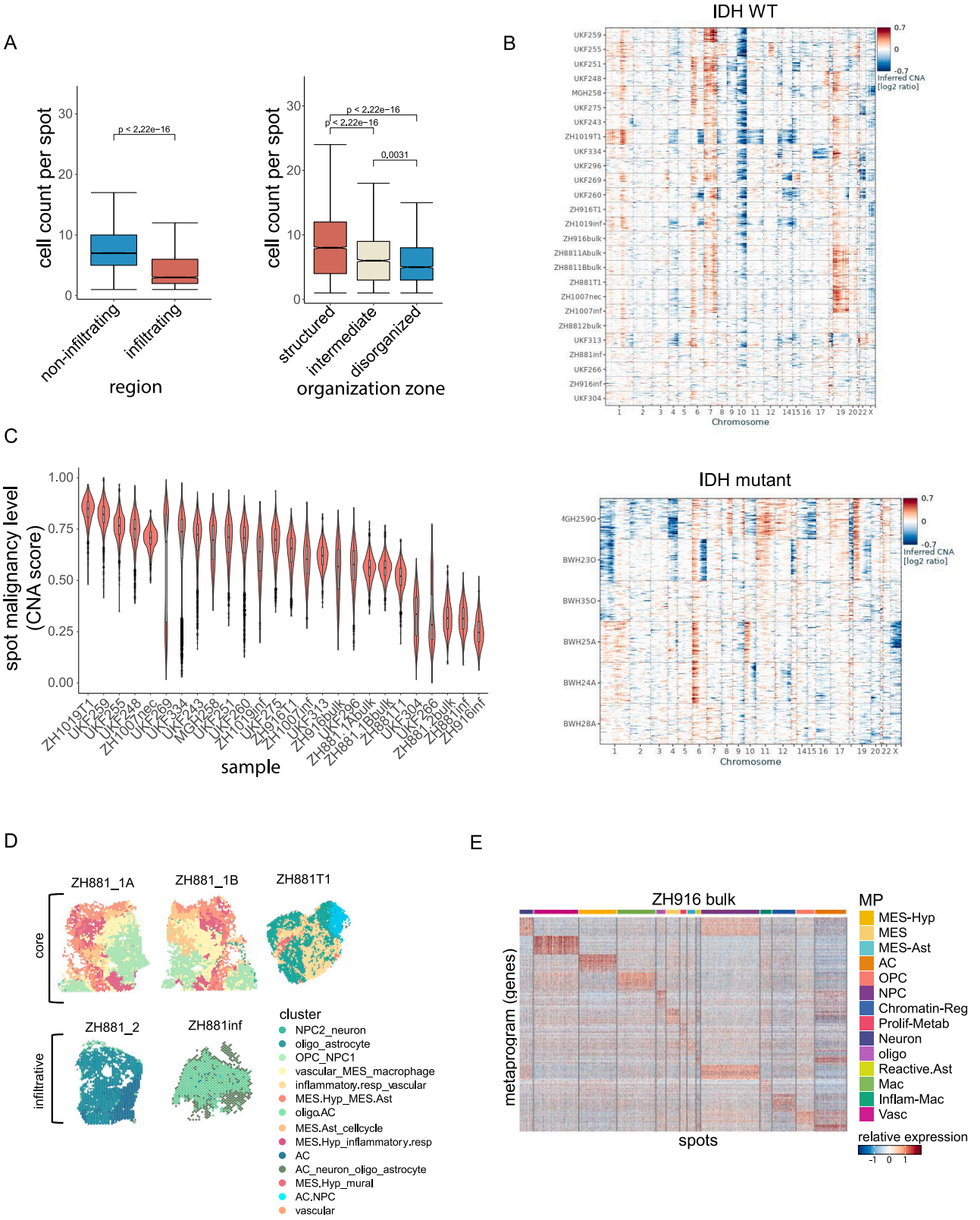
Please cite this article in press as: Greenwald et al., Integrative spatial analysis reveals a multi-layered organization of glioblastoma, Cell (2024), https://doi.org/10.1016/j.cell.2024.03.029

(i.e., MES) next to cell types with very high spatial coherence (i.e. MES-Hyp). Next, we shuffled cell type labels *n*=500 times and calculated z-scores from the observed and shuffled (expected) interaction counts. Significance was reached if 95% of shuffled z-scores were larger or smaller (depending on directionality: coupled vs. uncoupled) than the observed value.

Regional composition

CODEX regional composition was calculated similarly to Visium regional composition on the single cell level, using the same radii as used for Visum. For each radius *i* (defined as the distance between *I* number of spots) ranging from 1 to 15, we defined a window around each cell. We then calculated the abundance of all MPs in the window and finally, calculated the correlations of all pairs' abundance across all the windows of the same size, omitting windows that included only one MP. We then repeated the process over 500 shuffled permutations of the spots positions to generate a random distribution from which we calculated significance for the correlation by Fisher's exact test.

Supplemental figures



(legend on next page)

Figure S1. Segmentation, CNA inference, and per-sample clustering, related to Figure 1

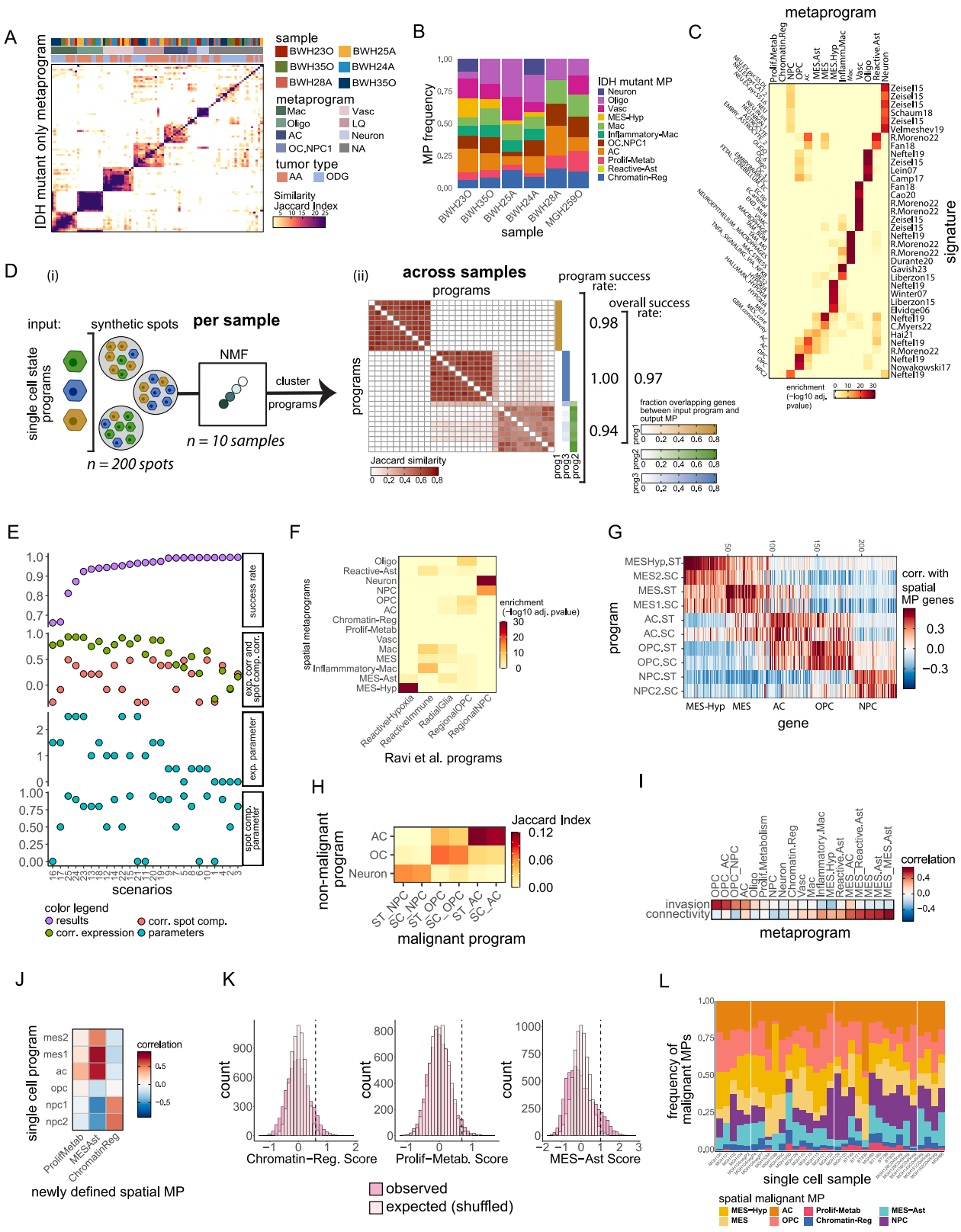
(A) Segmentation analysis. Left panel: cell density (cell count per 55 μm spot) for non-infiltrative and infiltrative samples. Right panel: cell density (cell count per 55 μm spot) for structured, intermediate, and disorganized regions as classified by spatial coherence analysis on CODEX pseudospots. Boxes indicate the median and the 1st and 3rd quartiles. The upper and lower whiskers extend to maximal and minimal values no further than 1.5 times the interquartile range from the 3rd and 1st quartiles. Data points outside the interval are represented as points, *p* values indicated on the plot calculated by t test.

(B) Copy-number aberrations (CNAs) across all samples for GBM WT (top) and IDH-mutant (bottom) cohorts. CNAs were inferred by average relative expression in windows of 150 analyzed genes. Rows correspond to spots arranged by malignancy level inference (see STAR Methods), columns correspond to genes arranged by chromosomal position.

(C) Spot CNA score by sample for GBM samples, boxes, and whiskers as defined in (A).

(D) In tumors for which there were multiple samples, Leiden clustering was performed per tumor jointly across the tissue sections (in addition to per sample).

(E) Spots vs. MP genes in one sample (ZH916), values are log-transformed normalized expression.



(legend on next page)

Figure S2. Validation and further analysis of spatial metaprograms, related to Figure 2

(A) All Leiden and NMF gene programs across all IDH-mutant samples clustered by gene overlap (Jaccard index) to derive IDH-mutant metaprograms.

(B) Composition of IDH-mutant samples by MP.

(C) Enrichments of spatial metaprograms (columns) with gene sets (rows) previously defined from studies indicated at the top. Enrichment values calculated by hypergeometric test; significance defined by $-\log_{10}$ FDR-adjusted p value < 0.01.

(D) (i) Scheme showing generation of synthetic spots and NMF simulation. (ii) Jaccard similarity matrix of generated output signatures and benchmarking results for the parameters: expression −0.3, composition −0.36 (STAR Methods).

(E) Summary of success rate and the generated correlations and input parameters of each simulation scenario. Success rates and correlations are averaged across 50 simulations.

(F) Enrichment of GBM spatial MPs with the five GBM spatial regional states defined by Ravi et al.[9,42] (enrichment calculated by hypergeometric test, significance defined as $-\log_{10}$ FDR-adjust p value < 0.01).

(G) Correlations between core malignant spatial metaprograms (per gene) with the equivalent metaprogram across each platform (ST, spatial transcriptomics; SC, single-cell RNA-seq). For each platform, GBM data of that type were used to calculate the correlations.

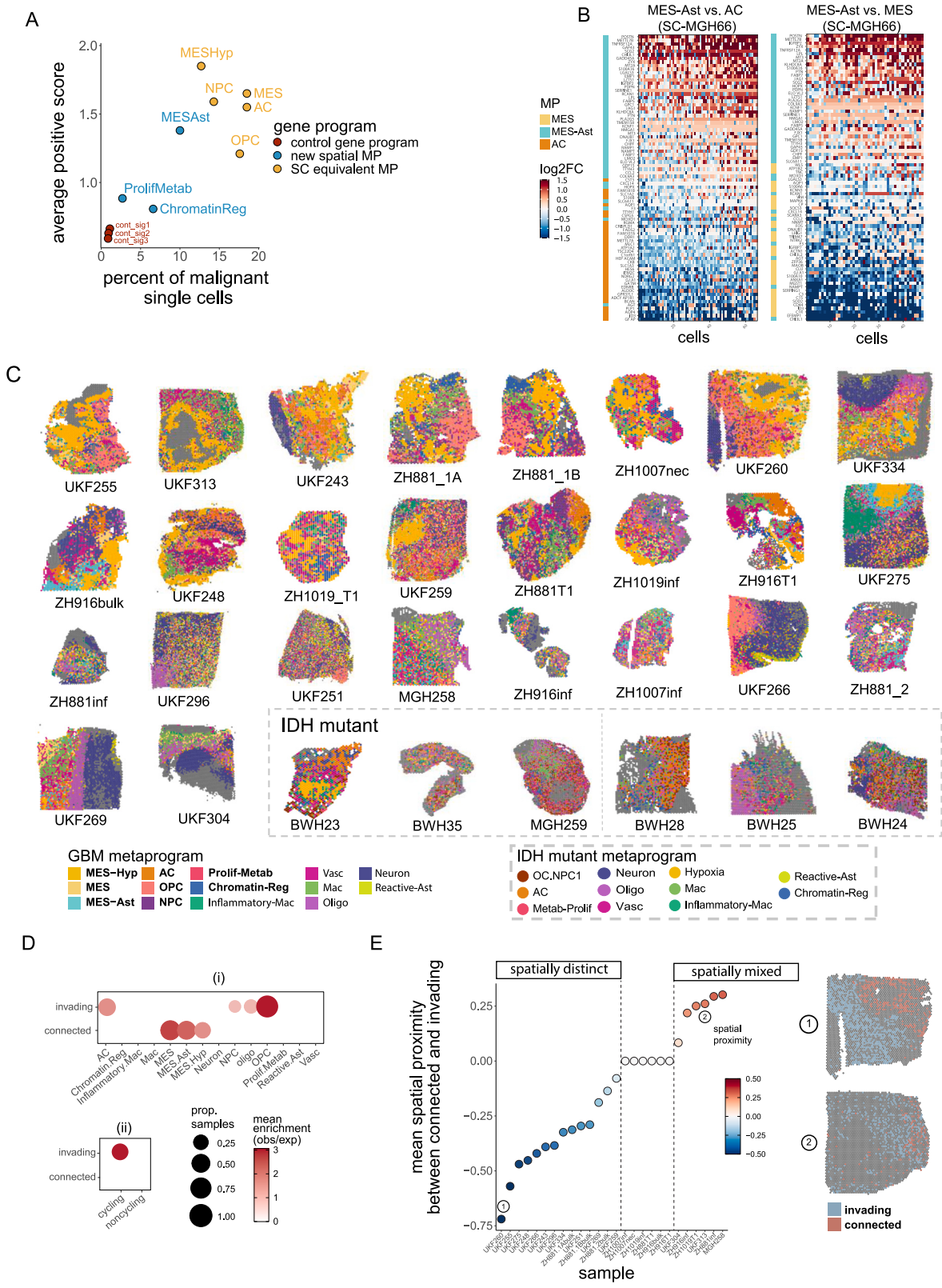
(H) Jaccard overlap of single-cell programs with consensus non-malignant single-cell programs for corresponding states and Jaccard overlap of spatial MPs with consensus non-malignant single-cell programs for corresponding states. ST, spatial transcriptomics; SC, scRNA-seq.

(I) Pearson correlation between program scores for spatial MPs and MP combinations with Connectivity and Invasion signatures[24,26] averaged across samples.

(J) Correlations between core malignant GBM single-cell states and newly described spatial malignant states in scRNA-seq data.

(K) Histograms showing the distribution of scores for newly defined spatial metaprograms in scRNA-seq data (dark pink) compared with a shuffled control distribution (light pink). Cells passing the threshold (dashed line, ≥99% percentile of the shuffled distribution) were classified as positive for that state.

(L) Distribution of malignant states (including newly defined spatial states) per sample in scRNA-seq data.



(legend on next page)

Figure S3. Mapping spatial metaprograms to cell states, related to Figure 2

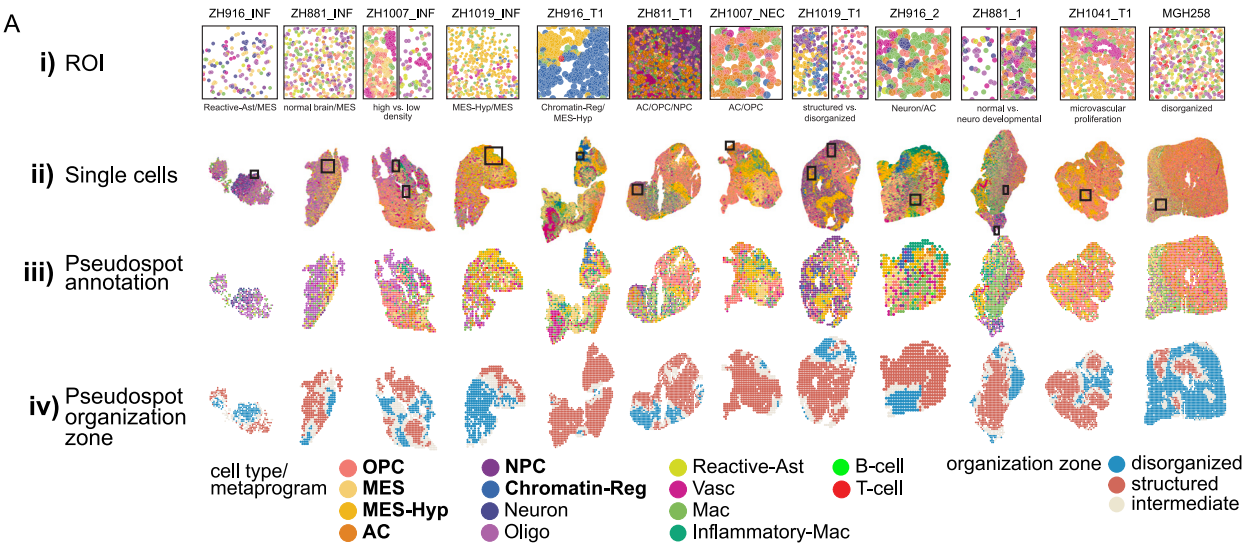
(A) Percent of malignant single cells in each state (x axis) vs. average positive score among cells classified to that state (y axis). Malignant states are colored by their category (core malignant GBM states = yellow; new spatial states = blue; control gene signatures = red).

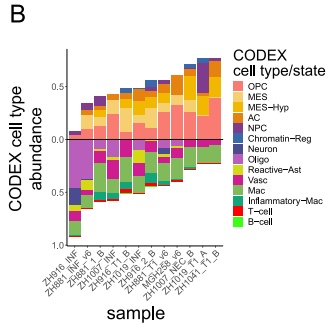
(B) Relative expression (log_2 fold-change) of AC and MES-Ast genes (left) and MES and MES-Ast genes (right) in scRNA-seq data from MGH66 in Neftel et al.[4]

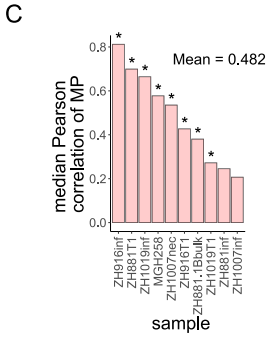
(C) Spatial maps showing spot annotation by MP. Spots were scored for MPs and annotated by their maximum score. Samples are ordered by hypoxia abundance.

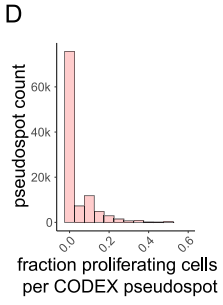
(D) Mean enrichment of Invading and Connected spots by spatial metaprogram; dot size represents the proportion of samples for which a given pair was significantly enriched ($p \leq 0.01$) by hypergeometric test. (ii) Mean enrichment of Invading and Connected spots by cell cycle; dot size represents the proportion of samples for which a given pair was significantly enriched ($p \leq 0.01$) by hypergeometric test.

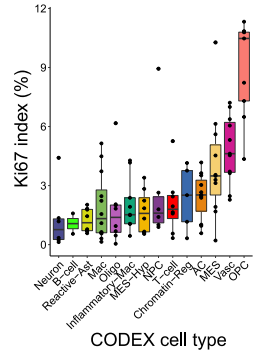
(E) Mean spatial proximity of Connected and Invading spots, across sliding windows. Spatial proximity is calculated by the regional composition calculation (STAR Methods) with spots annotated as Connected, Invading, or neither. "Spatially distinct" is defined as mean spatial proximity < 0 between Connected and Invading spots. An example of a sample in which Connected and Invading are spatially distinct is on the top (UKF260), and an example of a sample in which Connected and Invading are spatially mixed is on the bottom (UKF313).

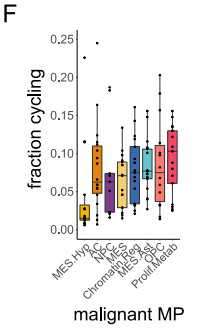


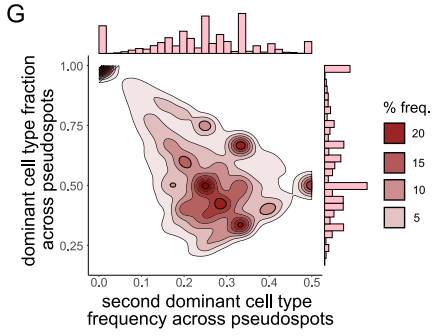


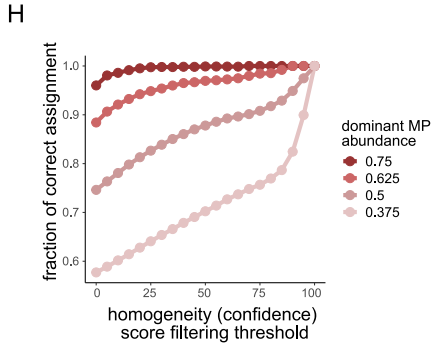


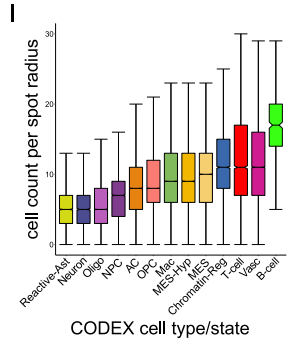


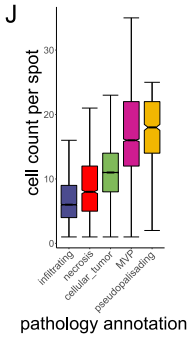












(legend on next page)

Figure S4. GBM profiling by CODEX, related to Figure 3

(A) Spatial maps of CODEX samples: (i) region of interest (ROI) highlighting pairwise cell state relationships. (ii) cells annotated by cell types/cell states. (iii) pseudospots annotated by the most abundant cell type per pseudospot. (iv) pseudospots annotated by organizational zone defined by spatial coherence.

(B) Stacked bar plot showing the relative cell state/type abundance per sample.

(C) Median Pearson correlation indices of MPs per sample in near-adjacent sections profiled by CODEX and Visium following alignment by STalign, statistical significance ($p < 0.05$) by t test.

(D) Histogram showing the distribution of pseudospots by fraction of proliferating cells.

(E) Boxplot of percent Ki67+ cells per sample across cells of each state. Boxes indicate the median and the 1st and 3rd quartiles. The upper and lower whiskers extend to maximal and minimal values no further than 1.5 times the interquartile range from the 3rd and 1st quartiles. Data points outside the interval are represented as points.

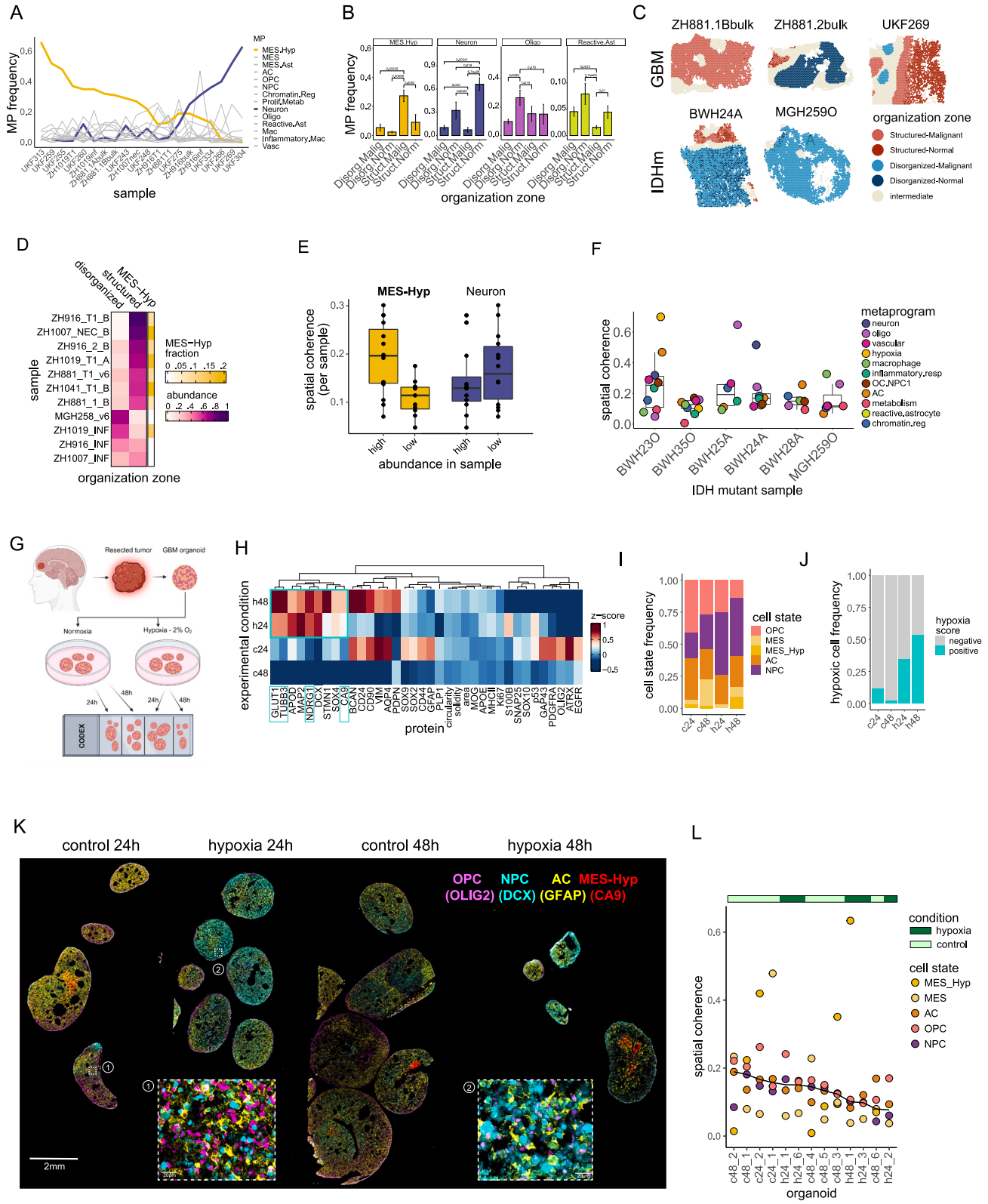
(F) Boxplot showing the percentage of Visium spots containing inferred cycling cells per sample, across spots annotated for a given malignant state, boxes, and whiskers as in (E).

(G) CODEX pseudospot frequencies of dominant cell types. The density map shows the joint frequency of the top two dominant cell types within a pseudospot. y axis corresponds to the top dominant cell type frequency within a pseudospot. x axis corresponds to the second most dominant cell type frequency within a pseudospot.

(H) Percent of correct assignment of single-cell generated pseudospots across homogeneity (i.e., confidence score) filtering thresholds (by quantiles).

(I) Boxplot of cell density (cell count per 55 μm diameter) per cell state, boxes, and whiskers as defined in (E).

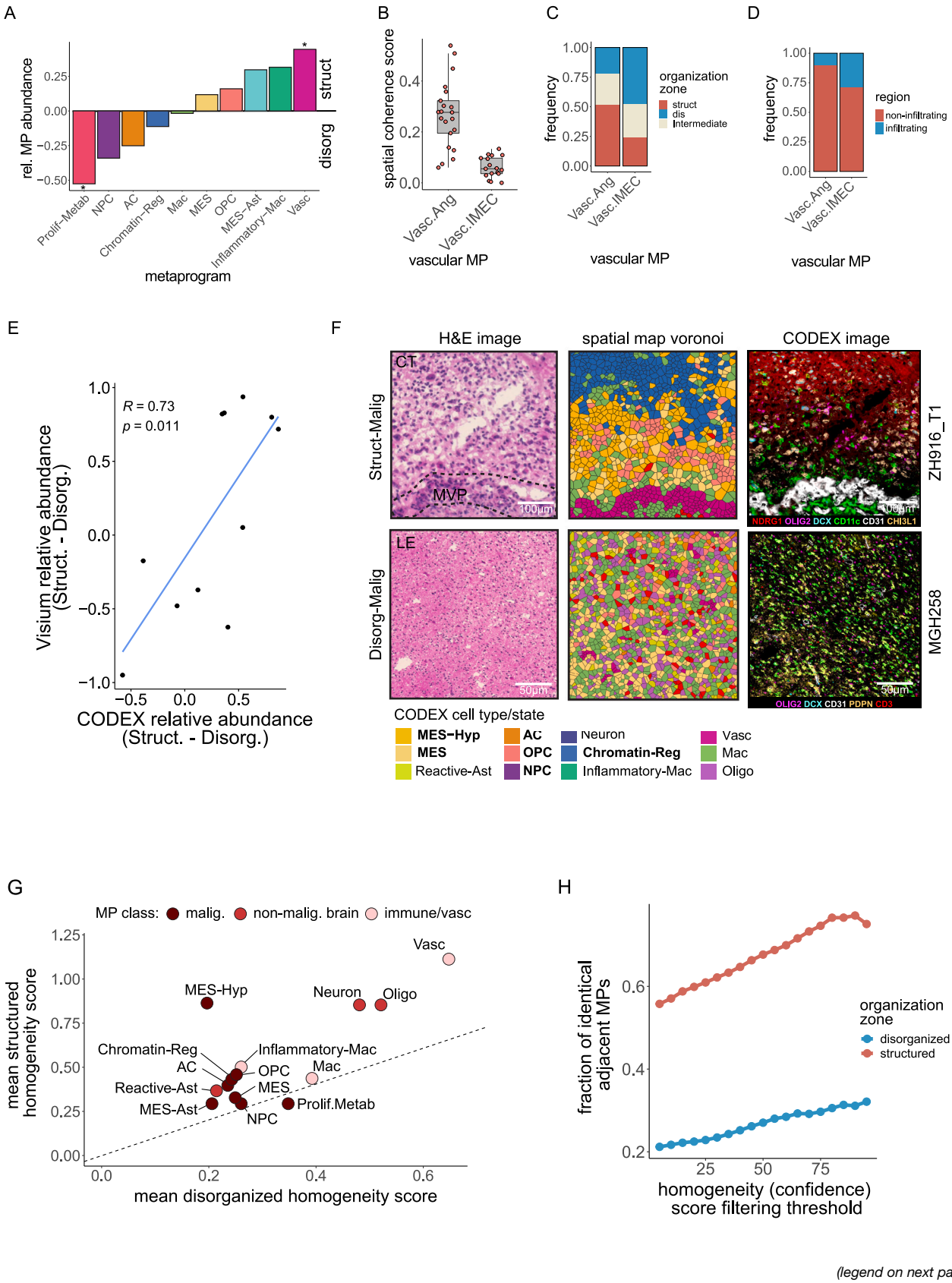
(J) Boxplot of cell density by histopathology annotation from H&E, boxes, and whiskers as defined in (E).



(legend on next page)

Figure S5. Association between spatial coherence and MES-Hyp, related to Figure 4

(A) Abundance of each MP across GBM Visium samples. MES-Hyp and Neuron MPs are colored. (B) Mean frequency of MES-Hyp, Neuron, Oligo, and Reactive-Ast spots per spatial organization zone across all GBM samples. Significant adjusted *p* values (<0.05) by the Wilcoxon rank-sum test are shown in the panel.
(C) Spatial maps of representative Visium samples annotated by organizational zones (Struct-Malig, Struct-Norm, Disorg-Malig, and Disorg-Norm) (STAR Methods).
(D) Heatmap showing relative abundance of structured vs. disorganized zones calculated by spatial coherence per CODEX sample. Annotation bar corresponds to MES-Hyp abundance.
(E) Boxplot where each point represents the mean spatial coherence of samples with high abundance (>10%) of MES-Hyp or Neuron vs. samples with low abundance (<10%) of MES-Hyp or Neuron. MES-Hyp or Neuron spots were removed from the calculation of mean spatial coherence. Boxes indicate the median and the 1st and 3rd quartiles. The upper and lower whiskers extend to maximal and minimal values. *p* values calculated by t test (0.0013, MES-Hyp effect; 0.35 Neuron effect).
(F) Boxplot of MP spatial coherence per sample in IDH-mutant samples, boxes as defined in (E). Each dot represents the mean spatial coherence of all spots annotated with a given MP per sample.
(G) Scheme depicting the workflow for GBM organoid hypoxia experiments and spatial profiling (generated with BioRender.com).
(H) Heatmap of the mean relative protein expression (by *Z* score) across all cells per condition. The highlighted proteins represent hypoxia markers (CA9 and GLUT1) or hypoxia-associated markers (NDRG1 and TUBB3) that were upregulated under hypoxic conditions (24 and 48 h).
(I) Stacked bar plot showing the cell state distribution per experimental condition (all organoids per condition are pooled).
(J) Stacked bar plot depicting the fraction of hypoxic cells as defined by hypoxia score (STAR Methods) per experimental condition.
(K) CODEX images of patient-derived GBM organoids under hypoxic conditions for 24 and 48 h with respective controls. Indicated markers representing major cancer cellular states are shown (OLIG2/OPC, DCX/NPC, GFAP/AC, and CA9/MES-Hyp). Scale bar is indicated in the images.
(L) Spatial coherence of cell states per organoid. Annotation bar indicates the experimental condition. Line indicates median.



(legend on next page)

Figure S6. Spatial coherence and organizational zones, related to Figure 4

(A) Compositional differences in structured vs. disorganized regions (across all samples). Vasc (enriched in structured, $p = 0.03$) and Prolif-Metab (enriched in disorganized, $p = 0.003$) were significant by t test.

(B) Boxplot of spatial coherence score per vascular subcluster. Each dot presents the mean spatial coherence score for spots of a given vascular subcluster per sample ($p < 0.001$ by t test). Boxes indicate the median and the 1st and 3rd quartiles; the upper and lower whiskers extend to maximal and minimal values no further than 1.5 times the interquartile range from the 3rd and 1st quartiles.

(C) Organization zone composition of each vascular subcluster (across all GBM samples).

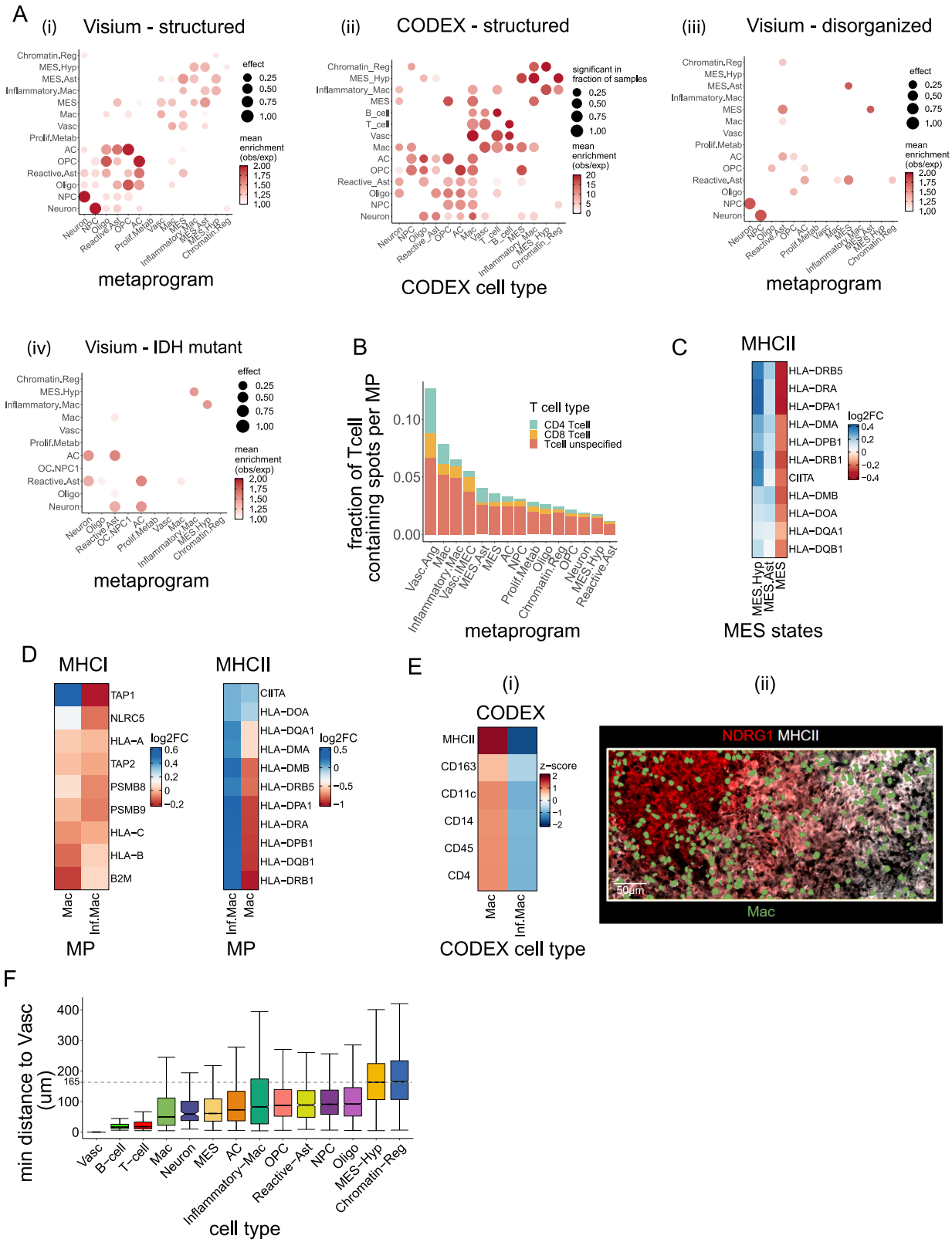
(D) Distribution of vascular subtypes in infiltrating vs. non-infiltrating areas (STAR Methods).

(E) Correlation between relative abundance of structured spots vs. disorganized spots in near-adjacent samples in Visium and CODEX ($R = 0.73$, $p = 0.011$ by Pearson correlation). Each point represents a sample.

(F) Example of a structured and disorganized region by H&E, single-cell spatial map (Voronoi) and CODEX image (corresponding to the image in Figure 4E). Scale bar is indicated in the image.

(G) Average homogeneity score (i.e., confidence score) per MP in disorganized spots vs. structured spots. p values across spots were significant for all MPs (<0.05 by t test).

(H) Scatter plot depicting the percent of spots with adjacent same-identity neighbors per organization zone across homogeneity score (i.e., confidence score) filtering thresholds by quantiles.



(legend on next page)

Figure S7. Spatial relationship analysis, related to Figure 5

(A) Colocalization of MP pairs within a spot for (i) structured and (iii) disorganized regions in Visium, structured regions in CODEX (ii), and IDH-mutant samples (iv). Mean enrichment values represent observed/expected spot colocalization, and dot size represents the proportion of samples for which a metaprogram pair was significantly enriched by Fisher's exact test.

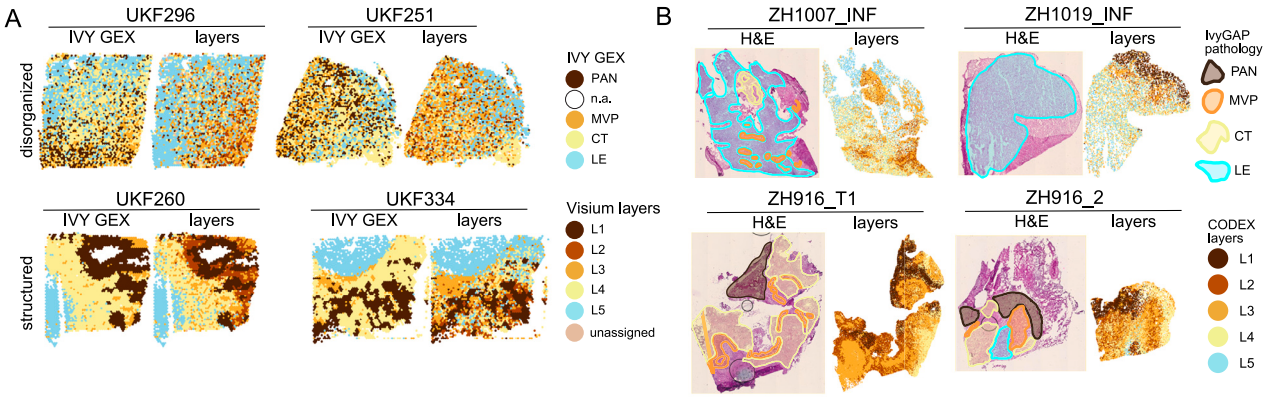
(B) Overall percentage of putative T-cell-containing spots per MP (across all GBM samples) and further categorization of the T-cell-containing spots (CD4, CD8, unspecified). T-cell-containing spots in Visium data were identified by the detection of counts for canonical T cell markers (STAR Methods).

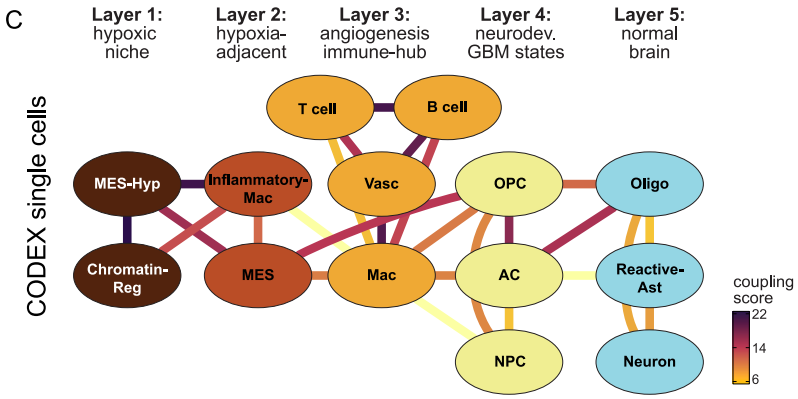
(C) Relative (log_2 fold-change) expression of MHC class II genes across the mesenchymal MPs.

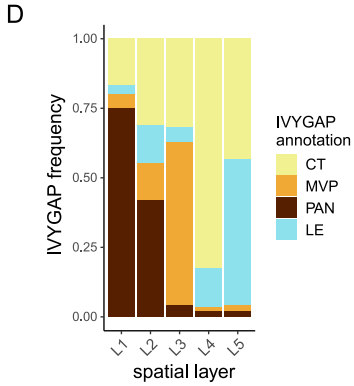
(D) Relative expression of MHC class I (left) and MHC class II (right) genes across myeloid MPs. Expression values represent average expression values across all GBM samples.

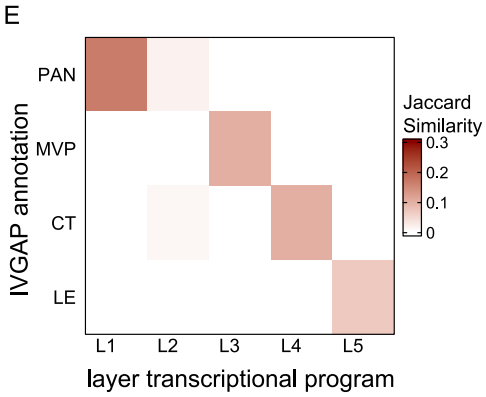
(E) (i) Relative differences in protein expression between Mac and Inflammatory-Mac for the indicated markers (by *Z* score). (ii) CODEX immunostaining showing macrophages with green segmentation masks. Inflammatory-Mac in hypoxic niche ($NDRG1^{high}$) shows downregulation of MHCII compared with Mac in the non-hypoxic area ($NDRG1^{low}$).

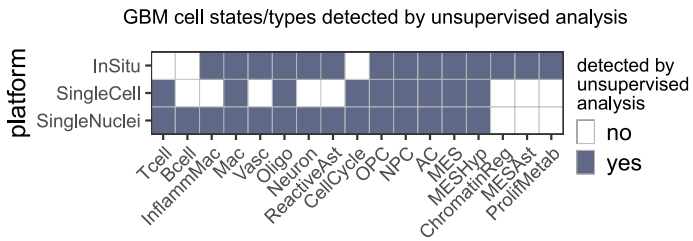
(F) Boxplot showing the distance to the closest Vasc per cell type (in μm). Median distance to MES-Hyp = 165 μm. Boxes indicate the median and the 1^{st} and 3^{rd} quartiles; the upper and lower whiskers extend to maximal and minimal values no further than 1.5 times the interquartile range from the 3^{rd} and 1^{st} quartiles.











(legend on next page)

Figure S8. GBM state layers, related to Figure 6

(A) Spatial maps of representative Visium samples annotated by IVY-Gap transcriptional programs corresponding to histopathological annotations (left) and by state layers (right).

(B) H&E stainings with histopathological annotation by neuropathologist (left) next to CODEX spatial maps annotated by the state layers (right).

(C) Network graph with nodes representing cell types/states and edges representing mean *Z* scores of spatial attraction derived from CODEX colocalization analysis. Only edges representing recurrent interactions with *Z* score >8 and present in >25% of samples are retained.

(D) Composition of state layers by their Ivy Gap histopathological program annotation (Visium data).

(E) Heatmap depicting Jaccard similarity between state layers gene expression programs and gene expression programs derived from histopathological features in the IVY Gap Atlas across all samples. (F) Scheme and heatmap depicting cell states and cell types in GBM discovered by unsupervised analysis using different transcriptomics platforms. snRNA-seq = single-nuclei sequencing, scRNA-seq=single-cell RNA sequencing, *in situ* = 10X Visium spatial transcriptomics. Single-cell cell states/cell types as defined by Neftel et al.[4]; Single-nuclei cell states/cell types as defined by Wang et al.[78] and Al-Dalahmah et al.,[79] *in situ* cell states/cell types as defined in this study.