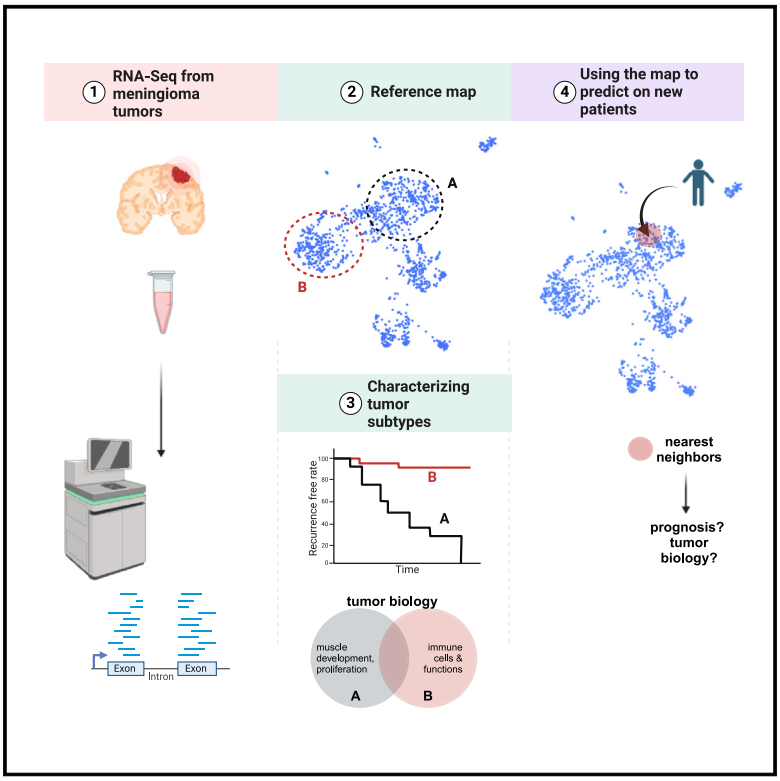
Cell Genomics

Article

Meningioma transcriptomic landscape demonstrates novel subtypes with regional associated biology and patient outcome

Graphical abstract



Highlights

- A meningioma reference map was generated using bulk RNA-seq from ~1,300 tumors
- Clinical and genomic patient data are regionally distributed across the reference map
- RNA-seq identifies tumor subtypes that associate with tumor biology and patient outcome
- The reference map may be used to predict tumor biology and outcome of new patients

Authors

H. Nayanga Thirimanne, Damian Almiron-Bonnin, Nicholas Nuechterlein, ..., Jay Shendure, Manuel Ferreira, Eric C. Holland

Correspondence

eholland@fredhutch.org

In brief

Using RNA-seq from over 1,000 meningioma tumors, Thirimanne et al. generated a reference map that delineates meningioma subtypes with different biologies and patient outcomes. Clinical and genomic patient data are regionalized across the map. The map can be used to overlay new patients and predict tumor biology and outcome.

Thirimanne et al., 2024, Cell Genomics *4*, 100566
June 12, 2024
https://doi.org/10.1016/j.xgen.2024.100566

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Meningioma transcriptomic landscape demonstrates novel subtypes with regional associated biology and patient outcome

H. Nayanga Thirimanne,[1] Damian Almiron-Bonnin,[2] Nicholas Nuechterlein,[3] Sonali Arora,[1] Matt Jensen,[1,4] Carolina A. Parada,[5] Chengxiang Qiu,[6] Frank Szulzewsky,[1] Collin W. English,[7] William C. Chen,[8] Philipp Sievers,[9,10] Farshad Nassiri,[11] Justin Z. Wang,[11] Tiemo J. Klisch,[7] Kenneth D. Aldape,[12] Akash J. Patel,[7] Patrick J. Cimino,[3] Gelareh Zadeh,[11] Felix Sahm,[9,10] David R. Raleigh,[8] Jay Shendure,[6] Manuel Ferreira,[5] and Eric C. Holland[1,4,13,*]

[1]Human Biology Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
[2]Department of Pathology, University of California, San Francisco, San Francisco, CA, USA
[3]Neuropathology Unit, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
[4]Seattle Translational Tumor Research Center, Fred Hutchinson Cancer Center, Seattle, WA, USA
[5]Department of Neurological Surgery, University of Washington Medical Center, Seattle, WA, USA
[6]Department of Genome Sciences, University of Washington, Seattle, WA, USA
[7]Department of Neurosurgery, Baylor College of Medicine, Houston, TX, USA
[8]Departments of Radiation Oncology, Neurological Surgery, and Pathology, University of California, San Francisco, San Francisco, CA, USA
[9]Department of Neuropathology, Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany
[10]Clinical Cooperation Unit Neuropathology, German Consortium for Translational Cancer Research (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[11]Department of Surgery, Division of Neurosurgery, University of Toronto, Toronto, ON, Canada
[12]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA
[13]Lead contact
*Correspondence: eholland@fredhutch.org
https://doi.org/10.1016/j.xgen.2024.100566

SUMMARY

Meningiomas, although mostly benign, can be recurrent and fatal. World Health Organization (WHO) grading of the tumor does not always identify high-risk meningioma, and better characterizations of their aggressive biology are needed. To approach this problem, we combined 13 bulk RNA sequencing (RNA-seq) datasets to create a dimension-reduced reference landscape of 1,298 meningiomas. The clinical and genomic metadata effectively correlated with landscape regions, which led to the identification of meningioma subtypes with specific biological signatures. The time to recurrence also correlated with the map location. Further, we developed an algorithm that maps new patients onto this landscape, where the nearest neighbors predict outcome. This study highlights the utility of combining bulk transcriptomic datasets to visualize the complexity of tumor populations. Further, we provide an interactive tool for understanding the disease and predicting patient outcomes. This resource is accessible via the online tool Oncoscape, where the scientific community can explore the meningioma landscape.

INTRODUCTION

Meningiomas are the most common primary intracranial tumors in humans and usually benign. However, some rapidly recur after multimodal treatment with surgery and radiotherapy, and can ultimately be fatal.[1] The histologic grading of the 2021 World Health Organization (WHO) identifies many of these aggressive tumors as grade 3, but some tumors identified as grades 1 or 2 are equally aggressive.[2] Improved risk classification systems for these tumors are needed, and several grading systems based on DNA methylation, copy number, or expression signatures have been proposed.[1,3–5]

Clues to the underlying biology of these tumors come from neurofibromatosis type 2 with germline loss of one copy of *NF2*, resulting in the formation of multiple meningiomas.[6] Consistent with this observation, DNA molecular analysis shows that the most common alteration of spontaneous meningiomas is loss of chromosome 22 harboring the *NF2* gene.[7,8] The majority of rapidly recurrent meningiomas are among those that show functional loss of *NF2*.[9,10] Additional recurrent genetic alterations in *NF2*-wild-type meningiomas include mutations in genes *TRAF7*, *KLF4*, *AKT1*, and *SMO* and are often restricted to benign meningiomas.[11,12]

NF2 encodes the protein Merlin, which is a tumor suppressor that regulates YAP1 via the Hippo signaling pathway.[7] Upon contact inhibition, the Hippo pathway phosphorylates YAP1, resulting in the inhibition of YAP1 activity.[13] In the absence of Merlin, YAP1 remains active and translocate into the nucleus,

binding the TEAD transcription factors and activating cell proliferation. In addition to chromosome 22 loss, the *NF2* allele can be inactivated due to either inactivating point mutations in the *NF2* sequence or inactivating gene fusions. An alternative route of YAP1 de-regulations can occur due to gene fusions involving the YAP1 gene, resulting in constitutively active YAP1 that is insensitive to Hippo pathway inactivation.[14] Mouse modeling experiments have shown that the expression of either constitutively active YAP1 or YAP1 gene fusions found in human meningiomas induce similar tumors in mice, functionally implicating de-regulated YAP activity in the pathobiology of meningiomas.[15]

Currently available therapeutic options for patients with aggressive meningiomas are limited to radiation and multiple surgeries. It is likely that rapid recurrence and aggressive behavior of some meningiomas reflect the underlying biology of these tumors, which is in turn largely reflected by its overall gene expression pattern. In the hope of understanding this aggressive subset of meningiomas and being able to predict which meningiomas will fall into that category, we performed RNA sequencing of 279 meningiomas from all grades. We then combined our data with multiple publicly available meningioma RNA sequencing (RNA-seq) datasets to generate one of the largest clinically annotated meningioma full RNA-seq datasets available to define the biology of the various meningioma subgroups to create a reference landscape using uniform manifold approximation and projection (UMAP) of 1,298 tumors with associated metadata.

The resulting reference map exhibits multiple clusters of tumors each represented by multiple datasets and indicating multiple RNA-seq-based meningioma subtypes, some of which are associated with distinct time to recurrence. These subgroups are distinguished from each other by gene expression similarities to developmental cell types and biological pathways. We observed several subtypes with particularly poor outcomes; the most common aggressive tumors of these showed high proliferation rates and RNA expression resembling muscle development. We also sought to develop a method to map new patients onto our UMAP landscape and predict tumor behavior and patient outcome based on the nearest neighboring tumors in the map. Oncoscape, an open-source online tool via which this reference map is accessible, provides a great visualization platform for the data shared in this article and allows interactive and analytical exploration of tumors along with various associated metadata. (Oncoscape is accessible on the Chrome search engine via the link https://oncoscape.sttrcancer.org/#project_meningiomaumap94. The main figure panels can be directly accessed using the dropdown menu "Figures from the paper" on the upper-right corner of the website.) We believe that this reference map with demographic and clinical data provides insight to the behavior of the multiple meningioma subtypes, and tools to map new patients onto it will be beneficial in clinical applications to determine patient outcome and therapeutic strategies.

RESULTS

Constructing the meningioma reference UMAP

We obtained 12 publicly available bulk RNA-seq meningioma datasets from nine institutions and five countries in North America, Europe, and Asia and combined them with 279 sequenced meningiomas from the University of Washington to create a set of 1,298 meningiomas (Table S1).[3,4,14,16–25] Raw sequencing data were collected from each dataset and aligned to human genome hg38 using the same pipeline (Figure 1A). To remove batch effects from different datasets, we used R function CombatSeq from the R package sva. Here, we removed batch effects specific to sample collection, processing, and sequencing inherent to each dataset (Figure S1A). Almost all datasets except the one from Yale had a similar distribution of benign and malignant tumors (Figures S1B and S1C). Gene expression values from combined datasets were normalized using the variance stabilizing transformation (VST) method.[26] We applied UMAP, a dimensionality reduction method, on batch-corrected, normalized transcript counts to create a reference UMAP (Figures 1A and 1B). This landscape is made up of multiple distinct clusters of different sizes, all of which are composed of a mix of the 13 datasets (Figure 1B). We were unable to assess the distribution of patients by their race/ethnicity due to lack of information. In addition to UMAP, we explored other dimension-reduction techniques (principal-component analysis [PCA] and t-distributed stochastic neighbor embedding [tSNE]) and found that UMAP better distinguished clusters that showed differences in clinical and genomic features (Figures S1D and S1E).[27] The collection of tumor samples included fresh-frozen (FF) tissue as well as formalin-fixed paraffin-embedded (FFPE) tissue. The FFPE samples distributed evenly across the landscape (Figure S1F). The UMAP landscape facilitates 2D or 3D visualization and is available for interactive analysis and visualization via the open-source, interactive online tool Oncoscape[28] (see Figure 1 in Oncoscape).

Known genetic aberrations are regionally distributed across the UMAP

We colored in the UMAP by DNA sequencing-based mutational metadata associated with a subset of the tumors. For any coloring scheme, we colored in only known values; tumors with no known values for a given data field were left gray. More than half of meningiomas (73% of tumors for which *NF2* status is available) exhibit functional loss of *NF2*, which is achieved via either the loss of chromosome 22, point mutations, or gene fusions (Figure S1G). We colored the map for tumors with known chromosome 22 loss (from collected metadata), which clearly highlighted a large subregion of the map (Figure 1C). Point mutations and gene fusions leading to inactivation of *NF2* also cluster with the tumors with chr22 loss (Figures 1D and 1E). Coloring the UMAP for all three mechanisms of *NF2* inactivation demonstrated a near-complete loss of *NF2* across this region of the map, correlating with overall downregulated expression of *NF2* (Figures 1F and S1E). *NF2* wild-type YAP1 fusion-positive meningiomas also mapped onto the same region, indicating that they resemble *NF2* mutant meningiomas on a gene expression level (Figure 1G).[15,29] Other recurrent non-*NF2* mutations, including *TRAF7* and *SMO*, were found distributed across the *NF2*-wild-type region of the map, while *KLF4* and *AKT1* additionally showed high regionality for recurrent mutations (Figures 1H and S1H). The regionality of these

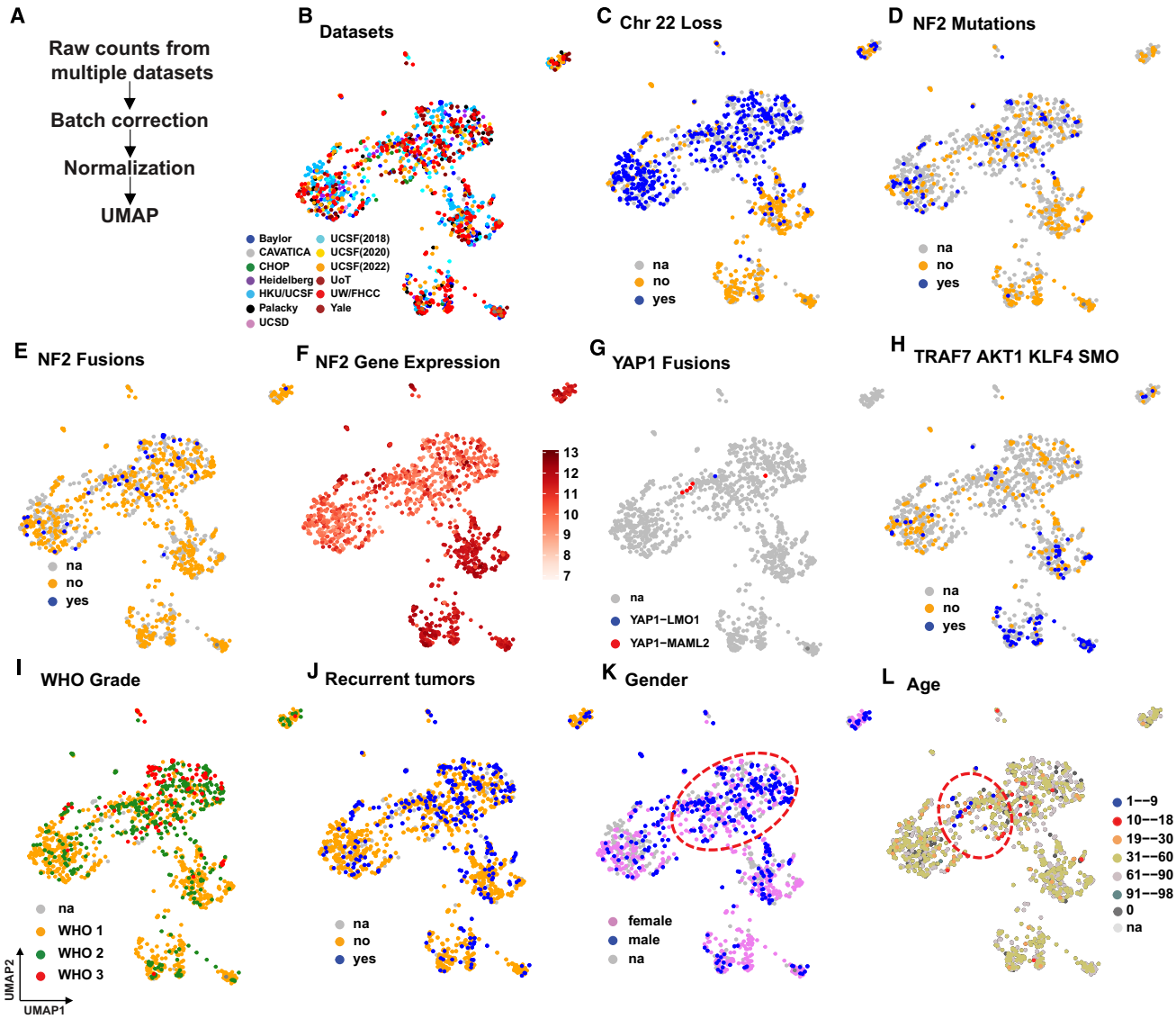


Figure 1. Generating the meningioma reference UMAP and coloring by clinical and genomic metadata

(A) Method overview.

(B) UMAP colored by datasets.

(C) Tumors with (blue) and without (yellow) loss of chromosome 22. (D) Tumors with (blue) and without (yellow) *NF2* mutations.

(E) Tumors with (blue) and without (yellow) *NF2* gene fusions.

(F) *NF2* expression.

(G) YAP1 gene fusions.

(H) Tumors with (blue) and without (yellow) mutations in *TRAF7*/*KLF4*/*AKT1*/*SMO*.

(I) WHO grade: grade 1 (yellow), 2 (green), and 3 (red).

(J) Recurred (blue) and primary (yellow) tumors.

(K) Patients' gender (female pink, male blue). Region 1 marked by the red dashed line.

(L) Age at sample acquisition. Region 2 marked by red dashed line. na, not available. Figure 1 in Oncoscape. See also Figure S1 and Table S1.

genetic alterations is consistent with the known unique biology for the meningiomas. For example, meningiomas that harbor both *TRAF7* and *KLF4* mutations were predominantly regionalized to one cluster (Figure S1I).[30] *TRAF7*-mediated cell transformation is enhanced by loss of *KLF4* in a subset of meningiomas.[31]

Aggressive tumors are regionally concentrated

Most of the samples in our dataset have a WHO grade associated with them, and coloring in the UMAP by that grade shows nonrandom distribution. A subset of the region characterized by *NF2* loss had an increased concentration of WHO 2 and 3 tumors relative to the rest of the map (Figure 1I). However, the

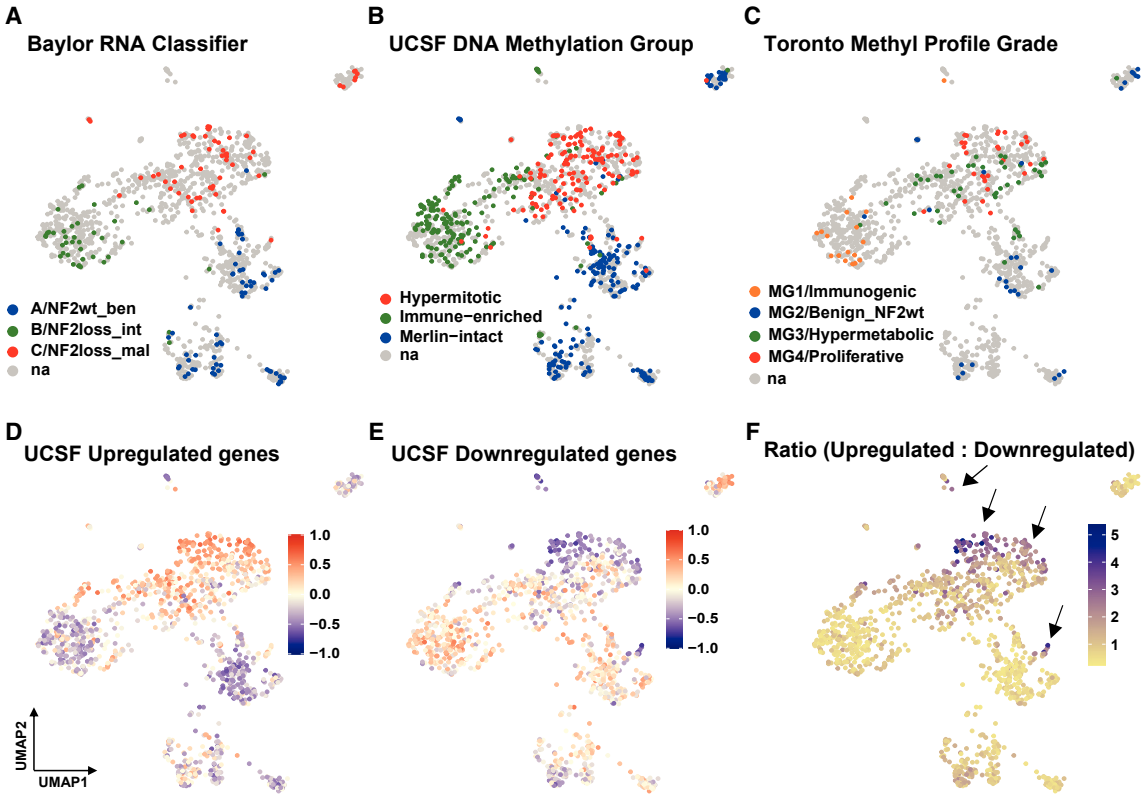


Figure 2. Various grading systems show regional patterns across the UMAP

(A) UMAP colored by the Baylor RNA classification: A/NF2wt_ben = NF2 wild-type benign (blue), B/NF2loss_int = NF2 lost intermediate (green), C/NF2loss_mal = NF2 lost malignant (red).

(B) UMAP colored in by UCSF DNA methylation-based classification of a subset of tumors: hypermitotic, red; immune-enriched, green; Merlin-intact, blue.

(C) UMAP colored in by the Toronto methylation profile of a subset of tumors: MG1/immunogenic, orange; MG2/benign_NF2wt, (benign NF2 wild type) blue; MG3/hypermetabolic, green; MG4/proliferative, red.

(D and E) UMAP colored by GSVA scores calculated using UCSF gene set (D) upregulated and (E) downregulated in most aggressive meningioma. 1 suggests upregulation, −1 suggests downregulation of the respective gene set.

(F) Ratio of GSVA scores from upregulated and downregulated gene sets. Black arrows indicate the regions with the most aggressive tumors marked by the ratio. Figure 2 in Oncoscape. See also Figure S2.

region with the highest concentration of WHO 2 and 3 tumors still contained tumors of all three grades. Tumors known to be recurrent at the time of resection are generally also concentrated in the same region as tumors with a higher grade (Figure 1J). For some of the patients, there were data regarding the time between the surgery generating the sample and the next or previous resection. Coloring in the map with these data also showed that a short time to recurrence was enriched in that same area, as was higher average grade and increased likelihood of being a recurrent tumor (Figure S1J).

Regional age and gender distribution

Nearly all the samples had records of age and gender. Consistent with what is known, the majority of the map comprised older patients that were predominantly female (median age of 58 years and 66% female) (Figure 1K and 1L).[32,33] By contrast, there were two regions of the UMAP that varied from this general rule. One region with the most aggressive tumors was largely male (61% male vs. 31% male in the rest of the UMAP), and a second adjacent region comprised a higher percentage of younger patients (<30 years) than the general populations of meningioma patients (22% vs. 5% in rest of the UMAP) (Figures S1K–S1M).

Various grading systems are consistent with the regional patterns across the UMAP

Because the WHO grading system does not identify all the meningiomas with aggressive behavior, several recent alternative grading systems have been proposed that use methylation patterns, copy-number alterations, and gene expression to place patients into specific groups associated with a time to recurrence.[3,4,21,34,35] We colored in this UMAP by metadata of these grading systems, and all of them correlate with UMAP subregions (Figures 2A–2C and S2A) (see Figure 2 in Oncoscape). Additionally, the expression of the 34 genes presented by the Raleigh lab as a signature to predict meningioma outcome was analyzed in correlation to our UMAP.[36] Upregulated and downregulated genes in the most aggressive meningiomas were divided into two gene sets, and the whole dataset was subjected

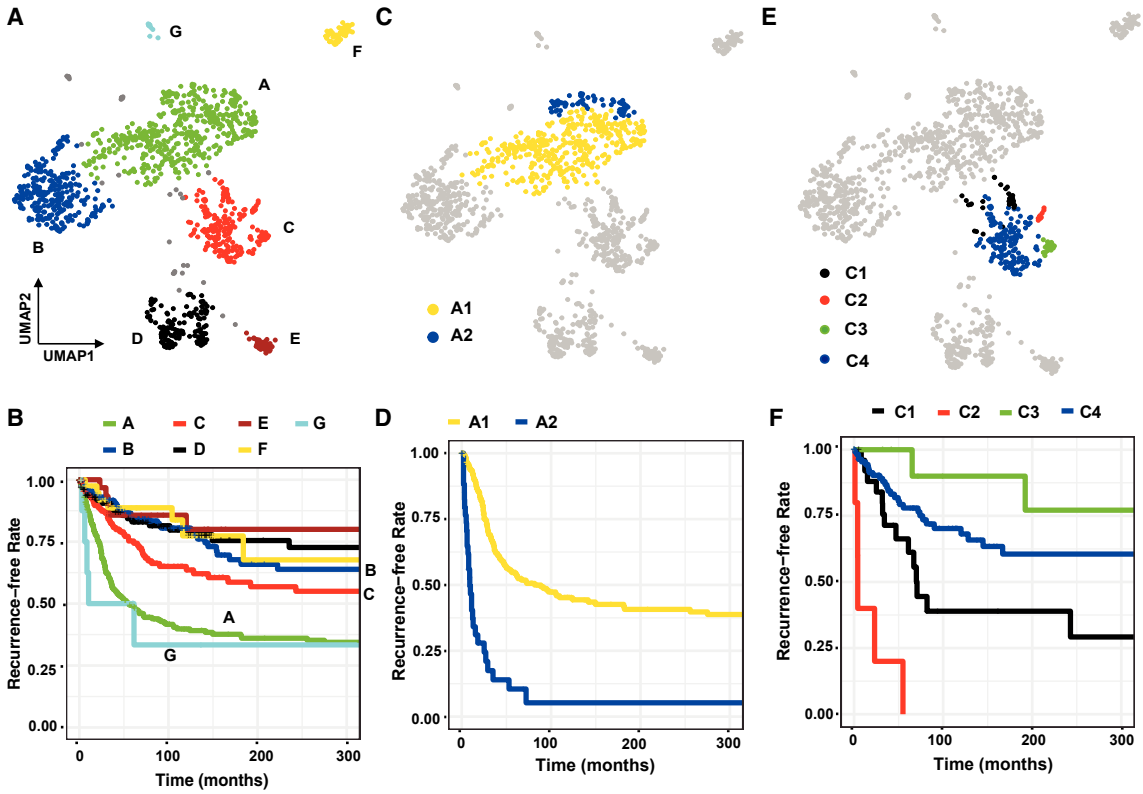


Figure 3. Meningioma subtypes with distinct time to recurrence

(A) Seven regions identified by DBSCAN denoting meningioma subtypes A–G. Unclustered samples (n = 37) are shown in gray.

(B) Kaplan-Meier plots for the seven regions based on time to recurrence (AvsB and AvsC $p < 0.0001$; CvsD and BvsC $p < 0.02$); n(A) = 317, n(B) = 172, n(C) = 131, n(D) = 101.

(C) Subclusters of cluster A (A1 and A2).

(D) Kaplan-Meier plots showing recurrence-free rates of cluster A subclusters (A1 vs. A2: $p < 0.0001$); n(A1) = 230, n(A2) = 36.

(E) Subclusters of cluster C (C1, C2, C3, and C4).

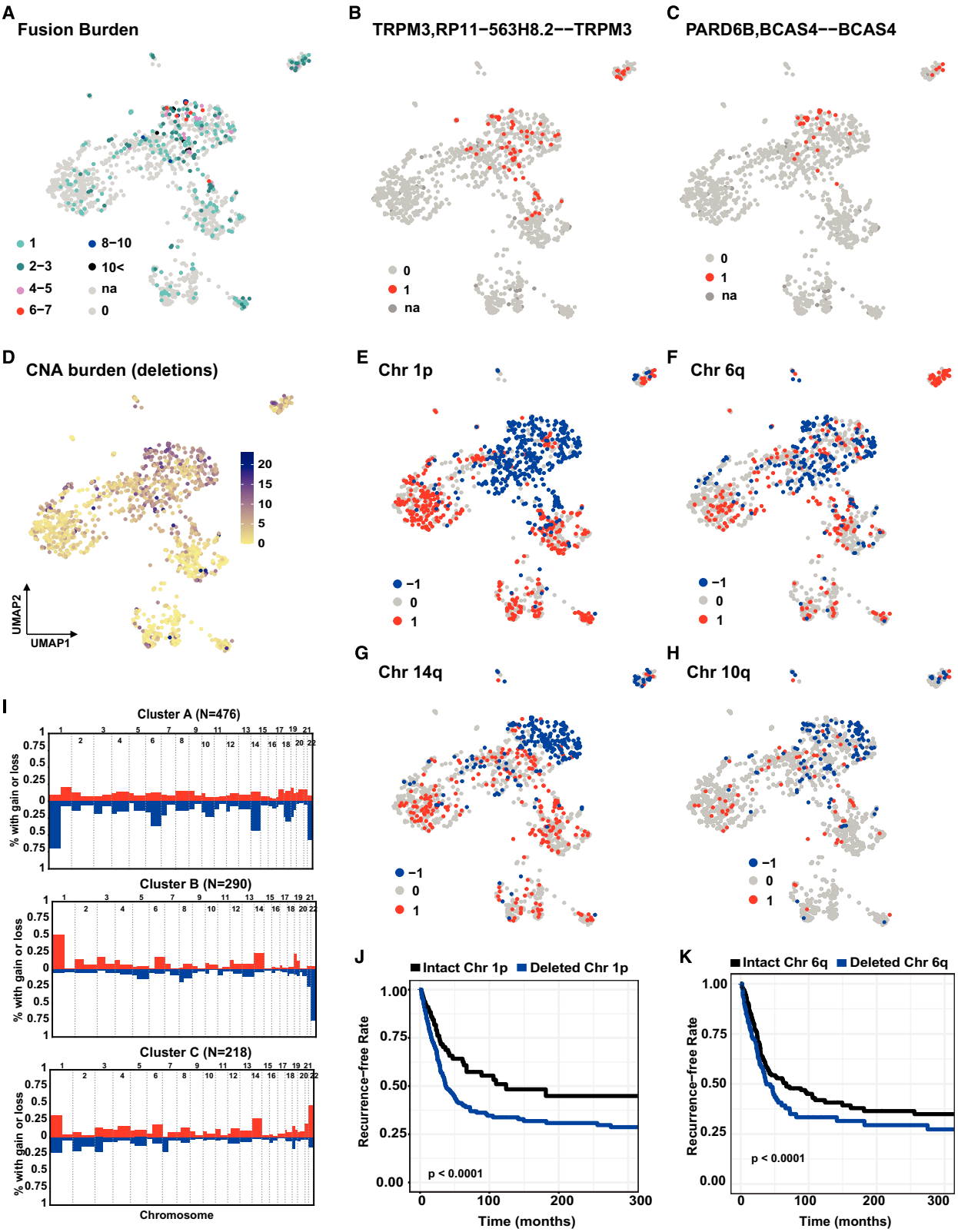
(F) Kaplan-Meier plots showing recurrence-free rates of cluster C subclusters (C1 vs. C2, $p < 0.0001$; C3 vs. C4, $p = 0.26$; C1 vs. C3, $p = 0.01$); n(C1) = 26, n(C2) = 5, n(C3) = 13, n(C4) = 100. Figure 3 in Oncoscape. See also Figure S3.

to gene set variation analysis (GSVA) using the two gene sets separately. The two sets of GSVA scores and then a ratio of them were used to color in the UMAP, and they highlighted the most aggressive region of our map in line with the University of California San Francisco (UCSF) findings (Figures 2D–2F). Furthermore, we generated a UMAP using only those 34 genes (Figure S2B) and found that basic regionalization of aggressive and recurrent tumors was in line with the UMAP generated using all protein-coding genes; however, meningioma subtypes were better segregated in the latter UMAP. The one classification system for meningiomas that did not correlate with map location was the Simpson grading scale, which is a measure of tumor resection completeness, suggesting that the ability to completely resect a tumor is not determined by expression pattern and underlying biology of the tumor (Figure S2C).[37]

Meningioma subtypes with distinct time to recurrence

Genomic and clinical metadata integrated into the UMAP revealed regionalization, suggesting potential meningioma subtypes. In order to define map regions with statistical confidence, we employed three methods: density-based spatial clustering of applications with noise (DBSCAN), k-means, and hierarchical clustering (Figures 3A, S3A, and S3B). There was overall overlap among clusters identified by the three methods. We chose DBSCAN due to its ability to delineate regional distinct clusters that corresponded well with metadata, identifying seven general clusters labeled A through G (Figure 3A) (see Figure 3 in Oncoscape). Notably, both k-means and hierarchical clustering outputs corroborated the UMAP-based intra-cluster subdivisions, as depicted below.

The region of the UMAP with functional loss of *NF2* (Figure S1G) comprised clusters A and B. Cluster A contained the highest density of aggressive tumors, and cluster B represented the remainder of the *NF2* loss region of the UMAP with relatively benign tumors. Clusters C and D were composed of mostly *NF2* wild-type tumors. The comparison of Kaplan-Meier plots of time to recurrence for these main clusters identified clusters A and G as the clusters with the shortest time to recurrence (Figure 3B). Patients in cluster C also perform significantly worse than those in clusters B, D, E, and F, all of which were similar.



(legend on next page)

Intra-cluster time-to-recurrence analysis suggests further subdivision of the map

The regional differences of time to recurrence within a cluster was seen for many of the clusters. The most striking was in clusters A and C. We identified two major subclusters within cluster A based on the regionalization of the most aggressive tumors and, most importantly, differences in patient outcome (Figures 3C and 3D). A2 harbored the largest population of patients with poor outcomes and was also identified by the 34-gene signature (Figure 2F). Cluster C can be similarly divided into four subclusters with significant differences in outcome (Figure 3E and 3F). Subcluster C2 is a small subcluster of eight tumors derived from six different datasets having a uniform short time to recurrence. Several of the other clusters can also be subdivided into regions with significant differences in outcome (Figure S3C).

Gene fusion calling from RNA-seq data shows high prevalence in aggressive regions

We identified gene fusions from the RNA-seq data using Arriba.[38] At high confidence, we were able to identify 171 gene fusions that have at least one coding gene partner and that recur at least twice within the dataset (Table S2). The regions of the map with highly aggressive tumors (cluster A and parts of cluster C) showed significantly high fusion burden (Figure 4A) (see Figure 4 in Oncoscape). Some of the tumors harbored multiple fusions, and some highly recurrent protein-coding gene fusions were found regionalized on the map (Figures 4B and 4C). For example, TRPM3,RP11-563H8.2-TRPM3 fusions enriched within cluster A and some parts of cluster C and F. In another example, PARD6B-BCAS4 fusions were enriched in cluster A and most concentrated within the subcluster A2. Using RNA-seq, we identified *NF2* fusions additional to those known from collected metadata, and they were predominant in cluster A (Figure S4A). It is worth noting that YAP1 fusions are mostly in pediatric patients (Figures 1G and 1L). YAP1-MAML2 is identified as a causal oncogenic driver in pediatric *NF2* wild-type meningiomas.[14] Furthermore, YAP1-MAML2, which leads to constitutive activation of YAP1, has been shown to be sufficient to induce meningiomas in mice.[15] The pediatric tumors achieve YAP1 activation by different mechanisms with fewer losses of chromosome 22 and more gene fusions such as those that activate YAP1.

Regionally enriched chromosome arm-level copy-number alterations

We estimated arm-length gains and losses of chromosomes using CaSpER on RNA-seq data (Table S3).[39] These were validated with the known copy-number alteration (CNA) data on 304 samples where DNA sequencing was available for chromosome 22 status, and 90% of 22q losses identified by CaSpER were confirmed by available metadata. Additionally, chromosome 22 losses correlate well with the expression of the *NF2* gene, which is harbored on chromosome 22 (Figure S4B). The CNA burden was highest in cluster A where the most aggressive tumors are located (Figure 4D). We observed specific gains and losses of chromosome arms regionalized on the UMAP in a cluster-specific manner (Figures 4E–4H and S4C). Consistent with published data, the aggressive region of the UMAP (cluster A) shows loss of 1p, 6q, 10q, and 14q (Figures 4E–4I).[24] Patients with loss of either 1p, 6q, 10q, or 14q were all associated with shorter time to recurrence than the patients without those CNAs (Figure 4J, 4K, and S4D–S4F). Chromosome 1p loss and 1q gain were specific to cluster A, while gain of 1p was seen frequently in the rest of the UMAP (Figures 4E–4I and S4G). We show loss of 10q, which harbors *PTEN*, and low expression of *PTEN* in the most aggressive region (subcluster A2) (Figures 4H and S4H). Other groups have demonstrated *PTEN* mutations in aggressive tumors.[3] We also observed loss of 9p, 18p, and 18q and gain of 17q enriched in the most aggressive regions (Figures 4I, S5, S6, and S7).

Meningioma subtypes frequently show expression patterns of developmental cell types

The meningioma reference landscape was generated using RNA-seq data and therefore presents an advantage of performing differential gene expression analysis and deciphering the underlying biology across meningioma subtypes. We first determined the differentially expressed genes in each cluster relative to the rest of the meningiomas. We then performed Gene Ontology (GO) analysis, and the most prevalent GO terms in each cluster were used to discern the underlying biological signature for each of them (Figures 5A–5C and S8A; Tables S4 and S5).[40] Cluster A was enriched for cell cycle, skeletal and cardiac muscle development, and DNA replication and repair, while cluster B was specific to immune cells and function (Figures 5A and 5B) (see Figure 5 in Oncoscape). Although cluster C had fewer GO terms that did not point toward a specific biological signature, *SMO* mutations were enriched in cluster C (Figure S8B). Accordingly, regulation of Smoothened signaling and SHH (Sonic hedgehog) pathway were upregulated within cluster C (Figures S8A and S8C). Similarly, cluster D had a broader collection of GO terms; however, these were enriched for *AKT1* mutations (Figures S8A and S1F). Clusters E and F enriched for epidermis development and vascular development, respectively (Figures 5C, and S8A). It is worth noting that KLF4, a transcription factor involved in skin development, was highly mutated in cluster E tumors—specifically, the K409Q mutation (Figure S1F).[41] Cluster G, which had one of the worst

Figure 4. Regionally enriched gene fusions and copy-number alterations

(A) Fusion burden in each tumor derived from high-confidence gene fusions called using RNA-seq.

(B and C) Examples for regionalized fusions.

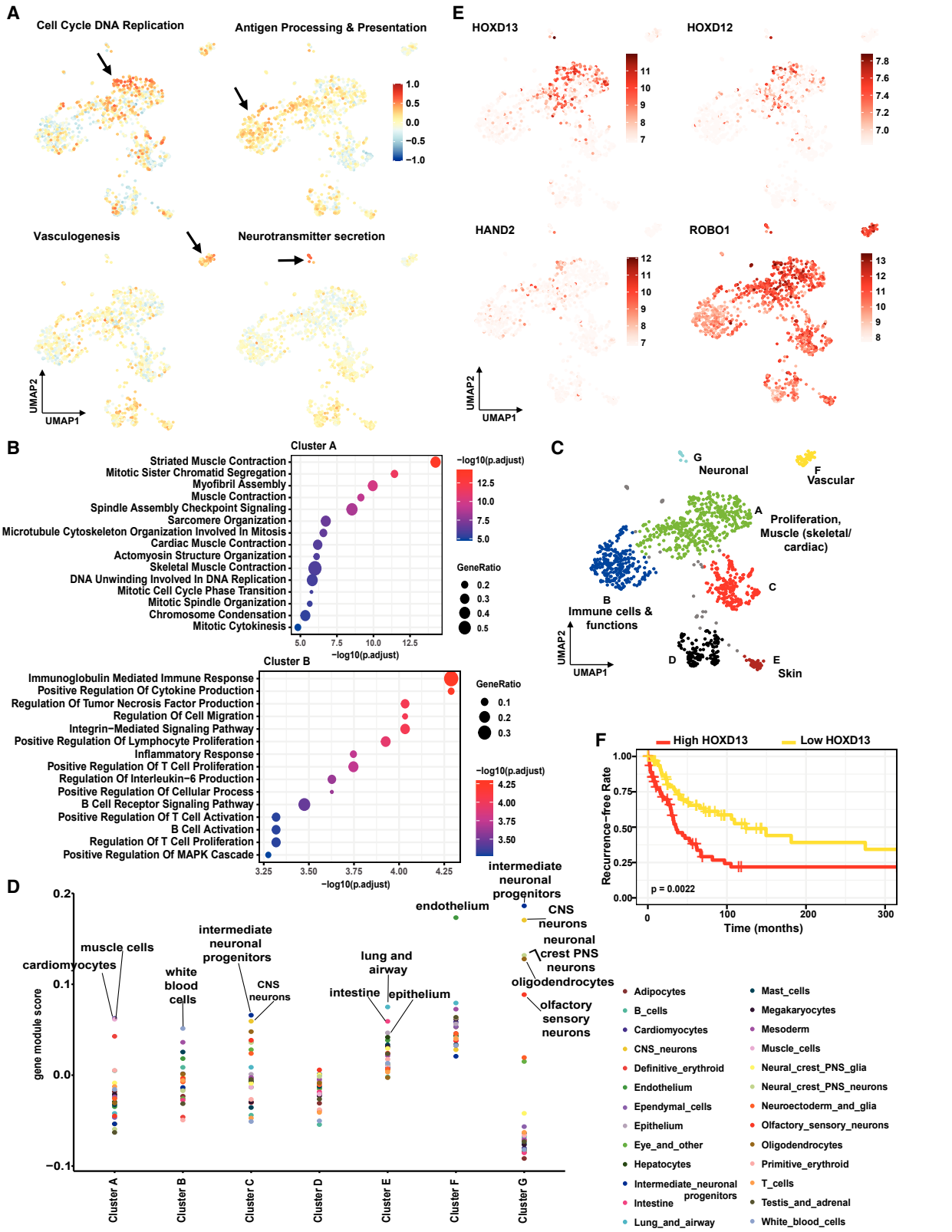
(D) Burden of CNA in each tumor (loss of chromosome arms).

(E–H) Loss (−1), gain (1), or intact (0) status of (E) chromosome 1p, (F) chromosome 6q, (G) chromosome 14q, and (H) chromosome 10q in each tumor.

(I) Manhattan plots showing losses (blue) and gains (red) of each chromosome arm in clusters A, B and C.

(J) Kaplan-Meier plot showing the recurrence-free rate of patients in cluster A with intact and deleted chr 1p; $p < 0.0001$, n(del) = 194, n(intact) = 86.

(K) chr 6q, $p < 0.0001$, n(del) = 96, n(intact) = 184. Figure 4 in Oncoscape. See also Figures S4–S7 and Tables S2 and S3.



(legend on next page)

outcomes, was enriched for neuronal functions including neurotransmitter/synaptic transmission and nervous system development.

Some of these clusters were notably enriched with developmental GO terms, indicating potential parallels between the biology of these tumors and embryonic development. Therefore, to further learn about the potential association of meningioma subtypes with developmental cell types, we compared the identified cluster-specific gene signatures to mouse embryonic cell types. We leveraged the transcription profiles of a series of mouse embryonic developmental stages and hundreds of cell types put together by the Shendure lab.[42] In line with what GO terms suggested, we found that cluster A was enriched for muscle progenitor cells and cardiomyocytes, while cluster B was enriched for immune cells (Figures 5D, S9, and S10; Table S6). Cluster C was enriched for neuronal cells; however, cluster D was not enriched for any specific embryonic cell type (Figures 5D, S9, and S10; Table S6). Consistent with GO terms, epithelial cells and endothelial cells were top hits in clusters E (skin related) and F (vascular related), respectively (Figures 5D, S9, and S10; Table S6). Similar to what GO terms suggested, cluster G was enriched for various neuronal cells (Figures 5D, S9, and S10; Table S6).

Cluster A, where the GO terms suggested muscle development had one of the worst outcomes, exhibited enrichment of *HOXD12/13*, *HAND2*, and *ROBO1* genes (Figure 5E). Relative enrichment of expression of these genes suggests the possibility that the underlying biology of this cluster may resemble embryonic limb development.[43] Alternatively, biology of this cluster could be related to the entirety of embryonic state in which the limb is developing rather than limb development specifically. High-*HOXD13*-expressing tumors in cluster A had a significantly shorter time to recurrence than the low-*HOXD13*-expressing tumors of that same cluster, which may further suggest that limb-related development correlates with the most common aggressive meningiomas (Figure 5F).

Recurrent tumors remain largely in the cluster from which they arise

There were several cases where samples were resected from multiple tumors from the same patient. We identified three scenarios: recurred tumors, multiple individual tumors from different brain regions, and progressed tumors due to incomplete surgical resection. We evaluated their location on the UMAP to further understand how their biology and outcome might differ with time. Recurrent tumors remained within the clusters in which they were found originally, and vectors (between two tumors of the same patient) do not point toward a more aggressive region of the map, suggesting that the recurred tumors' biology and outcome do not vastly differ from the initial tumor (Figure 6A; Table S7A). In four different cases, multiple tumors had occurred in different brain regions, and most of them were patients with *NF2* loss (Figures 6B; Table S7B). While some of the tumors presented similar biology and outcome, some were vastly different from each other. Two cases where the tumors progressed due to previous incomplete surgical resection mapped within the same cluster (Figures 6C; Table S7C).

Overlaying new patients onto an existing reference UMAP

The above data have shown that the biology and outcomes of meningiomas are regionally located in our UMAP reference landscape. Therefore, the nearest neighbors on the UMAP to a given tumor can serve as references from which a tumor's biology and likely outcome can be inferred. However, for us to make such inferences, we must be able to reliably map a new patient onto our reference UMAP. To this end, we developed and validated an algorithm that accurately places new patients on our map.

Our placement method uses a weighted, nearest-neighbors approach that leverages an ensemble of UMAP models. We pre-trained 100 UMAP models with different initializations on our reference dataset (Figure 7A) and used each model to map a new patient to a distinct two-dimensional embedding (Figure 7B). We then used the location of the new patient in each embedding to determine which reference samples are the 100 nearest neighbors of the new patient in each embedding within a radius determined using cross-validation (Figure 7C). This results in 100 sets of nearest neighbors from the reference dataset. This information is used to determine how frequently each reference sample in our reference dataset is a nearest neighbor of the new patient (Figure S11A). Finally, using this frequency information and the coordinates of the reference samples on our reference UMAP, we computed the centroid of the coordinates of the reference samples on our reference UMAP weighted by the frequency with which these samples were nearest neighbors of the new patient (Figures 7D and 7E). We used this centroid as the final placement location of a new patient on our reference UMAP.

To establish the reliability of our placement algorithm, we used leave-one-out cross-validation to assess how far samples in the reference dataset moved when they were removed from the dataset and mapped back onto the reference UMAP. First, we considered each reference sample's location in the reference UMAP as ground truth (Figure 7F). Next, we iteratively removed each sample, retrained our UMAP models without that sample, and used our placement method to place each sample back

Figure 5. Biological significance of meningioma subtypes

(A) Visualization of GSVA scores across the UMAP for selected Gene Ontology Biological Processes (GO BP) terms. 1 suggests upregulation, −1 suggests downregulation of the respective gene set.

(B) Top 15 GO BP terms enriched in clusters A and B.

(C) Summary of biological significance of each cluster.

(D) Mouse embryonic cell types enriched in each cluster (top hits). Cell type similarities are as listed. Welch's two sample t test; $p < 2.2e{-}6$.

(E) Expression profiles for genes known to be involved in embryonic limb development.

(F) Kaplan-Meier plots showing correlation between recurrence-free rate and HOXD13 levels; $p = 0.0022$, n(high) = 66, n(low) = 69. Figure 5 in Oncoscape. See also Figure S8–S10 and Tables S4, S5, and S6.

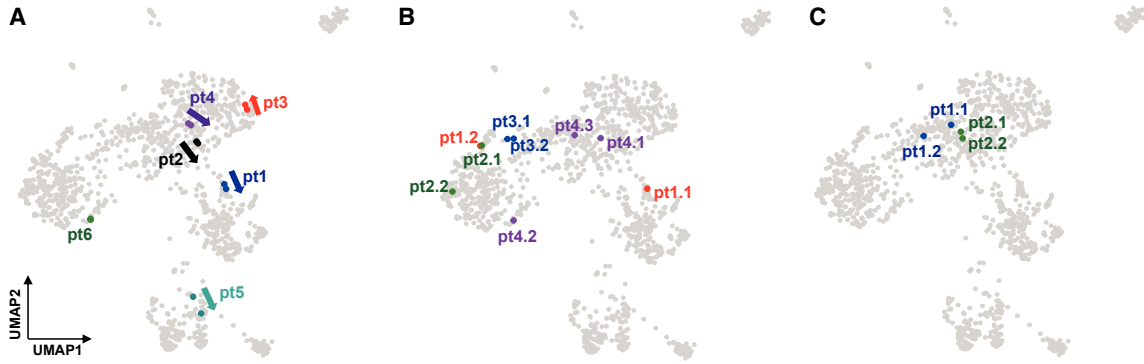


Figure 6. Evolution of multiple tumors from the same patient

(A) Primary and recurred tumors from the same patient. Arrows show the direction from the first tumor to the second tumor of a specific patient. Tumors from a single patient are distinguished by the colors. pt, patient.

(B) Multiple individual tumors occurred within the same patient. Each patient distinguished by different colors (e.g., pt1.1 = patient 1 tumor 1, pt1.2 = patient 1 tumor 2).

(C) Primary and progressed tumors (e.g., pt1.1 = patient 1 tumor 1, pt1.2 patient 1 tumor 2 [progressed]). See also Table S7.

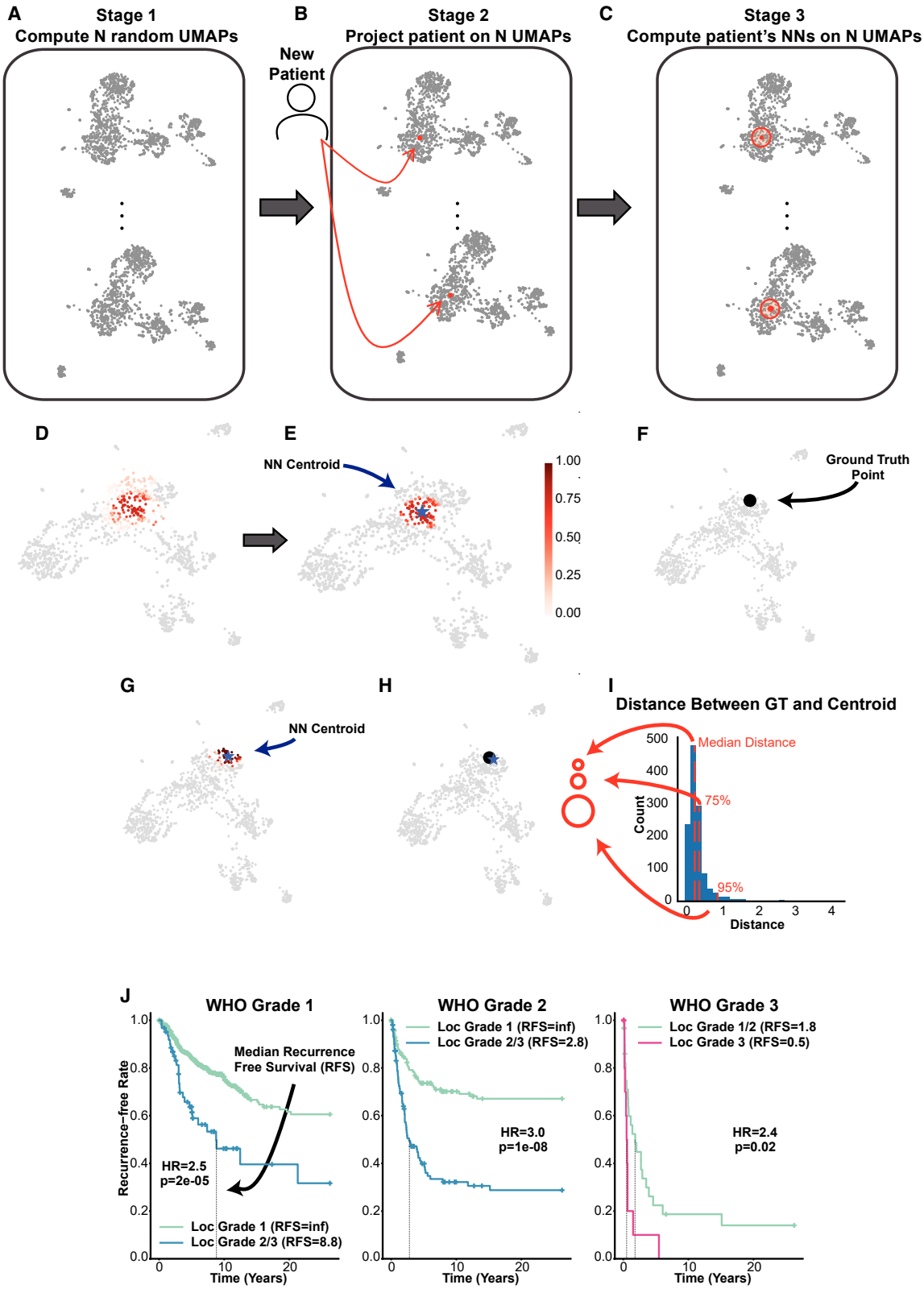
onto the reference map (Figure 7G). Last, we computed the Euclidean distance between each reference sample's ground-truth position and its predicted location (Figure 7H). Nearly all reference points were mapped within a small radius of their true location (Figure 7I). The reliability of our placement method was also confirmed by evaluating its predictive power. Cross-validated results showed that our method was able to predict patient cluster membership accurately (AUC = 0.98) by simply predicting the cluster most common in the samples around which the patient was placed (Figures S11B–S11D).

Cross-validated results also demonstrated the potential prognostic utility of our reference landscape. To leverage a patient's location on our reference map, we assigned a location-based tumor grade to each sample in our reference dataset that corresponded to the WHO grade most common in the sample's surrounding samples once remapped onto our reference map. Our results indicated that our predicted location grade was a superior risk indicator compared to WHO grade within WHO grade 1 and WHO grade 2 meningiomas and, to a lesser extent, WHO grade 3 tumors. In univariate analyses, WHO grade 1 tumors were separated into predicted location grade 1 and predicted location grade 2/3 tumors with dramatically different recurrence-free survival (hazard ratio [HR] = 2.5, p = 2e−05, log rank); similarly, WHO grade 2 tumors classified as location grade 1 had significantly better recurrence-free survival compared to tumors classified as location grade 2 or 3 (HR = 3.0, p = 1e−08, log rank) (Figures 7J and S11E). Additionally, among WHO grade 1 and 2 tumors, a multivariate analysis shows our predicted location grade is an independent predictor of recurrence-free survival compared to WHO grade in our reference dataset (p = 2e−13, Cox proportional hazard [CPH]) (Figure S11F). However, although WHO grade 3 tumors classified as location grades 1 or 2 had more favorable outcomes than those predicted to be location grade 3 (HR = 2.4, p = 0.02, log rank), all WHO grade 3 tumors experience short times to recurrence. Thus, despite the prognostic power of our UMAP landscape, histopathology plays a crucial role in assessing patient risk. We do not propose this location-based tumor grade as an alternative to current classification systems; rather, we present these predictions to highlight the predictive power of the reference landscape. The ability to place new patients on this landscape makes our findings relevant for clinical applications. Furthermore, we found that, despite notoriously poor sequencing quality, FFPE samples map onto the UMAP as well as the more common FF samples (Figure S11G).

In addition to leave-one-out cross-validation, we tested the reliability of our placement method by randomly removing 10% of samples from each data source in our dataset to form a separate test dataset (N = 128). We retrained 100 UMAP models on the remaining 90% of these data (N = 1,170) and used these models to individually place each test sample on a fixed UMAP reference map generated from these 1,170 training samples (Figure S12A). To quantify the performance of this assessment, we first used the nearest-neighbors approach described above to assign a location-based tumor grade to each test sample. As we saw during cross-validation, our predictions dramatically stratified survival in WHO grade 1 (HR = 4.2, p = 0.02, log rank) and WHO grade 2 (HR = 11.4, p = 0.004, log rank) samples (Figure S12B). To assess the accuracy of the placements of samples in the test set with a distance metric, we approximated the ground-truth location of the samples in the test set on the 1,170-sample reference UMAP. For each test sample, we considered its approximate ground-truth location on the 1,170-sample reference UMAP as the centroid of the samples that were the nearest neighbors of that test sample on the full 1,298-sample reference UMAP. We observed no significant difference between cross-validation error (difference between placement and ground truth on 1,298-sample UMAP) and test set error (difference between placement and approximated ground truth on 1,170-sample UMAP), an indication of similar accuracy (Figure S12C).

DISCUSSION

Most studies thus far have categorized meningiomas as malignant or benign. There are several classification systems currently



(legend on next page)

available that are based on histomorphology and a limited number of genetic alterations. Our goal was to determine the complexity of the meningioma population and decipher the underlying biology of these tumors using a comprehensive transcriptomic-based approach. We identified a complex landscape of multiple meningioma subtypes comprising seven general map regions and determined that there is more than one aggressive type (clusters A, C, and G). The reference UMAP regionalized meningioma subtypes based on *NF2* status and underlying biology.[3,4,21] More importantly, our study further extends the understanding of meningioma subtypes by adding a granular classification such as subtypes related to muscle, skin, vascular and neuronal development, and separate clusters within the *NF2* wild-type region (clusters C and D, which enrich for *SMO* and *AKT1* mutations, respectively). In addition to GO term analysis, we compared clusters against different cell types during mouse embryonic development, where we show that the biology of clusters depicts both pathways and developmental cell types. Our findings were in line with methylation-based classifications presented in previous studies because those classification cohorts mapped regionally on the UMAP. Tumor clusters identified by the UCSF group as hypermitotic, immunogenic, and Merlin-intact overlapped with cluster A; cluster B; and clusters C, D, E, and F, which enriched for proliferation, immune cells and function, and *NF2* wild-type tumors, respectively (Figure 2B).[4,21] Classification done by University of Toronto (UoT) was also congruent with our analysis (Figure 2C).[3] Furthermore, we show that the 34-gene signature presented by the UCSF group identifies UMAP regions with the most aggressive tumors (Figures 2E and 2F).

Within main clusters, we identified subregions that perform differently in terms of patient outcome. These subregions were identified based on visual inspection of how the most aggressive tumors group together and how they behave in the 3D format of the UMAP. Subregions were obvious, with clear differences in patient outcome. They were ascertained by the UMAP colored in by UCSF gene signature (Figure 2F). Moreover, we observed CNA patterns that highlighted subregions within clusters (for example, chr 10p and 10q in subcluster A2 and chr 5q, 19p, and 19q in subclusters B1/B2). Identification of such subregions underscores the importance of predicting the outcome of a new patient based on nearest neighbors within a subregion instead of overseeing it as a whole cluster. The addition of more samples to the map will likely allow delineation of these subregions with higher accuracy.

Our dataset contained 33 pediatric (<18 years old) and 49 young adults (19–30 years old). The majority of these young patients mapped onto the region between cluster A and B. Consistent with other tumor types, pediatric patients and young adults harbored more gene fusions as opposed to CNAs (data not shown). YAP1-MAML2 fusion was more prominent among younger patients, while NF2 fusions were enriched among adults. Further investigation is warranted comparing fusions in adults and young patients with a larger cohort of the latter. We observed a large fraction of fusions with non-coding genes. Although we limited our current study to fusions with protein-coding genes, it would be worth deciphering the role of non-coding fusions in meningiomas in future studies.

We developed a method to overlay prospective patients on our reference UMAP to infer their biology and clinical outcomes from their nearest neighbors. We used cross-validation to verify that patients were accurately placed on the reference UMAP both by measuring the distance our method places reference samples from their location in the reference UMAP and by assessing how well our method predicts cluster membership. In addition to cluster membership, we also used our method to assign patients a location grade based on the distribution of the WHO grade of their surrounding samples. Some tumors classified as WHO grade 1 were located in the most aggressive region of our UMAP reference landscape, which suggests that, despite the histopathological grading, their underlying biology and outcome are similar to more malignant tumors. Therefore, our transcriptomic-based UMAP landscape of meningioma may provide a better understanding of a patient's biology and outcome. If clinical and treatment data are available for the patients used to build the reference landscape, the nearest-neighbor analysis could also be used to identify what therapeutic approaches were successful in the patients with tumors most similar to a new patient placed on the reference landscape. In the case of meningiomas specifically, there are no targeted therapies that work in a subset of these tumors. However, in other tumor types, this kind of analysis could help distinguish those likely to respond based on the underlying biology of the disease type.

Although the majority of our data were derived from FF samples, we found that 114 FFPE samples blended into the map (Figure S1F). FFPE samples were mapped on to UMAP with a higher precision than the frozen samples (Figure S11G). Due to the low number of metadata associated with these FFPE samples, we were unable to reach statistical significance determining mapping accuracy comparing metadata of FFPE samples to their

Figure 7. Overlaying new patients on to the reference UMAP

(A) Two of 100 UMAP embeddings produced by 100 pre-trained UMAP models trained with different random states.

(B) New patient VST data are mapped onto all 100 UMAP embeddings using the pre-trained UMAP models.

(C) For each UMAP embedding, the nearest 100 neighbors are chosen subject to a radius R determined by cross-validation.

(D) Example plot of the reference UMAP with samples colored by the frequency of each reference sample in our reference dataset being a nearest neighbor of a new patient.

(E) Illustration of the placement of a new patient at the centroid of the nearest neighbors weighted by the frequency vector in (D) after outlier exclusion.

(F) The ground-truth location of a reference sample during cross-validation.

(G) The placement of a reference sample using our placement method during cross-validation.

(H) Comparison of the ground-truth placement and the centroid it is mapped to during cross-validation.

(I) The distribution of the distances between the ground-truth placement of a reference sample and its centroid placement for all reference samples during cross-validation.

(J) Kaplan-Meier curves for location grade predictions within WHO grade 1, 2, and 3 meningiomas. See also Figures S11 and S12.

nearest neighbors. The standard procedures for surgically derived tumor sample management is via formalin fixation and paraffin embedding. The ability to precisely incorporate data from these FFPE samples into reference landscapes generated from FF tissues opens up the possibility for any patient's tumor to be mapped onto reference landscapes like this one.

We combined multiple datasets from various sources and generated the largest meningioma reference landscape to date with comprehensive analysis of all protein-coding genes. One of our main goals was to better understand the biology of meningioma subtypes, and using transcriptomic data provided the ability to extract biological information such as expression of specific genes and pathways in a straightforward manner. To our knowledge, this is the first paper to put forth a reference map of a disease on an interactive online tool. Oncoscape is not only an attractive visualization platform for the figures but it also provides the opportunity for multifaceted exploitation of the map while mining for metadata. Overall, this study shows how we can harness the power of combining multiple datasets to extract further biological information of a particular disease. This approach may be useful in other tumor types, there is no reason to believe that success of this analysis will be unique to meningiomas.

Limitations of the study

We called gene fusions using RNA-seq raw data and recognize the lack of direct experimental evidence to support the identified fusions. However, we note that some of these fusions show regional localization on the reference map, indicating authenticity rather than noise. Apart from *NF2* mutational status known from original sources, we did not include DNA sequencing-based mutational data. A potential limitation of this study is that cross-validation can be biased toward over-optimistic results and can obscure prediction variation. We took steps to mitigate overfitting, including limiting parameter optimization search spaces and restricting the number of parameters we used cross-validation to optimize. We also randomly removed 10% of these data, trained our algorithm on the remaining 90%, and achieved similar survival and distance results when tested on the 10% of data uninvolved in training.

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 SUBJECT DETAILS
  - Collection of specimens and clinical data from University of Washington
- METHOD DETAILS
  - Specimen processing for RNA-Seq
  - Collection of publicly available RNA sequencing data
  - RNA-seq data processing and visualization
  - Clustering using DBSCAN
  - Obtaining gene fusion using RNA-Seq
  - Obtaining copy number alternations (CNAs) using RNA-Seq
  - Kaplan-Meier curves
  - Differential gene expression analysis
  - Pathway analysis
  - GSVA pathway analysis
  - Embryonic cell type analysis
  - Placing new patients on UMAP reference map
  - Oncoscape integration
  - Data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2024.100566.

ACKNOWLEDGMENTS

We thank Holland lab members for valuable discussions and FHCC core services: Specimen Processing & Research Cell Bank, Genomics, and Bioinformatics Services. This research was supported by funding from the FHCC (E.C.H), 1R35 CA253119-01A1 (E.C.H), and P30 CA015704 (Genomics & Bioinformatics Shared Resource RRID:SCR_022606).

AUTHOR CONTRIBUTIONS

Conceptualization, H.N.T., D.A.-B., N.N., S.A., F.S., and E.C.H.; methodology, H.N.T., D.A.-B., N.N., S.A., F.S., and E.C.H.; formal analysis, H.N.T., S.A., and N.N.; software, M.J.; investigation, H.N.T., D.A.-B., N.N., S.A., C.P., M.F., and C.Q.; resources, C.W.E., W.C.C., P.S., F.N., J.W., T.J.K., K.D.A., A.J.P., G.Z., C.P., F.S., M.F., and D.R.R.; data curation, H.N.T., D.A.-B., S.A., and N.N.; writing – original draft, H.N.T., N.N., and E.C.H.; writing – review & editing, S.A., F.S., P.J.C., D.R.R., J.S., and E.C.H.; visualization, H.N.T., N.N., and M.J.; supervision, J.S., M.F., and E.C.H.; funding acquisition, E.C.H.

DECLARATION OF INTERESTS

Although the majority of Oncoscape is open source, a subset of the technology and computational algorithms presented in this paper are covered by serial no. 63/595,717, and N.N., S.A., M.J., and E.C.H. are listed as inventors.

Received: January 26, 2024
Revised: April 16, 2024
Accepted: May 2, 2024
Published: May 23, 2024

REFERENCES

1. Sahm, F., Schrimpf, D., Stichel, D., Jones, D.T.W., Hielscher, T., Schefzyk, S., Okonechnikov, K., Koelsche, C., Reuss, D.E., Capper, D., et al. (2017). DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. Lancet Oncol. *18*, 682–694. https://doi.org/10.1016/S1470-2045(17)30155-9.
2. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. *23*, 1231–1251. https://doi.org/10.1093/neuonc/noab106.
3. Nassiri, F., Liu, J., Patil, V., Mamatjan, Y., Wang, J.Z., Hugh-White, R., Macklin, A.M., Khan, S., Singh, O., Karimi, S., et al. (2021). A clinically applicable integrative molecular classification of meningiomas. Nature *597*, 119–125. https://doi.org/10.1038/s41586-021-03850-3.
4. Choudhury, A., Magill, S.T., Eaton, C.D., Prager, B.C., Chen, W.C., Cady, M.A., Seo, K., Lucas, C.H.G., Casey-Clyde, T.J., Vasudevan, H.N., et al. (2022). Meningioma DNA methylation groups identify biological drivers and therapeutic vulnerabilities. Nat. Genet. *54*, 649–659. https://doi.org/10.1038/s41588-022-01061-8.

5. Wang, E.J., Haddad, A.F., Young, J.S., Morshed, R.A., Wu, J.P.H., Salha, D.M., Butowski, N., and Aghi, M.K. (2022). Recent advances in the molecular prognostication of meningiomas. Front. Oncol. *12*, 910199. https://doi.org/10.3389/fonc.2022.910199.

6. Evans, D.G.R. (2009). Neurofibromatosis type 2 (NF2): a clinical and molecular review. Orphanet J. Rare Dis. *4*, 16. https://doi.org/10.1186/1750-1172-4-16.

7. Petrilli, A.M., and Fernández-Valle, C. (2016). Role of Merlin/NF2 inactivation in tumor biology. Oncogene *35*, 537–548. https://doi.org/10.1038/onc.2015.125.

8. Agnihotri, S., Suppiah, S., Tonge, P.D., Jalali, S., Danesh, A., Bruce, J.P., Mamatjan, Y., Klironomos, G., Gonen, L., Au, K., et al. (2017). Therapeutic radiation for childhood cancer drives structural aberrations of NF2 in meningiomas. Nat. Commun. *8*, 186. https://doi.org/10.1038/s41467-017-00174-7.

9. Lee, S., Karas, P.J., Hadley, C.C., Bayley V, J.C., Khan, A.B., Jalali, A., Sweeney, A.D., Klisch, T.J., and Patel, A.J. (2019). The Role of Merlin/NF2 Loss in Meningioma Biology. Cancers *11*, 1633. https://doi.org/10.3390/cancers11111633.

10. Pemov, A., Dewan, R., Hansen, N.F., Chandrasekharappa, S.C., Ray-Chaudhury, A., Jones, K., Luo, W., Heiss, J.D., Mullikin, J.C., Chittiboina, P., et al. (2020). Comparative clinical and genomic analysis of neurofibromatosis type 2-associated cranial and spinal meningiomas. Sci. Rep. *10*, 12563. https://doi.org/10.1038/s41598-020-69074-z.

11. Brastianos, P.K., Horowitz, P.M., Santagata, S., Jones, R.T., McKenna, A., Getz, G., Ligon, K.L., Palescandolo, E., Van Hummelen, P., Ducar, M.D., et al. (2013). Genomic sequencing of meningiomas identifies oncogenic SMO and AKT1 mutations. Nat. Genet. *45*, 285–289. https://doi.org/10.1038/ng.2526.

12. Clark, V.E., Erson-Omay, E.Z., Serin, A., Yin, J., Cotney, J., Ozduman, K., Avşar, T., Li, J., Murray, P.B., Henegariu, O., et al. (2013). Genomic analysis of non-NF2 meningiomas reveals mutations in TRAF7, KLF4, AKT1, and SMO. Science *339*, 1077–1080. https://doi.org/10.1126/science.1233009.

13. Szulzewsky, F., Holland, E.C., and Vasioukhin, V. (2021). YAP1 and its fusion proteins in cancer initiation, progression and therapeutic resistance. Dev. Biol. *475*, 205–221. https://doi.org/10.1016/j.ydbio.2020.12.018.

14. Sievers, P., Chiang, J., Schrimpf, D., Stichel, D., Paramasivam, N., Sill, M., Gayden, T., Casalini, B., Reuss, D.E., Dalton, J., et al. (2020). YAP1-fusions in pediatric NF2-wildtype meningioma. Acta Neuropathol. *139*, 215–218. https://doi.org/10.1007/s00401-019-02095-9.

15. Szulzewsky, F., Arora, S., Arakaki, A.K.S., Sievers, P., Almiron Bonnin, D.A., Paddison, P.J., Sahm, F., Cimino, P.J., Gujral, T.S., and Holland, E.C. (2022). Both YAP1-MAML2 and constitutively active YAP1 drive the formation of tumors that resemble NF2 mutant meningiomas in mice. Genes Dev. *36*, 857–870. https://doi.org/10.1101/gad.349876.122.

16. Vasudevan, H.N., Braunstein, S.E., Phillips, J.J., Pekmezci, M., Tomlin, B.A., Wu, A., Reis, G.F., Magill, S.T., Zhang, J., Feng, F.Y., et al. (2018). Comprehensive Molecular Profiling Identifies FOXM1 as a Key Transcription Factor for Meningioma Proliferation. Cell Rep. *22*, 3672–3683. https://doi.org/10.1016/j.celrep.2018.03.013.

17. Prager, B.C., Vasudevan, H.N., Dixit, D., Bernatchez, J.A., Wu, Q., Wallace, L.C., Bhargava, S., Lee, D., King, B.H., Morton, A.R., et al. (2020). The Meningioma Enhancer Landscape Delineates Novel Subgroups and Drives Druggable Dependencies. Cancer Discov. *10*, 1722–1741. https://doi.org/10.1158/2159-8290.CD-20-0160.

18. Magill, S.T., Vasudevan, H.N., Seo, K., Villanueva-Meyer, J.E., Choudhury, A., John Liu, S., Pekmezci, M., Findakly, S., Hilz, S., Lastella, S., et al. (2020). Multiplatform genomic profiling and magnetic resonance imaging identify mechanisms underlying intratumor heterogeneity in meningioma. Nat. Commun. *11*, 4803. https://doi.org/10.1038/s41467-020-18582-7.

19. Vasudevan, H.N., Choudhury, A., Hilz, S., Villanueva-Meyer, J.E., Chen, W.C., Lucas, C.H.G., Braunstein, S.E., Oberheim Bush, N.A., Butowski, N., Pekmezci, M., et al. (2022). Intratumor and informatic heterogeneity influence meningioma molecular classification. Acta Neuropathol. *144*, 579–583. https://doi.org/10.1007/s00401-022-02455-y.

20. Clark, V.E., Harmancı, A.S., Bai, H., Youngblood, M.W., Lee, T.I., Baranoski, J.F., Ercan-Sencicek, A.G., Abraham, B.J., Weintraub, A.S., Hnisz, D., et al. (2016). Recurrent somatic mutations in POLR2A define a distinct subset of meningiomas. Nat. Genet. *48*, 1253–1259. https://doi.org/10.1038/ng.3651.

21. Choudhury, A., Chen, W.C., Lucas, C.H.G., Bayley, J.C., Harmanci, A.S., Maas, S.L.N., Santagata, S., Klisch, T., Perry, A., Bi, W.L., et al. (2023). Hypermitotic meningiomas harbor DNA methylation subgroups with distinct biological and clinical features. Neuro Oncol. *25*, 520–530. https://doi.org/10.1093/neuonc/noac224.

22. Patel, A.J., Wan, Y.W., Al-Ouran, R., Revelli, J.P., Cardenas, M.F., Oneissi, M., Xi, L., Jalali, A., Magnotti, J.F., Muzny, D.M., et al. (2019). Molecular profiling predicts meningioma recurrence and reveals loss of DREAM complex repression in aggressive tumors. Proc. Natl. Acad. Sci. USA *116*, 21715–21726. https://doi.org/10.1073/pnas.1912858116.

23. Slavik, H., Balik, V., Kokas, F.Z., Slavkovsky, R., Vrbkova, J., Rehulkova, A., Lausova, T., Ehrmann, J., Gurska, S., Uberall, I., et al. (2022). Transcriptomic Profiling Revealed Lnc-GOLGA6A-1 as a Novel Prognostic Biomarker of Meningioma Recurrence. Neurosurgery *91*, 360–369. https://doi.org/10.1227/neu.0000000000002026.

24. Viaene, A.N., Zhang, B., Martinez-Lage, M., Xiang, C., Tosi, U., Thawani, J.P., Gungor, B., Zhu, Y., Roccograndi, L., Zhang, L., et al. (2019). Transcriptome signatures associated with meningioma progression. Acta Neuropathol. Commun. *7*, 67. https://doi.org/10.1186/s40478-019-0690-x.

25. Lilly, J.V., Rokita, J.L., Mason, J.L., Patton, T., Stefankiewiz, S., Higgins, D., Trooskin, G., Larouci, C.A., Arya, K., Appert, E., et al. (2023). The children's brain tumor network (CBTN) - Accelerating research in pediatric central nervous system tumors through collaboration and open science. Neoplasia *35*, 100846. https://doi.org/10.1016/j.neo.2022.100846.

26. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106. https://doi.org/10.1186/gb-2010-11-10-r106.

27. Arora, S., Szulzewsky, F., Jensen, M., Nuechterlein, N., Pattwell, S.S., and Holland, E.C. (2023). Visualizing genomic characteristics across an RNA-Seq based reference landscape of normal and neoplastic brain. Sci. Rep. *13*, 4228. https://doi.org/10.1038/s41598-023-31180-z.

28. McFerrin, L.G., Zager, M., Zhang, J., Krenn, G., McDermott, R., Horse-Grant, D., Silgard, E., Colevas, K., Shannon, P., Bolouri, H., and Holland, E.C. (2018). Analysis and visualization of linked molecular and clinical cancer data by using Oncoscape. Nat. Genet. *50*, 1203–1204. https://doi.org/10.1038/s41588-018-0208-7.

29. Sievers, P., Sill, M., Blume, C., Tauziede-Espariat, A., Schrimpf, D., Stichel, D., Reuss, D.E., Dogan, H., Hartmann, C., Mawrin, C., et al. (2021). Clear cell meningiomas are defined by a highly distinct DNA methylation profile and mutations in SMARCE1. Acta Neuropathol. *141*, 281–290. https://doi.org/10.1007/s00401-020-02247-2.

30. Reuss, D.E., Piro, R.M., Jones, D.T.W., Simon, M., Ketter, R., Kool, M., Becker, A., Sahm, F., Pusch, S., Meyer, J., et al. (2013). Secretory meningiomas are defined by combined KLF4 K409Q and TRAF7 mutations. Acta Neuropathol. *125*, 351–358. https://doi.org/10.1007/s00401-013-1093-x.

31. Najm, P., Zhao, P., Steklov, M., Sewduth, R.N., Baietti, M.F., Pandolfi, S., Criem, N., Lechat, B., Maia, T.M., Van Haver, D., et al. (2021). Loss-of-Function Mutations in TRAF7 and KLF4 Cooperatively Activate RAS-Like GTPase Signaling and Promote Meningioma Development. Cancer Res. *81*, 4218–4229. https://doi.org/10.1158/0008-5472.CAN-20-3669.

32. Amoo, M., Henry, J., Farrell, M., and Javadpour, M. (2023). Meningioma in the elderly. Neurooncol. Adv. *5*, i13–i25. https://doi.org/10.1093/noajnl/vdac107.

33. Sun, T., Plutynski, A., Ward, S., and Rubin, J.B. (2015). An integrative view on sex differences in brain tumors. Cell. Mol. Life Sci. *72*, 3323–3342. https://doi.org/10.1007/s00018-015-1930-2.

34. Bayley, J.C., 5th, Hadley, C.C., Harmanci, A.O., Harmanci, A.S., Klisch, T.J., and Patel, A.J. (2022). Multiple approaches converge on three biological subtypes of meningioma and extract new insights from published studies. Sci. Adv. *8*, eabm6247. https://doi.org/10.1126/sciadv.abm6247.

35. Parada, C.A., Osbun, J.W., Busald, T., Karasozen, Y., Kaur, S., Shi, M., Barber, J., Adidharma, W., Cimino, P.J., Pan, C., et al. (2020). Phosphoproteomic and Kinomic Signature of Clinically Aggressive Grade I (1.5) Meningiomas Reveals RB1 Signaling as a Novel Mediator and Biomarker. Clin. Cancer Res. *26*, 193–205. https://doi.org/10.1158/1078-0432.CCR-18-0641.

36. Chen, W.C., Choudhury, A., Youngblood, M.W., Polley, M.-Y.C., Lucas, C.-H.G., Mirchia, K., Maas, S.L.N., Suwala, A.K., Won, M., Bayley, J.C., et al. (2023). Targeted gene expression profiling predicts meningioma outcomes and radiotherapy responses. Nat. Med. *29*, 3067. https://doi.org/10.1038/s41591-023-02586-z.

37. Chotai, S., and Schwartz, T.H. (2022). The Simpson Grading: Is It Still Valid? Cancers *14*, 2007. https://doi.org/10.3390/cancers14082007.

38. Uhrig, S., Ellermann, J., Walther, T., Burkhardt, P., Fröhlich, M., Hutter, B., Toprak, U.H., Neumann, O., Stenzinger, A., Scholl, C., et al. (2021). Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. *31*, 448–460. https://doi.org/10.1101/gr.257246.119.

39. Serin Harmanci, A., Harmanci, A.O., and Zhou, X. (2020). CaSpER identifies and visualizes CNV events by integrative analysis of single-cell or bulk RNA-sequencing data. Nat. Commun. *11*, 89. https://doi.org/10.1038/s41467-019-13779-x.

40. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. *14*, 128. https://doi.org/10.1186/1471-2105-14-128.

41. Ghaleb, A.M., and Yang, V.W. (2017). Kruppel-like factor 4 (KLF4): What we currently know. Gene *611*, 27–37. https://doi.org/10.1016/j.gene.2017.02.025.

42. Qiu, C., Martin, B.K., Welsh, I.C., Daza, R.M., Le, T.M., Huang, X., Nichols, E.K., Taylor, M.L., Fulton, O., O'Day, D.R., et al. (2023). A single-cell transcriptional timelapse of mouse embryonic development, from gastrula to pup. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.05.535726.

43. Salsi, V., Vigano, M.A., Cocchiarella, F., Mantovani, R., and Zappavigna, V. (2008). Hoxd13 binds in vivo and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. Dev. Biol. *317*, 497–507. https://doi.org/10.1016/j.ydbio.2008.02.048.

44. Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

45. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

46. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

47. Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom. Bioinform. *2*, lqaa078. https://doi.org/10.1093/nargab/lqaa078.

48. Ester, M.,K.,H.-P., Sander, J., and Xu, X.; others (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise (KDD-96 Proceedings).

49. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

50. Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 1–15. https://doi.org/10.1186/1471-2105-14-7.

51. Hie, B., Cho, H., DeMeo, B., Bryson, B., and Berger, B. (2019). Geometric Sketching Compactly Summarizes the Single-Cell Transcriptomic Landscape. Cell Syst. *8*, 483–493.e7. https://doi.org/10.1016/j.cels.2019.05.003.

52. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
In-house sequenced human meningioma tumor samples	Department of Neurological Surgery, University of Washington Medical Center	GSE252291
Deposited data		
Raw and analyzed RNA-Seq data	This paper	GEO: GSE252291
Heidelberg	Sahm Lab, University Hospital Heidelberg, Germany[14]	N/A
UoT	University of Toronto, Canada[3]	European Genome Archive (EGA): EGAS00001004982
CAVATICA	CAVATICA[24]	N/A
UCSF(2018)	University of California San Francisco[16]	GEO: GSE101638
Baylor	Baylor College of Medicine[22]	GEO: GSE136661
UCSD	University of California San Diego[17]	GEO: GSE139652
UCSF(2020)	University of California San Francisco[18]	GEO: GSE151921
UCSF (2022)	University of California San Francisco[4]	GEO: GSE183656
Yale	Yale University School of Medicine[20]	GEO: GSE85133
Palacky	Palacky University and University Hospital Olomouc[23]	NIH BioProject ID: PRJNA705586
CHOP	Children's' Hospital of Philadelphia[25]	CBTN
HKU/UCSF	University of California San Francisco and University of Hong Kong[21]	GEO: GSE212666
Gene sets for gene ontology biological processes	MSigDB	https://www.gsea-msigdb.org/gsea/msigdb/
scRNA-Seq data of mouse embryos	Qiu et al.[42]	N/A
Human reference genome	Gencode GRCh38.primary_assembly	https://www.gencodegenes.org/human/release_39.html
Software and algorithms		
SRA-Toolkit 2.11.0 (fastq-dump)		https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software
FASTQC		https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MULTIQC	Ewels et al.[44]	N/A
STAR2(v2.7.7a)	Dobin et al.[45]	N/A
HTSeq (v0.11.0)	Anders et al.[46]	N/A
ComBat-seq (sva package)	Zhang et al.[47]	N/A
DBSCAN	Ester et al.[48]	N/A
Arriba (v2.1.0)	Uhrig et al.[38]	N/A
CaSpER	Serin et al.[39]	N/A
R package "survival" (v3.5.7)		https://github.com/therneau/survival
Python package lifelines (v 0.27.7)		https://lifelines.readthedocs.io/en/latest/index.html
edgeR	Robinson et al.[49]	N/A
EnrichR	Chen et al.[40]	N/A
R package "GSVA"	Hanzelmann et al.[50]	N/A
geosketch	Hie et al.[51]	N/A
Seurat/v3	Stuart et al.[52]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DESeq2	Anders et al.[26]	N/A
Custom code used in this study		https://github.com/FredHutch/MeningiomaLandscape-HollandLab Zenodo (https://doi.org/10.5281/zenodo.11069609)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Eric C. Holland (eholland@fredhutch.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

RNA-Seq raw data from meningiomas collected from UW/FHCC are deposited in GEO GSE252291. All custom code used in this study are available at https://github.com/FredHutch/MeningiomaLandscape-HollandLab and at Zenodo (https://doi.org/10.5281/zenodo.11069609).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Collection of specimens and clinical data from University of Washington

Studies were conducted in accordance with the U.S. Common Rule ethical guidelines. Tissue was collected from study participants admitted at the University of Washington Medical Center, Department of Neurological Surgery. The respective clinical data was extracted from the University of Washington, School of Medicine clinical database. Data and specimen collection were reviewed and approved by the University of Washington Institutional Review Board and Human Subjects Division. Written-informed consent was obtained from all subjects. Patients underwent surgery at the University of Washington Medical Center between January 1, 1998 and June 1, 2023. Samples were collected and stored in −80°C. Clinical data were gathered regarding history, demographics, imaging, neuropathology reports, operative information, adjuvant treatment, and outcomes. Resected tumors were graded according to the current criteria (https://doi.org/10.1093/neuonc/noab106) and correlated with clinical information, when advised. Histologic subtype, mitoses, Ki-67/MIB, sheeting, macronuclei, hypercellularity, invasion, necrosis, TERT promoter mutations (https://doi.org/10.1093/jnci/djv377)) and CDKN2a/b homozygous deletions (https://doi.org/10.1007/s00401-020-02188-w) were recorded. Specimens were reviewed by three neuropathologists and neurosurgeons. Total resection was defined as absence of residual enhancement on post-operative MRI within 48 h of surgery. Recurrence was defined as at least 1 cm of enhancement on subsequent MRI. Progression was considered to be at least 1 cm of growth of residual tumors detected on MRI after surgery.

METHOD DETAILS

Specimen processing for RNA-Seq

RNA was extracted using the Qiagen RNeasy Plus mini kit. Total RNA integrity was checked using an Agilent 4200 TapeStation (Agilent Technologies, Inc., Santa Clara, CA) and quantified using a Trinean DropSense96 spectrophotometer (Caliper Life Sciences, Hopkinton, MA). Samples with RIN <5 were removed from further analysis. RNA was normalized to 50ng/ul and 500ng was submitted for library preparation. RNA-seq libraries were prepared from total RNA using the TruSeq Stranded mRNA kit (Illumina, Inc., San Diego, CA, USA). Sequencing was performed using an Illumina NovaSeq 6000 employing a paired-end, 50 base read length (PE50) sequencing strategy. RNA extraction was performed by the Specimen Processing & Research Cell Bank at Fred Hutch. Library preparation and sequencing was performed by the Genomics and Bioinformatics Core Services at Fred Hutch.

Collection of publicly available RNA sequencing data

Raw sequencing data of human meningioma samples were obtained from respective public data repositories (Table S1) except Heidelberg dataset which was obtained from the data repository of the Dept. of Neuropathology at the University Hospital Heidelberg. SRA files downloaded from GEO were converted to fastq files using fastq-dump from the SRA-Toolkit (v2.11.0).

RNA-seq data processing and visualization

Quality check on raw RNA sequencing data was done using FastQC (v0.11.9) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (v1.9) tools.[44] RNA sequencing reads were aligned to the Gencode GRCh38.primary_assembly genome using STAR2 (v2.7.7a) and then using HTSeq (v0.11.0) reads were counted for each associated gene using the Gencode V39 primary assembly annotations.[45,46] Raw gene counts from each dataset were combined and corrected for batch effects using ComBat-seq from the R package "sva".[47] Gene expression values from combined datasets were normalized using the Variance Stabilizing Transformation (VST) method.[26] Uniform Manifold Approximation and Projection (UMAP), a dimensionality reduction method, was applied on normalized counts from 19979 protein-coding genes to create the meningioma reference landscape.[27] UMAPs were constructed using the R package "umap" (https://cran.r-project.org/web/packages/umap/index.html).

Clustering using DBSCAN

DBSCAN (density-based spatial clustering of applications with noise) was used to confirm the clusters identified by UMAP.[48]

Obtaining gene fusion using RNA-Seq

Arriba (v2.1.0) was used to compute gene fusions from two-pass STAR-aligned RNA-Seq reads.[38] All fusion analyses were restricted to fusion calls Arriba indicated were high confidence. Using gencode.v38.annotation.gtf.gz from hg38 release 44 (GRCh38.p14) from https://www.gencodegenes.org/human/we determined whether a gene was protein coding and selected only fusions with at least one coding gene involved. Fusions that recur at least twice within the dataset were used to calculate fusion burden.

Obtaining copy number alternations (CNAs) using RNA-Seq

Large scale/chromosome arm level copy number alternations were estimated for all tumors using the package CaSpER (https://rpubs.com/akdes/673120) on bulk RNA-Seq data.[39] BAFExtract source code, genome list and genome pileup directory were downloaded from https://github.com/akdess/. hg38 cytoband and centromere information were downloaded from UCSC. GTEX RNA-Seq data from normal frontal cortex and hippocampus (dbGaP Accession: phs000424.v8.p2) were used as control samples in the CaSpER analysis (GTEX sample IDs: SRR1147618, SRR1334440, SRR1337431, SRR1342045, SRR1348360, SRR1354446, SRR1360128, SRR1375571, SRR1388305, SRR1402900, SRR1408368, SRR1413562, SRR1416477, SRR1435775, SRR1453341, SRR1471817, SRR1488367, SRR1488651, SRR1500868).

Kaplan-Meier curves

Kaplan-Meier curves were generated using the information on time to recurrence of each sample. To perform the calculations, we only selected tumors with known recurrence status (recurrence = yes/no) and known time to recurrence or last follow up. For tumors that were confirmed as non-recurrent however with no last follow up date, a default of 315 months was used as "Months of No Recurrence". Kaplan-Meier curves were plotted using the R package "survival" or Python package "lifelines" and p-values were calculated using the log -rank test (v3.5.7). Hazard ratios and p-vaues multivariate survival analyses were calculated using Cox Proportional Hazard (CPH) regression models.

Differential gene expression analysis

To determine the biological signature of each cluster, differential expression analysis was performed between each cluster and the rest of the UMAP samples using edgeR.[49] Upregulated genes in each cluster were identified based on FDR (<0.05) and log fold change (>0.6) or fold change of 1.5 cut-off. Downregulated genes in each cluster were identified based on FDR (<0.05) and log fold change (<−0.6). Upregulated and downregulated genes in each cluster are listed in Table S5A.

Pathway analysis

Top 500 significant upregulated genes in each cluster (with the exception of cluster C = ∼700 genes) were analyzed using Enrichr to determine the enriched biological signature.[40] Gene Ontology Biological Processes (GO BP) terms that were statistically significant (adjusted p value <0.05) were manually curated to remove GO BP terms that had gene sets >80% overlapping. Dot plots were generated using ggplot2 (top 15 GO terms included). Complete lists of GO terms were attached as supplement information.

GSVA pathway analysis

Gene sets for pathways from Gene Ontology Biological Processes were downloaded from Molecular Signature Databases (MSigDB) (v7.2) (https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). Gene Set Variation Analysis (GSVA) was performed on batch corrected VST normalized counts from all 1298 samples. GSVA scores obtained from 1 and -1 for each sample were visualized using ggplot2. Similarly, GSVA analysis was performed using UCSF gene set (34 genes upregulated or downregulated in aggressive meningioma). Upregulated and downregulated genes were considered as two separate gene sets.

Embryonic cell type analysis

The scRNA-seq data of mouse embryos was downloaded from Qiu et al.[42] In the study, over 11 million cells were profiled from mouse embryos during organogenesis and fetal development, with every 6 h temporal resolution ranging from embryonic day 8 (E8) to

postnatal day 0 (P0). This resulted in the identification of ~190 cell types. To save computational time and memory, the dataset was downsampled to ~1 million cells using geosketch.[51] Top 500 upregulated protein-coding genes in each cluster (with the exception of cluster C = ~700 genes) were obtained from the previously mentioned differential expression analysis. A gene module score was calculated for individual groups of genes using the *AddModuleScore* function implemented in Seurat/v3.[52] Gene module score was calculated for each cell type and the mean score of each major cell cluster was calculated for Figure 5D. Welch two sample t-test was performed to statistically confirm the top cell clusters enriched in each meningioma cluster.

Placing new patients on UMAP reference map

The stages of our algorithm, which places new patients on our UMAP reference map, are described below. These stages consist of pre-training UMAP models, mapping new patients to embeddings generated by pre-trained UMAP models, determining nearest neighbors on UMAP embeddings, aggregating all sets of UMAP embedding-derived nearest neighbors, and using the frequency of these nearest neighbors to compute a centroid. This last process also involves removing outliners before computing the centroid.

UMAP pretraining

We train $K = 100$ UMAP models on the 1298 samples in our RNA-Seq reference dataset D using a different random state for each UMAP model. We denote each of the K pre-trained UMAP models by the function.

$$UMAP_k : R^{\#genes} \rightarrow R^2$$

which maps any VST sample to two dimensions. In this way, we can represent the embedding of the reference dataset given by each pre-trained UMAP model as $UMAP_k(D)$. We denote the x-coordinate and y-coordinate of every reference sample p_i in an embedding given by $UMAP_k$ as $UMAP_k(p_i)_1$ and $UMAP_k(p_i)_2$, respectively. Similarly, we will refer to the x-coordinate and y-coordinate of every reference sample p_i in the reference landscape as $UMAP(p_i)_1$ and $UMAP(p_i)_2$. All UMAP embeddings given by $UMAP_k$ are normalized by centering the embeddings and scaling points so that the average distance between points is 1.

Mapping samples to embeddings given by pre-trained UMAP models

Given K pre-trained UMAP models, we can place any new patient P onto a UMAP embedding $UMAP_k(D)$ by passing the VST data of that patient (P_{VST}) through the UMAP model $UMAP_k$. We denote this position in the embedding as $UMAP_k(P_{VST})$, where we will refer to the x-coordinate of $UMAP_k(P_{VST})$ as $UMAP_k(P_{VST})_1$ and y-coordinate as $UMAP_k(P_{VST})_2$.

Computing 100 sets of nearest neighbors from pre-trained UMAP embeddings

To get a set of nearest neighbors of a patient P on an embedding generated by a pre-trained UMAP model $UMAP_k$, we first compute the distances between the position of the new patient P in the embedding and the position of all other reference samples p_i. Thus, for every patient p_i in the reference dataset and every pre-trained UMAP model $UMAP_k$ we compute the square of the Euclidian distances $d_k(p_i, P)$ between $UMAP_k(P_{VST})$ and the 1298 positions $UMAP_k(p_i)$ on each of the 100 embeddings generated by the 100 $UMAP_k$ maps. The square root of $d_k(p_i, P)$ is not taken for efficiency. We define $d_k(p_i, P)$ as follows:

$$d_k(p_i, P) = \left(UMAP_k(p_i)_1 - UMAP_k(P_{VST})_1\right)^2 + \left(UMAP_k(p_i)_2 - UMAP_k(P_{VST})_2\right)^2$$

To obtain a list of nearest neighbors for the patient P on each UMAP model $UMAP_k$, we choose the 100 samples p_i with the smallest $d_k(p_i, P)$ under the constraint $d_k(p_i, P) < \alpha$, where $\alpha = 0.05$ is the square root of the Euclidian radius chosen so that the list of nearest neighbors would not span multiple unconnected clusters. We denote the set of nearest neighbors to a patient P on for a pre-trained UMAP model $UMAP_k$ as $S^k_{NN}(\mathrm{P})$.

Computing the frequency vector used to weight the centroid calculation

We construct a matrix M whose values represent whether a sample in the reference dataset p_i is a nearest neighbor of the new patient P in the k th UMAP embedding $UMAP_k(D)$. Formally, this can be expressed as

$$M_{i,k} = 1\left\{p_i \in S^k_{NN}(P)\right\}$$

where 1 is the indicator function. To find the frequency vector f^P of new patient P, we average the values of M across the K UMAP embeddings. This results in a vector of length 1298 whose value f^P_i represents the frequency sample p_i was a nearest neighbor of P across the K pre-trained UMAP embeddings.

Compute centroid

The weighted centroid of a frequency vector f^P and the x-coordinate $UMAP(p_i)_1$ and y-coordinate $UMAP(p_i)_2$ of each sample on the reference UMAP is compute as follows:

$$C\left(f^P\right) = \left(\frac{\sum_{i=1}^{N} f^P_i \cdot UMAP(p_i)_1}{\sum_{i=1}^{N} f^P_i}, \frac{\sum_{i=1}^{N} f^P_i \cdot UMAP(p_i)_2}{\sum_{i=1}^{N} f^P_i}\right)$$

Remove outliers in the frequency vector

To prevent samples p_i which may have been included in some set of nearest neighbors $S^k_{NN}(P)$ by chance, we create an adjusted frequency vector $\hat{f}^P$ from f^P by setting elements f^P_i to zero for all such p_i. We first set $\hat{f}^P$ equal to f^P. Next, we adjust $\hat{f}^P$ by setting values less than 0.25 to 0 and compute the weighted centroid $C(\hat{f}^P)$ of $\hat{f}^P$ (described above). Then we compute the distance of all

samples p_i with non-zero $\hat{f}_i^P$ values to $C(\hat{f}^P)$ and remove those with distances greater than the 95% quantile. Last, we recompute the centroid $C(\hat{f}^P)$ with the updated $\hat{f}^P$ and set the values of $\hat{f}^P$ for samples farther than $R = 0.75$ from the centroid to zero. he radius R was chosen using cross-validated results and visual inspection.

Final placement of the patient

To place a new patient P on our reference UMAP, we simply compute $\hat{f}^P$ and place the patient at the coordinates given by $C(\hat{f}^P)$ as described above.

Uncertain placements

We compute a score to quantify the quality of our placements based on the distribution of frequencies in f^P, which is computed as follows

$$s(f_i^P) = \frac{1}{N}\sum_{i=1}^{N}(1 - f_i^P)$$

Here $s(f_i^P)$ describes the average frequency every reference sample p_i is a nearest neighbor of P on the UMAP embeddings $UMAP_k(D)$ for samples that are nearest neighbors at least once (i.e., p_i such that $f_i^P \neq 0$). Lower values of $s(f_i^P)$ indicate that the sets of nearest neighbors were more consistent over all $S_{NN}^k(P)$ and suggest more reliable predictions. For this reason, we consider predictions for which $s(f_i^P) < 0.75$. This threshold was established empirically using cross-validation.

Cross-validation

To conduct cross-validation, we repeated the following procedure for each patient p_i in our reference dataset D. First, we pre-trained 100 UMAP models on $D \backslash p_i$, the reference dataset without the sample p_i. Afterward, we treated p_i as a new patient and placed p_i on our reference UMAP at the location $C(f^{p_i})$ as described above. Finally, we computed the distributions of the Euclidean distances between the predicted centroid $C(\hat{f}^{p_i})$ and the ground truth location of p_i on the reference UMAP for each p_i

$$\left\{\left\|UMAP(p_i), C(\hat{f}^{p_i})\right\|_2 \middle| i \in \{1, \ldots, 1298\}\right\}$$

where

$$\left\|UMAP(p_i), C(\hat{f}^{p_i})\right\|_2 = \sqrt{\sum_{j=1}^{2}\left(UMAP(p_i)_j - C(\hat{f}^{p_i})_j\right)^2}$$

90/10 *training/testing validation*

To establish a 90/10 split of our dataset into a training set and test set, we randomly removed 10% of samples from each data source in our dataset to form a separate test dataset ($N = 128$); the remaining samples ($N = 1170$) made up our training set. We then pre-trained 100 UMAP models on the training set and used these UMAP models to place each test sample on a fixed UMAP reference map generated from the training data. To assess the performance of our method on the test set, we first used the nearest-neighbors approach described below in the "WHO grade prediction" section to assign a location-based tumor grade and evaluated the extent to which these predictions stratified outcomes. We also assessed the accuracy of our method on the samples in the test set using a distance metric. We first approximated the ground truth location of the samples in the test set on the training reference UMAP. For each test sample, we considered its approximate ground truth location on the training reference UMAP as the centroid of the samples that were the nearest neighbors of that test sample on the full 1298-sample reference UMAP. Then we examined the distribution of the distances between the placement location and the approximated ground truth location for each test sample.

Cluster prediction

To predict the cluster membership of a new patient, we mapped the patient onto our reference UMAP and computed its nearest neighbors.

We used a majority vote strategy which predicted the most common cluster found in this set of nearest neighbors. Cluster predictions given in the main text were computed via cross-validation.

WHO grade prediction.

We predicted WHO grade using the same procedure we used to predict cluster membership. We mapped a new patient onto our reference UMAP and computed its nearest neighbors. A majority vote strategy issued the most common WHO grade among the set of nearest neighbors as the WHO grade prediction. WHO grade predictions given in the main text were computed via cross-validation.

Oncoscape integration

Matrix and clinical data were prepared for Oncoscape by converting them to cBioPortal formats (cbioportal.org). Custom settings, including colorings and precalculated views to match the paper's figures, were stored in JSON in an Oncoscape updates.txt file. See https://oncoscape.sttrcancer.org/assets/uploading.html for details.

Data analysis

All analysis including statistics and visualization were done in R version 4.2.0 (2022-04-22) as implemented in Rstudio (2022.05.999). Plots were generated using R basic graphics and ggplot2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Kaplan-Meier curves were generated using the information on time to recurrence of each sample. To perform the calculations, we only selected tumors with known recurrence status (recurrence = yes/no) and known time to recurrence or last follow up. For tumors that were confirmed as non-recurrent however with no last follow up date, a default of 315 months was used as ''Months of No Recurrence''. Kaplan-Meier curves were plotted using the R package ''survival'' (for Figures 3 and 4, S3 and S4) or Python package ''lifelines'' and p-values were calculated using the log -rank test (v3.5.7) (for Figures 7, S11 and S12). Details on statistical analysis is mentioned in the STAR Methods. Number of samples and p-values are listed on the figure or in figure legend. Pathway enrichment analysis was done using Enrichr and adjusted p-value was calculated using the inbuilt Benjamini-Hotchberg method (Figures 5, S8). In the mouse embryonic cell type analysis, Welch two sample t-test was performed to statistically confirm the top cell clusters enriched in each meningioma cluster. FDR and fold change cut off in differential expression analysis are mentioned above in methods.