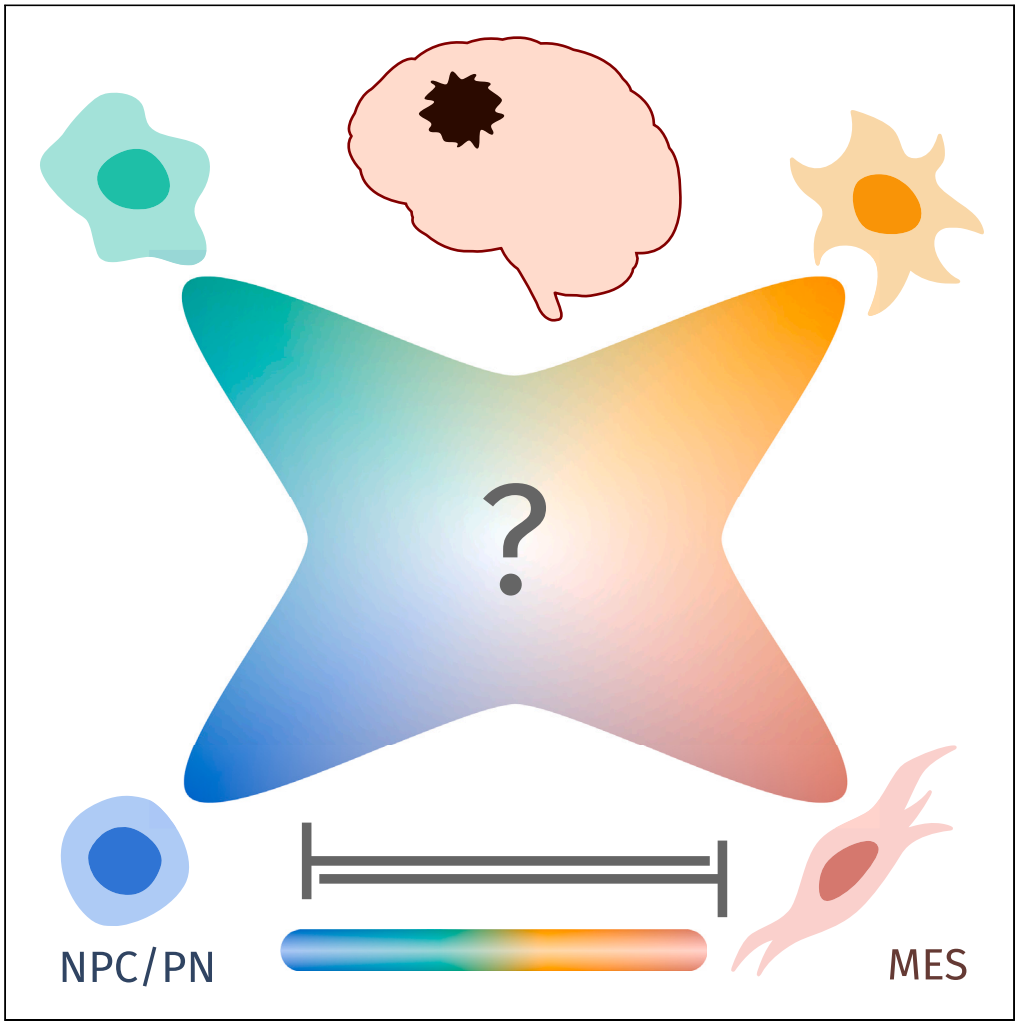
iScience

CellPress
OPEN ACCESS

Article

Proneural-mesenchymal antagonism dominates the patterns of phenotypic heterogeneity in glioblastoma



Harshavardhan BV, Mohit Kumar Jolly

mkjolly@iisc.ac.in

Highlights

Challenges the assumption of four mutually exclusive transcriptomic subtypes in GBM

Reveals proneural-mesenchymal antagonism as the dominant axis of GBM heterogeneity

Unravels functional differences in terms of cell cycle, metabolism and immune evasion

BV & Jolly, iScience *27*, 109184
March 15, 2024
https://doi.org/10.1016/j.isci.2024.109184

iScience

CellPress
OPEN ACCESS

Article

Proneural-mesenchymal antagonism dominates the patterns of phenotypic heterogeneity in glioblastoma

Harshavardhan BV[1] and Mohit Kumar Jolly[2,3,*]

SUMMARY

The aggressive nature of glioblastoma (GBM) – one of the deadliest forms of brain tumors – is majorly attributed to underlying phenotypic heterogeneity. Early attempts to classify this heterogeneity at a transcriptomic level in TCGA GBM cohort proposed the existence of four distinct molecular subtypes: Proneural, Neural, Classical, and Mesenchymal. Further, a single-cell RNA sequencing (scRNA-seq) analysis of primary tumors also reported similar four subtypes mimicking neurodevelopmental lineages. However, it remains unclear whether these four subtypes identified via bulk and single-cell transcriptomics are mutually exclusive or not. Here, we perform pairwise correlations among individual genes and gene signatures corresponding to these proposed subtypes and show that the subtypes are not distinctly mutually antagonistic in either TCGA or scRNA-seq data. We observed that the proneural (or neural progenitor-like)-mesenchymal axis is the most prominent antagonistic pair, with the other two subtypes lying on this spectrum. These results are reinforced through a meta-analysis of over 100 single-cell and bulk transcriptomic datasets as well as in terms of functional association with metabolic switching, cell cycle, and immune evasion pathways. Finally, this proneural-mesenchymal antagonistic trend percolates to the association of relevant transcription factors with patient survival. These results suggest rethinking GBM phenotypic characterization for more effective therapeutic targeting efforts.

INTRODUCTION

Heterogeneity in GBM exists at many levels: functional, molecular, interpatient, and intratumoural.[1] It is a major clinical challenge due to varied susceptibilities of cellular subpopulations to treatment. Additionally, GBM cells exhibit a remarkable degree of plasticity, allowing them to adapt and evolve rapidly in response to diverse treatments.[2] Similar to observations in breast cancer, lung cancer and melanoma,[3–8] GBM subpopulations can also undergo genetic and/or non-genetic (reversible) cell-state transitions,[9,10] often driving drug resistance and eventual GBM recurrence. Understanding the molecular underpinnings of such plasticity and heterogeneity is, therefore, essential.

An earlier key study characterizing heterogeneity in GBM used transcriptomic data from the TCGA-GBM cohort and reported 4 subtypes: Proneural (TCGA-PN), Neural (TCGA-NL), Classical (TCGA-CL), and Mesenchymal (TCGA-MES).[11] Except for the neural subtype, other subtypes were associated with specific gene abnormalities: EGFR alterations in classical, NF1 mutations in mesenchymal, and PDGFRA and IDH1 mutations in proneural. Upon quantifying the enrichment of gene expression profiles of neural cell types in GBM, the proneural samples were highly enriched for oligodendrocytic signature, while the mesenchymal ones were strongly associated with cultured astroglial signature. Further analysis suggested that the neural subtype may be associated with contamination of normal neural cells in the GBM sample, indicating that proneural, mesenchymal and classical phenotypes may be the GBM-specific ones.[12,13] Multi-region tumor sampling and single-cell RNA-sequencing (scRNA-seq) demonstrated that these different molecular subtypes can co-exist in the same tumor specimen.[13–15] These observations are reminiscent of intra-tumor heterogeneity along the proliferative-invasive spectrum in melanoma samples[16] and along the epithelial-mesenchymal axis in carcinomas.[17–19]

To further characterize intra-tumor heterogeneity, scRNA-seq of GBM tumors from 28 adult and pediatric patients were used, and 4 meta-modules mimicking neuro-developmental lineages were identified: neural-progenitor-like (NPC-like), oligodendrocyte-progenitor-like (OPC-like), astrocyte-like (AC-like), and mesenchymal-like (MES-like) states. Each tumor analyzed contained cells in at least 2 of these 4 cell-states, with varied relative ratios. These subtypes had amplifications of CDK4, PDGFRA, EGFR, and NF1 respectively. Consistently, AC-like and MES-like states were found to correspond to TCGA-CL and TCGA-MES subtypes, respectively, while TCGA-PN subtype corresponded to a combination of OPC-like and NPC-like cell-states.[20] Together, these landmark studies in GBM heterogeneity posited the (co-)

[1]IISc Mathematics Initiative, Indian Institute of Science, Bengaluru, Karnataka 560012, India
[2]Department of Bioengineering, Indian Institute of Science, Bengaluru, Karnataka 560012, India
[3]Lead contact
*Correspondence: mkjolly@iisc.ac.in
https://doi.org/10.1016/j.isci.2024.109184

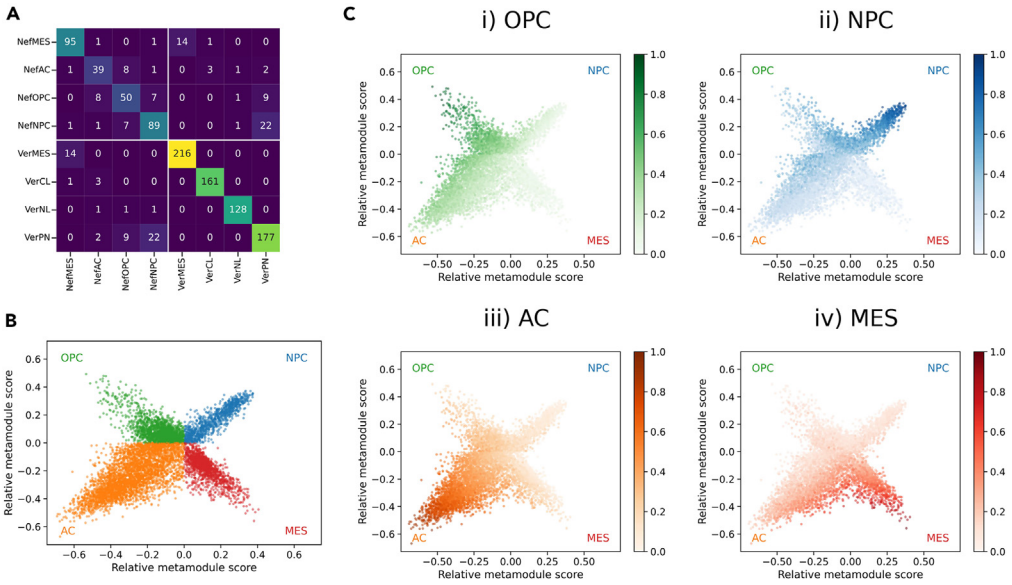


Figure 1. Subtypes defined for GBM

(A) Representation of number of genes within each gene set, highlighting the common genes between different gene sets.

(B) Recreation of the two-dimensional representation of cellular states as defined in Neftel et al.[20]

(C) Two-dimensional representation colored by scores for each subtype: (i) OPC-like, (ii) NPC-like, (iii) AC-like, (iv) MES-like.

existence of four distinct cellular states. However, whether these four cell-states are mutually exclusive or antagonistic among one another or not remains elusive.

Here, we first quantify pairwise correlations among the scores of all four proposed subtypes in TCGA and scRNA-seq datasets based on which these subtypes were proposed and demonstrate that only proneural (or neural-progenitor-like)-mesenchymal pair shows a strong antagonistic pattern. Next, we show that this antagonism is also present at the individual gene level among these two states. Finally, a meta-analysis of more than 100 bulk and scRNA-seq GBM datasets endorses this analysis and reveals functional differences between proneural and mesenchymal states in terms of their association with cell cycle, metabolic status and immune evasion. Our results thus suggest that the proneural-mesenchymal switch dominates the patterns of GBM phenotypic heterogeneity.

RESULTS

Scoring of subtype-specific signatures reveal that not all cell states are mutually antagonistic

The four proposed sub-types were classified at a transcriptomic level and thus characterized by sets of genes highly expressed in each of them.[11,20] Thus, we quantified the enrichment of different gene sets in each sample in a given cohort. If the proposed subtypes correspond to distinct cell-states, we expect to see antagonism or independence in the expression patterns of these gene sets. Mutually exclusive or antagonistic gene expression profiles can emerge from one set of genes/master regulators of one phenotype suppressing those from the other sets, as witnessed in other cancers exhibiting phenotypic plasticity.[21,22] Thus, antagonism in gene regulatory networks is expected to manifest as a negative correlation between the enrichment scores of such gene sets. However, if the gene modules corresponding to different phenotypes do not regulate one another, we expect to see no correlation among their scores. As expected, the gene sets corresponding to these four subtypes - (TCGA-CL, TCGA-MES, TCGA-PN, TCGA-NL)[11] or (NPC-like, APC-like, OC-like, and MES-like)[20] - do not show much overlap (Figure 1A).

We first recreated the graphical representation presented earlier[20] using corresponding enrichment scores of the signatures corresponding to the proposed four subtypes (Figure 1B). This projection adopts a discontinuous x axis, which, while being a valid stylistic choice, introduces a visual separation that may inadvertently emphasize the distinctions between cell states, leading to the impression that the four states are all mutually antagonistic and that they correspond to four independent dimensions. Moreover, the utilization of the "max" operation between sets of subtypes for the y axis could contribute to this exaggeration. Thus, we questioned whether this portrayal could be inadvertently obscuring the true and intricate relationships between the identified states. First, we quantified the scores of individual subtypes on this projection (Figure 1C). The scores of a particular subtype are relatively high in the "arm" corresponding to that subtype but not exclusive; also, the discontinuity of the x axis becomes clearly visible with a break in the gradient. This analysis suggested that the use of a discontinuous x axis was at least partly distorting the true relationship among the proposed subtypes in the underlying gene expression space, and thus, unlike the visual impression created, the four subtypes may not be as mutually exclusive among one another.

We investigated the two source datasets: GSE131928 and TCGA-GBM, from which the Neftel signature and Verhaak signature were derived, respectively. In both these groups of four gene sets each, we noticed that not all pairwise correlations are negative; instead, many

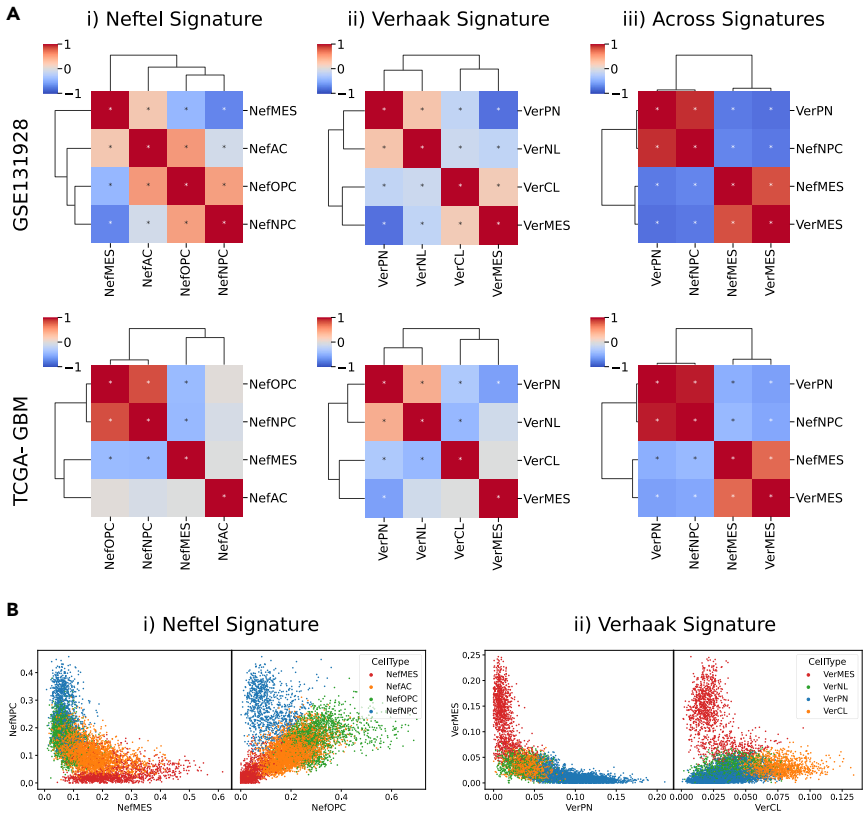


Figure 2. Trends revealed by the Signature scores

(A) Heatmaps visualizing spearman correlation of ssGSEA and AUCell scores for subtypes in the (i) Neftel signature, (ii) Verhaak signature, and (iii) comparison between subtypes across these signatures across (top) GSE131928 and (bottom) TCGA-GBM datasets. The presence of "∗" indicates that the correlation has a p value < 0.05.

(B) Scatterplots showing AUCell scores from the GSE131928 dataset, comparing scores between gene sets that exhibit negative and positive correlations. For the (i) Neftel signature, the comparison is between NPC-like-vs-MES-like and NPC-like-vs-OPC-like, and for the (ii) Verhaak signature, the comparison is between MES-vs-PN and MES-vs-CL.

pairs show positive correlation as well, indicating that they are not distinct (Figure 2A). Particularly, we noticed that the Neftel NPC-like and MES-like signatures are negatively correlated and, hence, antagonistic. Similarly, for the Verhaak signatures, the PN and MES pair are antagonistic. When comparing across the Neftel and Verhaak signatures, we had the following observations: (1) Neftel NPC-like and Verhaak PN are positively correlated with each other, (2) Neftel MES-like and Verhaak MES are positively correlated with one another, and (3) both Neftel NPC and Verhaak PN are negatively correlated with the MES signatures from both the gene set cohorts, further endorsing previous observations about NPC-like and PN may correspond to similar cell-states.[20]

The difference between the pairs of states that are negatively and positively correlated is further illustrated through the scatter between the pairs of scores colored by the maximum score in the GSE131928 dataset (Figure 2B). The cells with high PN/NPC enrichment tend to have very low or no expression of MES genes. Interestingly, the cells with high enrichment of AC-like or OPC-like in Neftel signatures and CL or NL in Verhaak signatures lie in between the gradient of scores in these axes. On the other hand, the cells that are highly enriched for OPC-like tend to have comparable expression of NPC-like as well. This trend is observed for the other pairs as well (Figure S1A) and in the TCGA-GBM dataset (Figure S1B).

Together, our results indicate that across the two datasets used for identifying the two sets of four subtypes,[11,20] the proneural (PN) (or equivalently NPC-like) and mesenchymal (or equivalently MES-like) show strongest antagonistic trends, despite little overlap in gene signatures of PN and NPC-like or MES and MES-like.

NPC/PN-MES antagonism is prevalent at the gene level as well

After investigating pairwise correlations at the gene signature level, we then focused on the gene level to confirm that our observations were not an artifact of the gene set enrichment scoring methods. Thus, we calculated pairwise correlation among all genes involved in the four Neftel gene sets together and noticed that most MES genes correlated positively with one another, most NPC genes correlated positively with each other, but most MES genes were antagonistic to most NPC ones (Figure 3A). Similar observations were made for Verhaak signatures

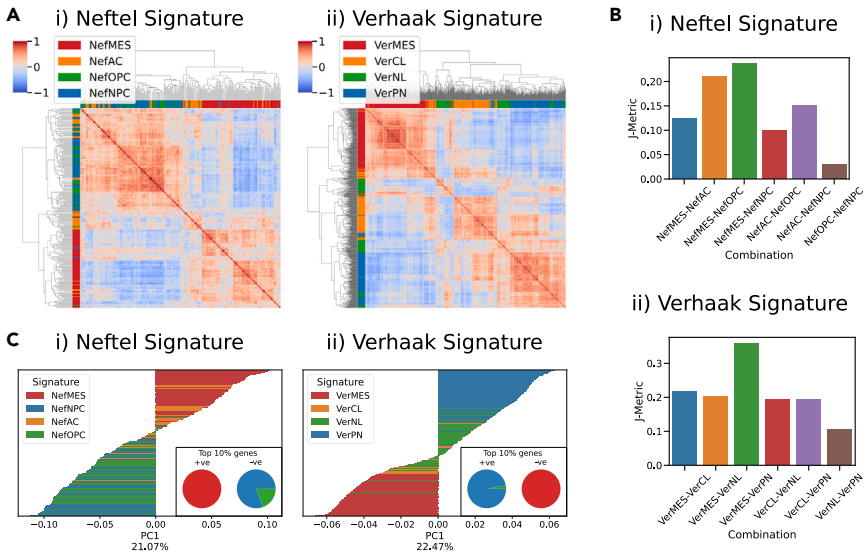


Figure 3. Analysis for antagonism of signature genes in the TCGA-GBM dataset

(A) Heatmaps visualizing spearman correlation of gene expression levels in the (i) Neftel signature and (ii) Verhaak signature. The colors beside the heatmap represent the gene set to which each gene belongs.

(B) J-Metric values for various combinations of subtypes for (i) Neftel signature and (ii) Verhaak signature.

(C) PC1 loadings of PCA based on the gene expression levels in the (i) Neftel signature and (ii) Verhaak signature. The inset shows the gene set to which the top 10% of genes belong for both positive and negative loadings.

as well – where genes from PN, MES, and CL clustered separately and for both the source datasets – GSE131928 and TCGA-GBM (Figures 3A and S2A). These observations suggest that antagonism seen among the enrichment scores of MES vs. NPC or PN emerges from coordinated expression profiles of genes showing a negative correlation.

To quantify these associations, we define J-metric (Equation 2) that quantifies the extent of positive correlation among genes in the same gene set and extent of negative correlation among those across gene sets.[23] J-metric calculations further support our observations that the MES-NPC or the MES-PN gene sets are the most antagonistic ones among all six pairwise correlations (Figures 3B and S2B) (Neftel: MES-AC, MES-OPC, MES-NPC, AC-OPC, AC-NPC, OPC-NPC, and Verhaak: MES-CL, MES-NL, MES-PN, CL-NL, CL-PN, NL-PN).

Next, we interrogated whether this antagonistic relationship also corresponds to the highest variance. To investigate this, we conducted Principal component analysis (PCA) on the gene expression levels within each signature. The first principal component (PC1) represents the axis that captures the greatest variance. Looking at the PC1 loadings for TCGA and GSE131928 (Figures 3C and S2C), we noticed that for the Neftel signatures, MES genes dominate on loadings with positive PC1 coefficient and NPC genes on the negative loadings. Similarly, for the Verhaak signatures, the PN genes dominate on the positive loadings and the MES genes on the negative loadings. Together, these results suggest that the MES- NPC/PN antagonism seen at signature levels percolates to individual genes, and among all the six possible pairwise analyses, MES and NPC (or PN) genes show distinct mutually exclusive patterns.

Meta-analysis across multiple transcriptomic datasets reveal the functional consequences of NPC/PN-MES antagonism

To test for the generalizability of our results, we investigated if the trends we observe hold up across multiple transcriptomic datasets. We examine 80 bulk RNA-sequencing datasets with samples spanning across patient tumor samples, mouse models and cell-lines to capture the maximum variability. For datasets with samples across different organisms (mouse models, patient samples) and contexts (*in vitro*, *in vivo*), we segregated them accordingly and combined the samples, totaling to 90 consolidated samples across all datasets. In each consolidated sample, we calculated the correlation coefficient between the scores of each pair of subtypes. For the correlations between Neftel MES-like and Neftel NPC-like, out of the 38 samples that show a significant correlation, 33 samples (86.8%) are negatively correlated, with just 5 being positive. This pair shows the most bias for being negatively correlated among all the possible combinations, with a few of the other pairs showing bias toward being positively correlated (Figure 4A). Similarly, for the correlations between Verhaak MES and Verhaak PN, of the 48 samples that show a significant correlation, 45 (93.8%) of them are negatively correlated, with just 3 being positive. This pair shows the most bias for negatively correlated among all possible pairwise comparisons.

However, when analyzing bulk transcriptomic datasets, positive correlations could also arise due to the co-occurrence of those subtypes in a tumor and thus may not reflect their genuine transcriptomic associations. To address this, we have looked at 26 scRNA-seq datasets, with 30 consolidated samples as well. Again, here we see the Neftel MES and Neftel NPC pair to have the most bias for negatively correlated of 13 (86.7%) out of 15 significantly correlated, which is the highest bias for negatively correlated (Figure 4B). For the Verhaak signature, the MES-PN pair have the most bias for negatively correlated of 23 (100%) out of 23 significantly correlated.

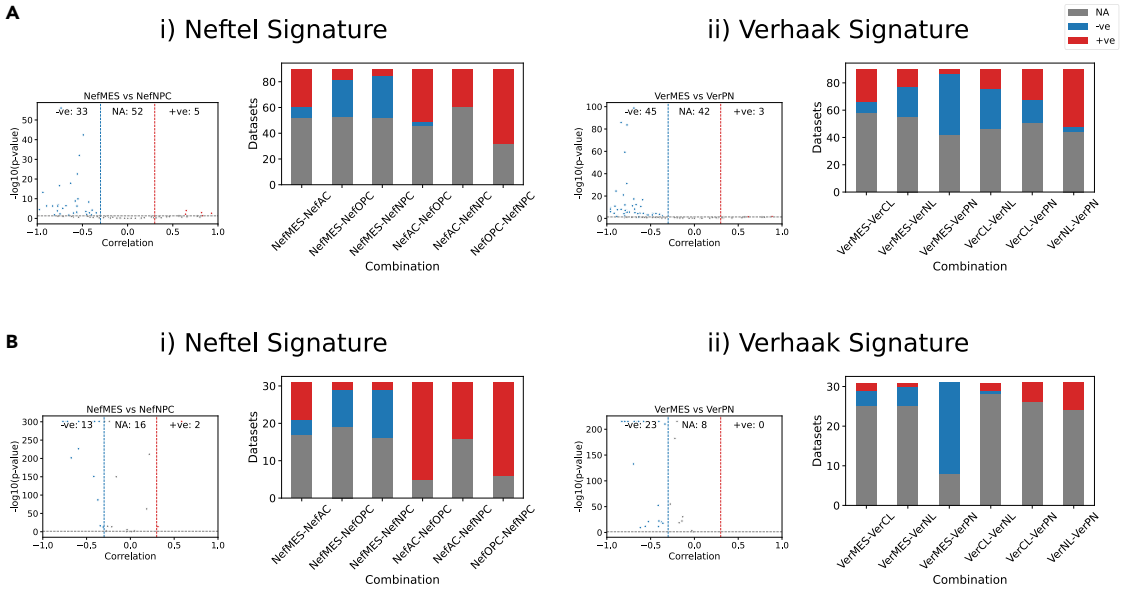


Figure 4. Meta-analysis for trends in correlation of subtypes

(A) Analysis of bulk RNA-seq datasets.

(B) Analysis of scRNA-seq datasets. The left panel displays the volcano plot of correlation between NPC-vs-MES and PN-vs-MES for (i) Neftel Signature and (ii) Verhaak Signature. A dataset with correlation coefficient < -0.3 and p value < 0.05 is categorised as negatively correlated, with correlation coefficient > 0.3 and p value < 0.05 is categorised as positively correlated and NA otherwise. The right panel displays the stacked bar plot of these counts for correlation between different combinations of subtypes.

Next, we checked if the antagonism between MES-NPC/PN has any functional relevance. To characterize the immune infiltration, we looked at the correlation of PD-L1 signature scores with the subtype specific scores. With the bulk datasets, Neftel NPC and Verhaak PN showed the highest negative correlation with PD-L1 of 33 out of 34 (97%) and 35 out of 38 (92.1%) respectively. Meanwhile, the Neftel and Verhaak MES showed the highest positive correlation with PD-L1 of 49 out of 55 (89.1%) and 55 out of 60 (91.7%) respectively (Figure 5A). To characterize the metabolic activity, we looked at the correlation of Glycolysis signature scores with the subtype specific scores. With the bulk datasets, Neftel NPC and Verhaak PN showed the highest negative correlation of 24 out of 39 (61.5%) and 29 out of 42 (69%) respectively, while, the Neftel and Verhaak MES showed the highest positive correlation of 60 out of 61 (98.4%) and 57 out of 60 (95%) respectively (Figure 5B). Similarly, for proliferative activity, we looked at the correlation with KEGG cell cycle signature score. With the bulk datasets, Neftel NPC and Verhaak PN showed the highest positive correlation of 18 out of 28 (64.3%) and 34 out of 40 (85%) respectively, while, the Neftel and Verhaak MES showed the highest negative correlation of 19 out of 38 (50%) and 24 out of 60 (63%) respectively (Figure S3A). Overall, when analyzing individual datasets, Neftel NPC and Verhaak PN showed consistently negative correlation with PD-L1 signature scores in a majority of datasets, in contrast to the trends seen for Neftel and Verhaak MES signatures (Figure 5C). With the glycolysis signatures, Neftel and Verhaak MES signatures predominantly exhibited a positive correlation, while both the Neftel NPC and Verhaak PN were mostly negatively correlated (Figure 5). However, the Neftel OPC, Neftel AC, Verhaak CL and Verhaak NL signatures did not show strong patterns.

Similar trends in general are observed with the single cell datasets as well, however, these should be treated with caution given the low number (< 10) of cases showing significant correlation. Generally, the MES signatures are more negatively correlated with KEGG cell cycle (Figure S3B), and, positively with PD-L1 (Figure S3C) and glycolysis (Figure S3D) while, NPC/PN are positively with KEGG cell cycle and negatively with PD-L1 and glycolysis. Together these results points toward NPC/PN and MES occupying opposing ends of the functional spectrum.

NPC/PN - MES associated transcription factors capture clinical significance

Next, we looked to identify transcription factors (TFs) associated with the NPC/PN and MES states, respectively. We looked at the correlation of all human TFs[24] with the enrichment scores for Neftel NPC and Neftel MES, Verhaak PN, and Verhaak MES signatures in scRNA-seq data (GSE131928). Most TFs are positively correlated with NPC/PN and negatively correlated with MES or positively correlated with MES and negatively correlated with NPC/PN (Figure 6A). While some TFs do not follow this trend, their correlation with either MES or NPC/PN enrichment scores is not strong (<0.2) to be considered. The top 10 TFs that are highly positively correlated with MES lie on the extreme end of the second quadrant, and similarly, the top 10 TFs that are highly positively correlated with NPC/PN lie on the extreme end of the fourth quadrant. These top TFs are also common for the similar states between the Neftel and Verhaak signatures (Table S3). This TF list includes RUNX1, FOSL2, and BHLHE40 that have been previously reported to associate with a mesenchymal state,[25–27] and TCF4, TCF12, and MYT1 are known to drive a proneural state.[28–30] When the expression levels of these TFs were correlated in a pairwise manner, we noticed "teams-like" behavior[23,31,32]

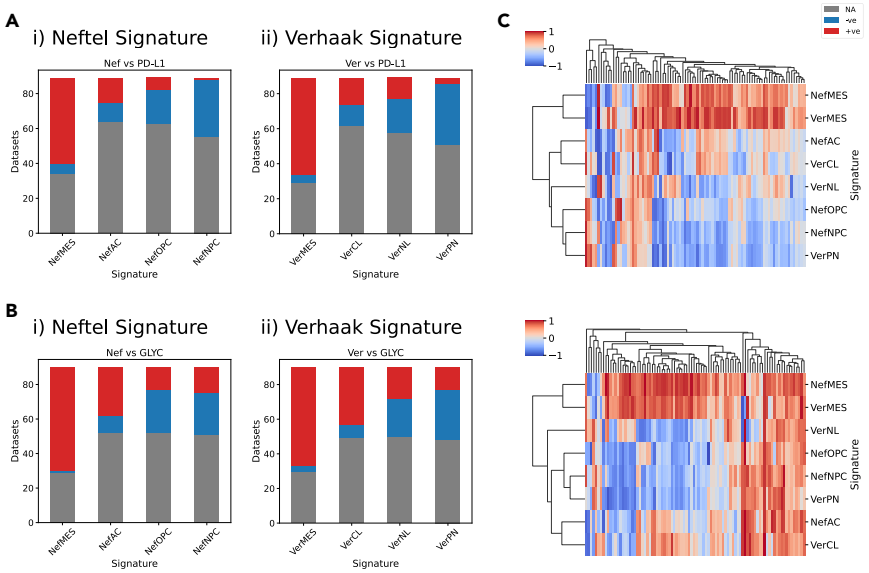


Figure 5. Meta-analysis for trends in correlation with immune and metabolic signatures for bulk datasets

(A) Counts of correlation of subtypes with PD-L1 Signature score.

(B) Counts of correlation of subtypes with Hallmark Glycolysis score. The stacked bar plot gives the counts for correlation between different subtypes of the (i) Neftel Signature and (ii) Verhaak Signature. A dataset with correlation coefficient < −0.3 and p value < 0.05 is categorized as negatively correlated, with correlation coefficient > 0.3 and p value < 0.05 is categorized as positively correlated and NA otherwise.

(C) Heatmap of correlation of subtypes with (top) PD-L1 and (bottom) Glycolysis signature scores. Each column represents a dataset, and the rows represent each subtype.

such that TFs that positively correlate with Neftel or Verhaak MES signatures also are positively correlated among themselves, while correlated negatively with the TFs that correlate positively with Verhaak PN or Neftel NPC signatures (Figure 6B), reminiscent of the antagonism we observed above between the genes in the signatures (Figure 3A). These trends were recapitulated in bulk RNA-seq data for TCGA as well (Figures S4A and S4B).

Finally, we assessed the clinical significance of these TFs by performing a survival analysis on the TCGA GBM patient data. We noticed that for progression-free interval (PFI), higher expression of TFs associated with a proneural state often exhibited better prognosis (Figure 6C). On the contrary, the enrichment of TFs associated with a mesenchymal state showed a worse prognosis. Similar trends were observed for overall survival (Figure S4C); particularly, TCF12 and CXXC4 were associated with better survival (HR = 0.626, $p < 0.001$ for TCF12; HR = 0.684, $p < 0.05$ for CXXC4), while RUNX1 and BHLHE40 were associated with worse survival (HR = 1.42, $p < 0.05$ for RUNX1; HR = 1.63, $p < 0.001$ for BHLHE40) (Figure 6D). Together, this analysis helps identify putative transcriptional drivers and repressors of proneural-mesenchymal transition (PMT) and their clinical significance.

DISCUSSION

Our in-depth analysis underscores that not all four proposed GBM subtypes can be unequivocally categorized as distinct entities. It is apparent that, within this classification, some subtypes exhibit a level of similarity that prevents us from asserting their distinctiveness. However, we noticed consistently that the proneural/mesenchymal axis (as per Verhaak et al.[11] classification) or neural progenitor-like/mesenchymal-like axis (as per Neftel et al.[20] classification) are the most notably antagonistic pairs. This antagonistic relationship, together with their inverse relationships with functional attributes (cell cycle, immune evasion, metabolism) as identified via our meta-analysis, implies a clear demarcation between these two subtypes, rendering them genuinely distinct states.

Our results are reminiscent of previous single-cell/single-nucleus RNA-seq observations of a cohort of primary tumors showing that phenotypes of proliferating GBM cells lie along a single axis of variation ranging from proneural to mesenchymal.[33] Diverse biological phenotypes being mostly explained by principal component 1 (PC1) axis has also been observed for other instances of cancer cell plasticity: epithelial-mesenchymal transition in carcinomas, and proliferative-invasive switch in melanomas,[34] suggesting this feature to be a more generic occurrence than only in GBM. Transcriptional variability along the proneural-mesenchymal axis also impacts drug sensitivity,[35] similar to the association of EMT with resistance to chemotherapy, targeted therapy, and immune checkpoint blockade therapy.[3,36,37] Consistently, recurrent GBMs are enriched in a mesenchymal state.[10] This variability also has implication in the development of targeted therapy. For example, a drug aimed at the NPC-like state can also impact cells in the OPC-like state, given their transcriptional similarity. Conversely, the MES-like state is expected to be refractory to the drug targeting NPC-like state. Thus, combinatorial strategies targeting the MES-like and NPC-like states individually are likely to be more effective than those targeting only one end of the spectrum.

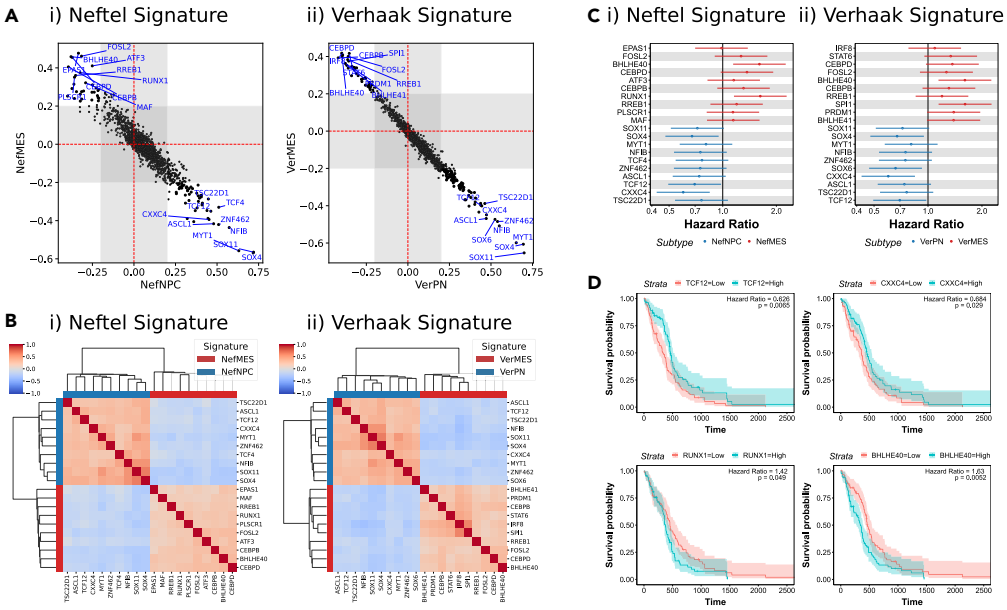


Figure 6. Transcription factors associated with the NPC/PN-MES antagonism in GSE131928 and their clinical implication

(A) Scatterplot of correlation of TFs with MES vs. correlation with NPC/PN. The gray background is for correlation between −0.2 and 0.2. The highly correlated TFs are annotated.

(B) Pairwise correlation of the TFs highly correlated with NPC/PN and MES.

(C) Hazard ratios of the high to low expression of TFs for progression-free interval of TCGA-GBM patients. A filled dot represents the hazard ratio with a p value < 0.05.

(D) Kaplan-Meier plots of the overall survival probability of TCGA-GBM patients by expression levels of TCF12, CXXC4, RUNX1, and BHLHE40. The top row is associated with NPC/PN, and the bottom row is associated with MES.

Our meta-analysis suggests that mesenchymal subtype is relatively enriched in glycolytic traits, a behavior validated in metabolic differences between proneural and mesenchymal tumor-initiating cells.[38,39] The observed association of PD-L1 with mesenchymal subgroup in GBM[40–42] also reinforces our analysis; functionally, PD-L1 can trigger more aggressive GBM behavior through the downstream Ras/Erk signaling. The independent association of both PD-L1 and glycolysis with worse patient outcomes in GBM[41,43] suggests more aggressive behavior of MES phenotype, given the enrichment of both glycolysis and PD-L1 levels in MES subtype. Concurrent enrichment of glycolysis and PD-L1 was found to be associated with worse overall survival across multiple cancer types as well.[44] The NPC/PN-MES antagonism is also reflected in our clinical data analysis, where TFs associated with MES phenotype exhibited an association with worse survival, while those driving an NPC/PN state associated with better survival trends.

Our results establish important functional differences between proneural and mesenchymal states and suggest rethinking GBM phenotypic classification. Thus, future efforts to characterize GBM phenotypic plasticity and heterogeneity should incorporate both genetic and non-genetic (transcriptional, epigenetic, metabolic) components, given increasing evidence about the interplay of both components in cancer cell adaptation to therapeutic attacks.[45–47] Further, the mapping of underlying gene regulatory networks enabling phenotypic heterogeneity – as done in EMT, neuroendocrine differentiation in small cell lung cancer, and phenotypic switching in melanoma[32,48–51] - is still in its infancy in GBM.[52,53] A hallmark of these networks across the cancer types is the presence of "teams" of mutually inhibitory players that can allow for co-existence of cell-states in a population and most variability in phenotypic space residing along the PC1 axis.[23,32,34,54] Whether such "teams" exist for proneural vs. mesenchymal phenotypes remains elusive as of yet, although our preliminary analysis of TFs supports this hypothesis.

Overall, the results presented here across bulk and single-cell transcriptomics demonstrate that proneural-mesenchymal axis dominates the patterns of phenotypic heterogeneity in GBM, and cells lying along this axis have varied functional traits in terms of metabolism and immune-evasion. Such extensive meta-analysis conducted for over 100 datasets offers an unbiased view of GBM heterogeneity, and highlights clinical implications of its accurate classification at both diagnostic and therapeutic aspects. Moreover, it can be a first key step to narrow down on the molecular mechanisms shaping these low-dimensional phenotypic heterogeneity patterns.

Limitations of the study

While our meta-analysis across multiple datasets demonstrates mutual exclusivity of proneural and mesenchymal antagonism in GBM, it suffers from several limitations. First, we have not identified underlying regulatory networks that can explain such antagonism and phenotypic plasticity. Second, our work only delves into transcriptional level interactions, and does not capture other modes of regulation - translation,

epigenetic, metabolic etc. Third, our work does not explain spatial aspects of phenotypic heterogeneity in GBM. Future efforts can focus on mapping underlying regulatory networks that can explain heterogeneity patterns.

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
- METHOD DETAILS
  - Datasets and preprocessing
  - Scoring of signatures
  - Correlation of gene expression
  - PCA of gene expression
  - Survival analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109184.

ACKNOWLEDGMENTS

H.B.V. is supported by the Prime Minister's Research Fellowship (PMRF). M.K.J. was supported by Ramanujan Fellowship (SB/S2/RJN-049/2018) by the Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, and by Param Hansa Philanthropies. Atchuta Srinivas Duddu is acknowledged for the artwork of graphical abstract.

AUTHOR CONTRIBUTIONS

M.K.J. conceptualized and designed the study. H.B.V. performed the analysis. H.B.V. and M.K.J. wrote and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used ChatGPT to enhance the clarity and coherence of parts of the discussion section. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Received: November 28, 2023
Revised: December 31, 2023
Accepted: February 6, 2024
Published: February 13, 2024

REFERENCES

1. Dymova, M.A., Kuligina, E.V., and Richter, V.A. (2021). Molecular Mechanisms of Drug Resistance in Glioblastoma. Int. J. Mol. Sci. *22*, 6385. https://doi.org/10.3390/ijms22126385.
2. Yabo, Y.A., Niclou, S.P., and Golebiewska, A. (2022). Cancer cell heterogeneity and plasticity: A paradigm shift in glioblastoma. Neuro. Oncol. *24*, 669–682. https://doi.org/10.1093/neuonc/noab269.
3. Sahoo, S., Mishra, A., Kaur, H., Hari, K., Muralidharan, S., Mandal, S., and Jolly, M.K. (2021). A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer *3*, zcab027. https://doi.org/10.1093/narcan/zcab027.
4. Oren, Y., Tsabar, M., Cuoco, M.S., Amir-Zilberstein, L., Cabanos, H.F., Hütter, J.C., Hu, B., Thakore, P.I., Tabaka, M., Fulco, C.P., et al. (2021). Cycling cancer persister cells arise from lineages with distinct programs. Nature *596*, 576–582. https://doi.org/10.1038/s41586-021-03796-6.
5. Ramirez, M., Rajaram, S., Steininger, R.J., Osipchuk, D., Roth, M.A., Morinishi, L.S., Evans, L., Ji, W., Hsu, C.-H., Thurley, K., et al. (2016). Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. Nat. Commun. *7*, 10690. https://doi.org/10.1038/ncomms10690.
6. Goyal, Y., Busch, G.T., Pillai, M., Li, J., Boe, R.H., Grody, E.I., Chelvanambi, M., Dardani, I.P., Emert, B., Bodkin, N., et al. (2023). Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. Nature *620*, 651–659. https://doi.org/10.1038/s41586-023-06342-8.
7. Subhadarshini, S., Sahoo, S., Debnath, S., Somarelli, J.A., and Jolly, M.K. (2023). Dynamical modeling of proliferative-invasive plasticity and IFNγ signaling in melanoma reveals mechanisms of PD-L1 expression heterogeneity. J Immunother Cancer *11*, e006766. https://doi.org/10.1136/jitc-2023-006766.

8. Su, Y., Wei, W., Robert, L., Xue, M., Tsoi, J., Garcia-Diaz, A., Homet Moreno, B., Kim, J., Ng, R.H., Lee, J.W., et al. (2017). Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc. Natl. Acad. Sci. USA *114*, 13679–13684. https://doi.org/10.1073/pnas.1712064115.
9. Stepanenko, A.A., Andreieva, S.V., Korets, K.V., Mykytenko, D.O., Baklaushev, V.P., Huleyuk, N.L., Kovalova, O.A., Kotsarenko, K.V., Chekhonin, V.P., Vassetzky, Y.S., et al. (2016). Temozolomide promotes genomic and phenotypic changes in glioblastoma cells. Cancer Cell Int. *16*, 36. https://doi.org/10.1186/s12935-016-0311-8.
10. 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.
11. Verhaak, R.G.W., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell *17*, 98–110. https://doi.org/10.1016/j.ccr.2009.12.020.
12. Wang, Q., Hu, B., Hu, X., Kim, H., Squatrito, M., Scarpace, L., deCarvalho, A.C., Lyu, S., Li, P., Li, Y., et al. (2017). Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell *32*, 42–56.e6. https://doi.org/10.1016/j.ccell.2017.06.003.
13. Gill, B.J., Pisapia, D.J., Malone, H.R., Goldstein, H., Lei, L., Sonabend, A., Yun, J., Samanamud, J., Sims, J.S., Banu, M., et al. (2014). MRI-localized biopsies reveal subtype-specific differences in molecular and cellular composition at the margins of glioblastoma. Proc. Natl. Acad. Sci. USA *111*, 12550–12555. https://doi.org/10.1073/pnas.1405839111.
14. 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.
15. Sottoriva, A., Spiteri, I., Piccirillo, S.G.M., Touloumis, A., Collins, V.P., Marioni, J.C., Curtis, C., Watts, C., and Tavaré, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc. Natl. Acad. Sci. USA *110*, 4009–4014. https://doi.org/10.1073/pnas.1219747110.
16. Pillai, M., Chen, Z., Jolly, M.K., and Li, C. (2022). Quantitative landscapes reveal trajectories of cell-state transitions associated with drug resistance in melanoma. iScience *25*, 105499. https://doi.org/10.1016/j.isci.2022.105499.
17. Brown, M.S., Abdollahi, B., Wilkins, O.M., Lu, H., Chakraborty, P., Ognjenovic, N.B., Muller, K.E., Jolly, M.K., Christensen, B.C., Hassanpour, S., and Pattabiraman, D.R. (2022). Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. Sci. Adv. *8*, eabj8002. https://doi.org/10.1126/sciadv.abj8002.
18. Bocci, F., Gearhart-Serna, L., Boareto, M., Ribeiro, M., Ben-Jacob, E., Devi, G.R., Levine, H., Onuchic, J.N., and Jolly, M.K. (2019). Toward understanding cancer stem cell heterogeneity in the tumor microenvironment. Proc. Natl. Acad. Sci. USA *116*, 148–157. https://doi.org/10.1073/pnas.1815345116.
19. Malagoli Tagliazucchi, G., Wiecek, A.J., Withnell, E., and Secrier, M. (2023). Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. Nat. Commun. *14*, 789. https://doi.org/10.1038/s41467-023-36439-7.
20. 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.
21. Hoek, K.S., Eichhoff, O.M., Schlegel, N.C., Döbbeling, U., Kobert, N., Schaerer, L., Hemmi, S., and Dummer, R. (2008). In vivo Switching of Human Melanoma Cells between Proliferative and Invasive States. Cancer Res. *68*, 650–656. https://doi.org/10.1158/0008-5472.CAN-07-2491.
22. Zhang, W., Girard, L., Zhang, Y.-A., Haruki, T., Papari-Zareei, M., Stastny, V., Ghayee, H.K., Pacak, K., Oliver, T.G., Minna, J.D., and Gazdar, A.F. (2018). Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. Transl. Lung Cancer Res. *7*, 32–49. https://doi.org/10.21037/tlcr.2018.02.02.
23. Chauhan, L., Ram, U., Hari, K., and Jolly, M.K. (2021). Topological signatures in regulatory network enable phenotypic heterogeneity in small cell lung cancer. Elife *10*, e64522. https://doi.org/10.7554/eLife.64522.
24. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.
25. Zhao, K., Cui, X., Wang, Q., Fang, C., Tan, Y., Wang, Y., Yi, K., Yang, C., You, H., Shang, R., et al. (2019). RUNX1 contributes to the mesenchymal subtype of glioblastoma in a TGFβ pathway-dependent manner. Cell Death Dis. *10*, 877. https://doi.org/10.1038/s41419-019-2108-x.
26. Qiu, W., Xiao, Z., Yang, Y., Jiang, L., Song, S., Qi, X., Chen, Y., Yang, H., Liu, J., and Chu, L. (2023). USP10 deubiquitinates RUNX1 and promotes proneural-to-mesenchymal transition in glioblastoma. Cell Death Dis. *14*, 207. https://doi.org/10.1038/s41419-023-05734-y.
27. Cooper, L.A.D., Gutman, D.A., Chisolm, C., Appin, C., Kong, J., Rong, Y., Kurc, T., Van Meir, E.G., Saltz, J.H., Moreno, C.S., and Brat, D.J. (2012). The tumor microenvironment strongly impacts master transcriptional regulators and gene expression class of glioblastoma. AmAm. J. Pathol. *180*, 2108–2119. https://doi.org/10.1016/j.ajpath.2012.01.040.
28. Zhang, J.-X., Zhang, J., Yan, W., Wang, Y.-Y., Han, L., Yue, X., Liu, N., You, Y.-P., Jiang, T., Pu, P.-Y., and Kang, C.-S. (2013). Unique genome-wide map of TCF4 and STAT3 targets using ChIP-seq reveals their association with new molecular subtypes of glioblastoma. Neuro. Oncol. *15*, 279–289. https://doi.org/10.1093/neuonc/nos306.
29. Zhu, G., Yang, S., Wang, R., Lei, J., Ji, P., Wang, J., Tao, K., Yang, C., Ge, S., and Wang, L. (2021). P53/miR-154 Pathway Regulates the Epithelial-Mesenchymal Transition in Glioblastoma Multiforme Cells by Targeting TCF12. Neuropsychiatr. Dis. Treat. *17*, 681–693. https://doi.org/10.2147/NDT.S273578.
30. Melhuish, T.A., Kowalczyk, I., Manukyan, A., Zhang, Y., Shah, A., Abounader, R., and Wotton, D. (2018). Myt1 and Myt1l transcription factors limit proliferation in GBM cells by repressing YAP1 expression. Biochim Biophys Acta Gene Regul. Mech. *1861*, 983–995. https://doi.org/10.1016/j.bbagrm.2018.10.005.
31. Sehgal, M., Nayak, S.P., Sahoo, S., Somarelli, J.A., and Jolly, M.K. (2024). Mutually exclusive teams-like patterns of gene regulation characterize phenotypic heterogeneity along the noradrenergic-mesenchymal axis in neuroblastoma. Cancer Biol. Ther. *25*, 2301802. https://doi.org/10.1080/15384047.2024.2301802.
32. Pillai, M., and Jolly, M.K. (2021). Systems-level network modeling deciphers the master regulators of phenotypic plasticity and heterogeneity in melanoma. iScience *24*, 103111. https://doi.org/10.1016/j.isci.2021.103111.
33. Wang, L., Babikir, H., Müller, S., Yagnik, G., Shamardani, K., Catalan, F., Kohanbash, G., Alvarado, B., Di Lullo, E., Kriegstein, A., et al. (2019). The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. Cancer Discov. *9*, 1708–1719. https://doi.org/10.1158/2159-8290.CD-19-0329.
34. Hari, K., Harlapur, P., Saxena, A., Haldar, K., Girish, A., Malpani, T., Levine, H., and Jolly, M.K. (2023). Low dimensionality of phenotypic space as an emergent property of coordinated teams in biological regulatory networks. Preprint at bioRxiv. https://doi.org/10.1101/2023.02.03.526930.
35. Meyer, M., Reimand, J., Lan, X., Head, R., Zhu, X., Kushida, M., Bayani, J., Pressey, J.C., Lionel, A.C., Clarke, I.D., et al. (2015). Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. Proc. Natl. Acad. Sci. USA *112*, 851–856. https://doi.org/10.1073/pnas.1320611111.
36. Dudas, J., Ladanyi, A., Ingruber, J., Steinbichler, T.B., and Riechelmann, H. (2020). Epithelial to Mesenchymal Transition: A Mechanism that Fuels Cancer Radio/Chemoresistance. Cells *9*, 428. https://doi.org/10.3390/cells9020428.
37. Gu, Y., Zhang, Z., and ten Dijke, P. (2023). Harnessing epithelial-mesenchymal plasticity to boost cancer immunotherapy. Cell. Mol. Immunol. *20*, 318–340. https://doi.org/10.1038/s41423-023-00980-8.
38. Seliger, C., Meyer, A.-L., Leidgens, V., Rauer, L., Moeckel, S., Jachnik, B., Proske, J., Dettmer, K., Rothhammer-Hampl, T., Kaulen, L.D., et al. (2022). Metabolic Heterogeneity of Brain Tumor Cells of Proneural and Mesenchymal Origin. Int. J. Mol. Sci. *23*, 11629. https://doi.org/10.3390/ijms231911629.
39. Mao, P., Joshi, K., Li, J., Kim, S.-H., Li, P., Santana-Santos, L., Luthra, S., Chandran, U.R., Benos, P.V., Smith, L., et al. (2013). Mesenchymal glioma stem cells are maintained by activated glycolytic

metabolism involving aldehyde dehydrogenase 1A3. Proc. Natl. Acad. Sci. USA *110*, 8644–8649. https://doi.org/10.1073/pnas.1221478110.

40. Ricklefs, F.L., Alayo, Q., Krenzlin, H., Mahmoud, A.B., Speranza, M.C., Nakashima, H., Hayes, J.L., Lee, K., Balaj, L., Passaro, C., et al. (2018). Immune evasion mediated by PD-L1 on glioblastoma-derived extracellular vesicles. Sci. Adv. *4*, eaar2766. https://doi.org/10.1126/sciadv.aar2766.
41. Nduom, E.K., Wei, J., Yaghi, N.K., Huang, N., Kong, L.-Y., Gabrusiewicz, K., Ling, X., Zhou, S., Ivan, C., Chen, J.Q., et al. (2016). PD-L1 expression and prognostic impact in glioblastoma. Neuro. Oncol. *18*, 195–205. https://doi.org/10.1093/neuonc/nov172.
42. Wu, C., Qin, C., Long, W., Wang, X., Xiao, K., and Liu, Q. (2022). Tumor antigens and immune subtypes of glioblastoma: the fundamentals of mRNA vaccine and individualized immunotherapy development. J. Big Data *9*, 92. https://doi.org/10.1186/s40537-022-00643-x.
43. Wang, Z., Zhang, C., Liu, X., Wang, Z., Sun, L., Li, G., Liang, J., Hu, H., Liu, Y., Zhang, W., and Jiang, T. (2016). Molecular and clinical characterization of PD-L1 expression at transcriptional level via 976 samples of brain glioma. OncoImmunology *5*, e1196310. https://doi.org/10.1080/2162402X.2016.1196310.
44. Muralidharan, S., Sehgal, M., Soundharya, R., Mandal, S., Majumdar, S.S., Yeshwanth, M., Saha, A., and Jolly, M.K. (2022). PD-L1 Activity Is Associated with Partial EMT and Metabolic Reprogramming in Carcinomas. Curr. Oncol. *29*, 8285–8301. https://doi.org/10.3390/curroncol29110654.
45. Fernandez-Mateos, J., Milite, S., Oliveira, E., Vlachogiannis, G., Chen, B., Yara, E., Cresswell, G.D., James, C., Patruno, L., Ascolani, G., et al. (2023). Epigenetic heritability of cell plasticity drives cancer drug resistance through one-to-many genotype to phenotype mapping. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.15.567140.
46. Salgia, R., and Kulkarni, P. (2018). The Genetic/Non-genetic Duality of Drug 'Resistance' in Cancer. Trends Cancer *4*, 110–118. https://doi.org/10.1016/j.trecan.2018.01.001.
47. Garofano, L., Migliozzi, S., Oh, Y.T., D'Angelo, F., Najac, R.D., Ko, A., Frangaj, B., Caruso, F.P., Yu, K., Yuan, J., et al. (2021). Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. Nat. Cancer *2*, 141–156. https://doi.org/10.1038/s43018-020-00159-4.
48. Steinway, S.N., Zañudo, J.G.T., Michel, P.J., Feith, D.J., Loughran, T.P., and Albert, R. (2015). Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. NPJ Syst. Biol. Appl. *1*, 15014. https://doi.org/10.1038/npjsba.2015.14.
49. Silveira, D.A., Gupta, S., and Mombach, J.C.M. (2020). Systems biology approach suggests new miRNAs as phenotypic stability factors in the epithelial-mesenchymal transition. J. R. Soc. Interface *17*, 20200693. https://doi.org/10.1098/rsif.2020.0693.
50. Ozen, M., and Lopez, C.F. (2023). Data-driven structural analysis of small cell lung cancer transcription factor network suggests potential subtype regulators and transition pathways. NPJ Syst. Biol. Appl. *9*, 55. https://doi.org/10.1038/s41540-023-00316-2.
51. Udyavar, A.R., Wooten, D.J., Hoeksema, M., Bansal, M., Califano, A., Estrada, L., Schnell, S., Irish, J.M., Massion, P.P., and Quaranta, V. (2017). Novel Hybrid Phenotype Revealed in Small Cell Lung Cancer by a Transcription Factor Network Model That Can Explain Tumor Heterogeneity. Cancer Res. *77*, 1063–1074. https://doi.org/10.1158/0008-5472.CAN-16-1467.
52. Pérez-Aliacar, M., Ayensa-Jiménez, J., Randelović, T., Ochoa, I., and Doblaré, M. (2023). Modelling glioblastoma resistance to temozolomide. Combination of spheroid and mathematical models to simulate cellular adaptation in vitro. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.24.568421.
53. Larsson, I., Dalmo, E., Elgendy, R., Niklasson, M., Doroszko, M., Segerman, A., Jörnsten, R., Westermark, B., and Nelander, S. (2021). Modeling glioblastoma heterogeneity as a dynamic network of cell states. Mol. Syst. Biol. *17*, e10105. https://doi.org/10.15252/msb.202010105.
54. Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A., and Jolly, M.K. (2022). Landscape of epithelial–mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. Elife *11*, e76535. https://doi.org/10.7554/eLife.76535.
55. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757. https://doi.org/10.1093/bioinformatics/btac757.
56. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.
57. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2023). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.
58. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.
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., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
60. Sahoo, S., Nayak, S.P., Hari, K., Purkait, P., Mandal, S., Kishore, A., Levine, H., and Jolly, M.K. (2021). Immunosuppressive Traits of the Hybrid Epithelial/Mesenchymal Phenotype. Front. Immunol. *12*, 797261.
61. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112. https://doi.org/10.1038/nature08460.
62. Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Analyzed Data	Neftel et al.[20]	GEO: GSE131928
Analyzed Data	Verhaak et al.[11]	TCGA: GBM
Software and algorithms		
GSEAPy	Fang et al.[55]	https://github.com/zqfang/GSEApy
AUCell	Aibar et al.[56]	https://bioconductor.org/packages/release/bioc/html/AUCell.html
Seurat	Hao et al.[57]	https://satijalab.org/seurat/
SciPy	Virtanen et al.[58]	https://scipy.org/

RESOURCE AVAILABILITY

Lead contact

Further information should be directed to and will be fulfilled by the lead contact, Mohit Kumar Jolly (mkjolly@iisc.ac.in).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table and Table S1.
- The scripts used for analysis and the processed data are available at: https://github.com/Harshavardhan-BV/GBM_4states.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

METHOD DETAILS

Datasets and preprocessing

All the datasets are publicly available and given in Table S1 for single-cell and bulk RNA sequencing datasets. The datasets with the GSE IDs are available and downloaded from the Gene Expression Omnibus (GEO) Website. Other datasets such as The Cancer Genome Atlas (TCGA), Chinese Glioma Genome Atlas (CGGA), Glioma Longitudinal AnalySiS (GLASS), Cancer Cell Line Encyclopedia (CCLE), QCell were downloaded from their respective website.

We consolidated all the samples from the same datasets into a single count matrix, except for instances where there were variations in the model system or sequencing platform, under which circumstances they were treated separately. All the bulk datasets were normalised to log transcripts per million ($log_2(TPM)$). For single-cell datasets, only the tumor cells were selected when metadata was available with this information. Quality control was done by removing outliers, identified as cells with gene counts exhibiting deviations beyond −5 and +5 mean - average deviations (MAD) or based on mitochondrial gene percentage beyond +3 MAD. Subsequently, the datasets were normalized using log normalization.

Scoring of signatures

The Verhaak signatures and the pathways signatures for Hallmark Glycolysis and KEGG cell-cyle were obtained from MSigDB.[59] The Neftel signatures and signature for PD-L1 were obtained from previous reports.[20,60] These signatures are given in Table S2.

The scoring of bulk datasets were done using ssGSEA[61] given by:

$$ES(G,S) = \sum_{i=1}^{N}\left[\underbrace{\sum_{r\in G, j\le i}\frac{|r_j|^{(1/4)}}{\sum_{r_j\in G}|r_j|^{(1/4)}}}_{\text{ECDF of genes in signature}} - \underbrace{\sum_{r\notin G, j\le i}\frac{1}{N-N_G}}_{\text{ECDF of background genes}}\right] \quad \text{(Equation 1)}$$

Where, G_i = Gene Set i, N_i = Number of genes in set i, r_j = Rank of gene j

The implementation of ssGSEA in GSEAPy[55] was used and the normalised expression scores were used for further analysis. The scoring of the single-cell datasets were done using AUCell from the SCENIC package in R.[56] Pairwise correlations between the scores were done with spearmanr function in SciPy.

Correlation of gene expression

The expression levels of the genes for a particular signature were selected. The pairwise spearman correlations were computed using the cor function in R. These correlations were then used to group together genes with similar expression patterns through hierarchical clustering, which was accomplished using the clustermap function in seaborn.

To quantify the antagonism between the different genesets, the J-Metric[23] was used given by:

$$J = \underbrace{\sum_{x,y \in G_1} \frac{\rho_r(x,y)}{4N_1^2} + \sum_{x,y \in G_2} \frac{\rho_r(x,y)}{4N_2^2}}_{\text{Correlation within genesets}} - \underbrace{\sum_{x \in G_1, y \in G_2} \frac{\rho_r(x,y)}{2N_1N_2}}_{\text{Correlation across genesets}} \quad \text{(Equation 2)}$$

Where, G_i = Gene Set i, N_i = Number of genes in set i, $\rho_r(x,y)$ = Spearman correlation of gene x with gene y

A higher value of J-metric corresponds to more antagonism between the genesets. A value closer to 1 indicates that the within geneset correlation is more positive and the across geneset correlation is more negative. Whereas, a value closer to 0 indicates the correlation within geneset and across geneset to be of similar levels.

PCA of gene expression

Principal component analysis (PCA) was done on the expression levels of all the genes for each signature using prcomp function in R. Subsequently, the loadings of the first PC1 were visualized in a bar plot, sorted by their values. To further analyze these results, the top 10% of genes were selected based on their loadings in PC1, both in the positive and negative directions. The counts of genes belonging to each geneset were then quantified.

Survival analysis

The curated survival data for TCGA-GBM patients were obtained from UCSC Xena.[62] The top transcription factors correlated with NPC and MES from Neftel signatures, and with PN and MES for the Verhaak signatures were selected. Patients were stratified into high-expression and low-expression groups based on the median gene expression for each of these transcription factors. The cox-regression model was fit using the coxph function from survival to the progression-free interval and overall survival with the low-expression group set as the baseline for each gene. The hazard ratios from this model are visualised as a forest plot for each gene. Kaplan-Meier (KM) analysis was done to the overall survival for select genes using the survfit function.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis were performed using R version 4.3.2, and using Python version 3.10.12 with SciPy version 1.11.4 as mentioned in the method details.