CHROMATIN BLUEPRINT OF GLIOBLASTOMA STEM CELLS REVEALS COMMON DRUG CANDIDATES FOR DISTINCT SUBTYPES

Paul Guilhamon[1], Michelle M Kushida[2], Graham MacLeod[3], Seyed AM Tonekaboni[1,4], Florence MG Cavalli[2], Fiona J Coutinho[2], Nishani Rajakulendran[3], Xinghui Che[2], Naghmeh Rastegar[2], Mona Meyer[2], Xiaoyang Lan[2], Nuno M Nunes[5], Uri Tabori[6], Michael D Taylor[2], Benjamin Haibe-Kains[1,4,7,8], Stephane Angers[3], Peter B Dirks[2,8,*], Mathieu Lupien[1,4,9,*]

[1]*Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada*
[2]*Developmental and Stem Cell Biology Program and Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada*
[3]*Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada*
[4]*Department of Medical Biophysics, University of Toronto, Toronto, ON M5S 1A8, Canada*
[5]*Program in Genetics and Genome Biology, Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada*
[6]*Division of Hematology and Oncology, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada*
[7]*Department of Computer Science, University of Toronto, Toronto, ON M5S 1A8, Canada*
[8]*Departments of Molecular Genetics and Surgery, University of Toronto, Toronto, ON M5S 1A8, Canada*
[9]*Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada*

*Corresponding authors: Mathieu Lupien (mlupien@uhnres.utoronto.ca) and Peter B Dirks (peter.dirks@sickkids.ca)

Summary

Glioblastoma (GBM) is an aggressive form of brain cancer with a median survival of 12.6 months[1] and for which the standard treatment of surgery, radiotherapy and temozolomide, provides only an additional 2.5 months in the small subset of responsive patients[2]. Despite extensive characterization and stratification of the bulk primary tumours, no targeted therapies have been successfully developed[1,3].

GBM tumours are rooted in glioblastoma stem cells (GSCs) that have self-renewal and tumour-initiating capacities[4]. GSCs also drive disease progression in vivo[5,6]. Although the mutational landscape of GSCs is well established[7,8], and their epigenetic profile, based on DNA methylation and histone modifications, has been described for a few samples[7–9], the variability in chromatin accessibility and resulting functional heterogeneity across GSCs has not been previously investigated. Indeed, GSCs derived from different patient tumours were shown to share numerous common features in their chromatin accessibility landscape[10]. However, phenotypic variability, such as differentiation capacity, was also reported between GSCs that exhibit specific differences in their chromatin accessibility[11], warranting a comprehensive assessment of heterogeneity across GSCs.

Here we reveal three novel and distinct GSC subtypes based on the integrative analysis of chromatin accessibility, DNA methylation and gene expression on a cohort of 27 patient tumour-derived GSCs. Each GSC subtype is regulated by a specific set of transcription factors, uniquely essential for growth in the respective subtypes. Through a single-cell clonal analysis, we show that a GBM tumour can harbour more than one GSC subtype. In addition, we not only identify subtype-biased growth inhibitors through our drug response screening assay but also show that all GSC subtypes commonly express the serotonin receptor 5-HT2 and are

sensitive to the dopamine/serotonin receptor ligand perphenazine. Overall, our results suggest that patients could benefit from serotonin receptor inhibitors or a combination therapy to address the GSC heterogeneity using drugs targeting the different GSC subtypes populating each GBM tumour.

In a cohort of 27 patient tumour-derived GSCs assessed for chromatin accessibility using ATAC-seq, over 16% of accessible regions were shared by over half the samples (Figure 1a). This indicates that a large set of chromatin accessibility features is shared within GSCs, as previously reported[10]. We also compared the full set of accessible regions in GSCs to those identified in human fetal neural stem cells (HFNS), the closest available normal tissue to the GSCs[5]. Over 38% of accessible genomic regions in GSCs were shared with HFNS (97,578 out of 255,890) (Supplementary Figure 1). This suggests that the apparent homogeneity in chromatin accessibility observed across the GSCs may in fact correspond to core features also required by their normal counterparts, and not necessarily linked to the shared tumourigenicity of the GSCs. Indeed, of the accessible regions shared between GSCs and HFNS, over 40% can be found in more than half the GSC lines (Figure 1b). Conversely, just over 1% of the ATAC-seq peaks exclusive to GSCs were shared by more than half of the GSCs (Figure 1c). However, over a quarter (26%) of those ATAC-seq peaks were common to subsets of three or more GSCs suggesting the presence of subtypes within the GSC cohort.

To identify GSC subtypes, we integrated chromatin accessibility (ATAC-seq) with data types previously used to subset GBMs, namely DNA methylation (Illumina EPIC array) and gene expression (RNA-seq) using Similarity Network Fusion (SNF)[12]. Each data type was first modelled individually as a separate network,

followed by fusion of the three networks to generate a combined network most stably supported by all three data types (Figure 1d,e). Spectral clustering on the fused network-based analysis identified three clusters revealing novel and distinct GSC subtypes (C1, C2, C3), supported overall by all three data types (Figure 1e) but with sample proximity within individual clusters better supported by specific data types. For example, sample proximity in the fused network within C1 and C3 is most strongly supported by chromatin accessibility and gene expression, while sample similarity within C2 is driven by DNA methylation and chromatin accessibility (Figure 1e). Comparing the final clusters obtained through the combined SNF to those generated from individual data types revealed that gene expression and DNA methylation largely support the final clustering, but fail to replicate the fused SNF-defined clusters (Figure 1f). However, chromatin accessibility alone is sufficient to recapitulate the fused SNF-defined subtypes. For comparison, we mapped the commonly reported[13] expression-based mesenchymal, classical, neuronal, and proneural classification derived from bulk GBM tumours onto the 27 GSCs. Although all four bulk GBM tumour signatures are represented in our cohort of GSCs, multiple signatures were significantly enriched within each GSC and within each subtype (Figure 1f). In addition, this expression-based classification of the GSCs does not match the expression-only SNF clusters of GSCs (Figure 1f), further highlighting the inadequacy of these signatures to subtype GSCs. Taken together, these results suggest that chromatin accessibility underlies the biological variation between these novel GSC subtypes, providing a unique data type to delineate the mechanisms driving the specific identity of each cluster.

To that end, we first assessed the subtype-specific coverage of chromatin accessibility across our cohort through a saturation analysis of the ATAC-seq data.

Using a self-starting nonlinear regression model, we predicted a 93%, 88%, and 71% saturation for the detection of regions of accessible chromatin across GSCs from the C1 (n=13), C2 (n=9), and C3 (n=5) subtypes, respectively (Figure 1g). Considering that accessible chromatin provides binding sites for transcription factors to regulate gene expression, we then performed DNA recognition motif enrichment analysis across subtypes to uncover regulators of GSC identity. We identified enriched DNA recognition motifs using HOMER[14] on regions exclusively accessible in each of the three GSC subtypes and grouped them into families with CIS-BP[15] (Figure 2a and Supplementary Figure 2). Interestingly, the most enriched DNA recognition motif families in each subtype were either depleted or showed only low-level enrichment in the other subtypes. The DNA recognition motifs for the interferon-regulatory factor (IRF) and Cys2-His2 zinc finger (C2H2 ZF) transcription factor families were enriched in the C1 subtype (Figure 2a). Regulatory factor X (RFX) and basic helix-loop-helix (bHLH) binding motifs were enriched in the C2 subtype, and the DNA recognition motif for the Forkhead family of transcription factors was enriched in the C3 subtype (Figure 2a). We subsequently tested the subtype-specific essentiality for growth of each transcription factor through genome-wide CRISPR-Cas9 screens in GSC lines representative of each subtype (Figure 2b,c) and validated their expression (Supplementary figure 3). Together, these analyses identified six transcription factors across the three GSC subtypes whose DNA recognition motif is exclusively enriched in a given subtype, that are expressed in GSCs from that subtype, and are exclusively essential for the proliferation of GSCs in that subtype: SP1 in C1, ASCL1, OLIG2, AHR, and NPAS3 in C2, and FOXD1 in C3 (Figure 2c).

These transcription factors were previously associated with crucial roles in GBM and/or GSC function. C1-associated SP1 is involved in cellular differentiation

and growth, apoptosis, response to DNA damage, chromatin remodelling[16], and the stimulation of *TERT* expression[17] in cancer stem cells, and increases stemness and invasion in GBM[18]. Although SP1 is essential for growth in only one of the three C1 samples tested, other members of the SP1 regulatory network are found to be exclusively essential in the other two C1 samples, suggesting the SP1 network as a whole is the key regulator of this GSC subtype. Of the C2-enriched essential transcription factors, OLIG2 is a known GSC marker[19], while ASCL1 is a critical regulator of GSC differentiation[11]. FOXD1, enriched in C3, is a known pluripotency regulator and determinant of tumourigenicity in GSCs where it regulates the transcriptional activity of the aldehyde dehydrogenase ALDH1A3, an established functional marker for mesenchymal GSCs[20,21]. In addition, all six transcription factors display significantly higher expression in GBM compared to normal brain (Figure 2d), further supporting their function as key regulators of tumour initiation and development.

Our data were generated on patient-derived GSC lines plated as mixed populations and enriched for self-renewal and tumour-initiating cells over several passages (Figure 3a). However, GBM tumours are reported to contain heterogeneous GSC populations[8,10,22], that could potentially correspond to a hierarchy of stem cell states. To assess the heterogeneity in GSC subtypes of the GBM tumours in our cohort, we delineated the subtypes of clonal populations derived from single cells initially sorted out from a fresh bulk tumour mixed population using cell-surface markers (Figure 3a). In total, we used six clonal populations derived from the same tumour region as the matched G498 GSC, four for G551, and six for G648. We then performed unsupervised hierarchical clustering using the ATAC-seq signal of the individual clones along with our cohort of 27 GSCs,

excluding the G498, G551 and G648 GSCs. These were removed to avoid any patient-specific bias in the clustering (Figure 3b-d). The G498 and G551 GSCs classify within the C1 subtype (Figure 1d-f). However, only 60% (n=10) of their matched clones clustered within C1 (3 of 6 clones matched to G498, and 3 of 4 matched to G551) (Figure 3b,d), while 40% clustered within the C2 subtype. Clones matched to G648 all clustered within C2, the same subtype as G648 (Figure 3c). The scarcity of C3 subtype GSC lines and available tumour tissue prevented the derivation of clonal populations from the original bulk matched tumours. Together, these data suggest that adult GBM tumours can rely on GSCs from more than one subtype.

Therapeutic options for GBM patients are currently extremely limited, and the subdivision of GSCs into the functional subtypes presented here offers new possibilities to identify effective compounds. We thus first re-analysed a published drug screen performed on GSCs[8] to identify compounds with subtype-specific effectiveness. Additional GSC populations were tested with nine of the more promising compounds: they had been tested in the original screen on populations from least one of the GSC subtypes and had shown no impact on the proliferation of normal HFNS cells[8]. In total, two drugs showed no impact on any of the subtypes (Supplementary Figure 4), and, as expected, we identified several compounds with subtype-biased effects, such as ML-9, DL-Cycloserine, Phorbol 12−myristate 13−acetate, and Tetraethylthiuram disulfide (Figure 4a).

Unexpectedly, however, three compounds inhibited proliferation by over 20% across all GSC subtypes. Two of these are neurotransmitter signalling disruptors,

shown to bind to a variety of receptors and neurotransmitter uptake channels: SB224289 hydrochloride and GBR-2909 dihydrochloride (Figure 4b). The only neurotransmitter-related gene expressed across more than 75% of the GSC populations and known to be bound by both compounds is the serotonin receptor 5-HT2 (Figure 4c), identifying serotonin signalling as a likely common pathway for targeting in all GSC subtypes. As neither of the identified compounds are clinically available, we tested the approved antipsychotic drug perphenazine on representative lines of each subtype. Perphenazine, with affinity to multiple serotonin receptors[23], has been previously shown to have no effect on the viability of normal cells up to concentrations of 50 μM[24]. However, end-point viability assays demonstrate that perphenazine has a strong negative proliferative effect on members of all three GSC subtypes (Figure 4d,g,i). Limiting dilution assays additionally revealed that, after two weeks of treatment, perphenazine significantly reduces the sphere-forming capacity of GSCs from all three subtypes (Figure 4e,f,h,i,k,l and Supplementary Figure 5). Through chromatin-based subtyping we have thus identified both compounds with subtype-biased efficacy and a clinically-approved drug that impacts the self-renewal capacity of GSCs in a subtype-agnostic manner.

Taken together, our work identifies three distinct GSC subtypes based on chromatin accessibility, DNA methylation, and gene expression data. Each GSC subtype relies on exclusive networks of essential transcription factors. Notably, our clonal analysis demonstrates that multiple GSC subtypes are present within individual tumours and we demonstrate the efficacy of perphenazine, a clinically-approved drug, in decreasing the self-renewal capacity of all GSCs.

ACKNOWLEDGEMENTS

We would like to thank the Princess Margaret Genomics Centre for their assistance in this study, and Aude Gerbaud for her help with figure formatting. This work was supported by the Princess Margaret Cancer Foundation and SU2C Canada Cancer Stem Cell Dream Team Research Funding (SU2C-AACR-DT-19-15) provided by the Government of Canada through Genome Canada and the Canadian Institute of Health Research, with supplemental support from the Ontario Institute for Cancer Research, through funding provided by the Government of Ontario. Stand Up To Cancer Canada is a Canadian Registered Charity (Reg. # 80550 6730 RR0001). Research Funding is administered by the American Association for Cancer Research International - Canada, the Scientific Partner of SU2C Canada. This study was also conducted with the support of the Ontario Institute for Cancer Research through funding provided by the Government of Ontario for the Brain Cancer Translational Research Initiative. M.L. holds an Investigator Award from the Ontario Institute for Cancer Research and a Canadian Institutes of Health Research (CIHR) New Investigator Award. P.G is supported by a CIHR Fellowship (MFE 338954). PBD is also supported by CIHR, OICR, the Terry Fox Research Institute, and the Hospital for Sick Children Foundation, Jessica's Footprint, Hopeful Minds, and the Bresler Family. PBD holds a Garron Chair in Childhood Cancer Research at the Hospital for Sick Children. S.A. is supported by the CIHR, the Terry Fox Institute and the Canadian Cancer Society.

AUTHOR CONTRIBUTIONS

ML, PG, and PBD conceptualized and designed the study assisted by BHK, SA, and FJC. PG conducted the genomics experiments and designed and/or implemented most of the computational and statistical approaches. MMK performed all the tissue

culture and in vitro drug screens, with the help of NR, XC, and MM under the supervision of PD. FMGC performed SNF, under the supervision of MDT. GM and SA contributed the essentiality screen data. SAMT contributed to the computational analysis, under supervision of BHK and ML. The manuscript was written by PG and ML with input from all other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests

REFERENCES

1. Carlsson, S. K., Brothers, S. P. & Wahlestedt, C. Emerging treatment strategies for glioblastoma multiforme. *EMBO Mol. Med.* **6,** 1359–1370 (2014).
2. Stupp, R. *et al.* Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352,** 987–996 (2005).
3. von Neubeck, C., Seidlitz, A., Kitzler, H. H., Beuthien-Baumann, B. & Krause, M. Glioblastoma multiforme: emerging treatments and stratification markers beyond new drugs. *Br. J. Radiol.* **88,** 20150354 (2015).
4. Venere, M., Fine, H. A., Dirks, P. B. & Rich, J. N. Cancer stem cells in gliomas: Identifying and understanding the apex cell in cancer's hierarchy. *Glia* **59,** 1148–1154 (2011).
5. Chen, J. *et al.* A restricted cell population propagates glioblastoma growth after chemotherapy. *Nature* **488,** 522–526 (2012).
6. Gallo, M. *et al.* MLL5 Orchestrates a Cancer Self-Renewal State by Repressing the Histone Variant H3.3 and Globally Reorganizing Chromatin. *Cancer Cell* **28,** 715–729 (2015).

7. Orzan, F. *et al.* Genetic Evolution of Glioblastoma Stem-Like Cells from Primary to Recurrent Tumor. *Stem Cells* (2017). doi:10.1002/stem.2703
8. Meyer, M. *et al.* Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* (2015). doi:10.1073/pnas.1320611111
9. Rheinbay, E. *et al.* An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep.* **3,** 1567–1579 (2013).
10. Lan, X. *et al.* Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature* **549,** 227–232 (2017).
11. Park, N. I. *et al.* ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* (2017). doi:10.1016/j.stem.2017.06.004
12. Wang, B. *et al.* Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **11,** 333–337 (2014).
13. Verhaak, R. G. W. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17,** 98–110 (2010).
14. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–589 (2010).
15. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158,** 1431–1443 (2014).
16. O'Connor, L., Gilmour, J. & Bonifer, C. The Role of the Ubiquitously Expressed Transcription Factor Sp1 in Tissue-specific Transcriptional Regulation and in

Disease. *Yale J. Biol. Med.* **89,** 513–525 (2016).

17. Liu, T., Yuan, X. & Xu, D. Cancer-Specific Telomerase Reverse Transcriptase (TERT) Promoter Mutations: Biological and Clinical Implications. *Genes* **7,** (2016).
18. Lee, W. S. *et al.* Specificity protein 1 expression contributes to Bcl-w-induced aggressiveness in glioblastoma multiforme. *Mol. Cells* **37,** 17–23 (2014).
19. Trépant, A.-L. *et al.* Identification of OLIG2 as the most specific glioblastoma stem cell marker starting from comparative analysis of data from similar DNA chip microarray platforms. *Tumour Biol.* **36,** 1943–1953 (2015).
20. Cheng, P. *et al.* FOXD1-ALDH1A3 Signaling Is a Determinant for the Self-Renewal and Tumorigenicity of Mesenchymal Glioma Stem Cells. *Cancer Res.* **76,** 7219–7230 (2016).
21. Koga, M. *et al.* Foxd1 is a mediator and indicator of the cell reprogramming process. *Nat. Commun.* **5,** 3197 (2014).
22. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344,** 1396–1401 (2014).
23. Sweet, R. A. *et al.* Pharmacologic profile of perphenazine's metabolites. *J. Clin. Psychopharmacol.* **20,** 181–187 (2000).
24. Kuzu, O. F., Gowda, R., Noory, M. A. & Robertson, G. P. Modulating cancer cell survival by targeting intracellular cholesterol transport. *Br. J. Cancer* **117,** 513–524 (2017).
25. Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx247
26. Pollard, S. M. *et al.* Glioma stem cell lines expanded in adherent culture have

tumor-specific phenotypes and are suitable for chemical and genetic screens. *Cell Stem Cell* **4,** 568–580 (2009).

27. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10,** 1213–1218 (2013).
28. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30,** 428–430 (2014).
29. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
30. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
31. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).
32. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15,** 554 (2014).
33. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3* **7,** 2719–2727 (2017).
34. Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163,** 1515–1526 (2015).

FIGURE LEGENDS

Figure 1 GSCs cluster into three distinct groups. a) Distribution of all ATAC-seq peaks across the 27 GSC samples. b) Distribution of ATAC-seq peaks shared between GSCs and HFNS across the 27 GSC samples. c) Distribution of GSC-exclusive ATAC-seq peaks across the 27 GSC samples. d) Sample networks determined by SNF on individual data types, modelled using Cytoscape. e) Fused SNF network, using DNA methylation (Methylation), gene expression (Expression), and chromatin accessibility (ATAC). f) Sample similarity as determined by SNF (central heatmap, darker blue = more similar, lighter blue = less similar); clustering of individual data types compared to the combined network (left, right and bottom); mapping of expression-determined GBM subtypes using GSEA signatures (top): any displayed signature was found significantly associated with that sample and signatures are ordered from top to bottom in order of increasing GSEA enrichment score. g) Saturation curves of accessible chromatin regions for all 27 GSCs and individual subtypes.

Figure 2 GSC subtypes are regulated by subtype-specific essential TFs. a) Motif family enrichment in each cluster; log2(Fold Enrichment) > 0.5 threshold selected based on the distribution of values in each cluster (Supplementary Figure 3). b) Schematic of drop-out essentiality screen using GSCs stably expressing Cas9 and gRNA libraries. c) z-score distribution of key essential genes in each cluster. Red line corresponds to the empirically determined threshold for essentiality in each tested line. d) Expression levels of key transcription factors in tumour and normal samples, analysed and displayed using GEPIA[25].

Figure 3 GSCs from multiple clusters are present in individual tumours. a) Schematic of the derivation of lines and clones from patient tumours. b) Results of hierarchical clustering of each of the six clones matched to G498 with the 24 GSCs. The G498 line clusters with C1, while three of the six matched clones cluster within C2. c) Results of hierarchical clustering of each of the six clones matched to G648 with the 24 GSCs. The G648 line clusters with C2, and all six matched clones cluster within C2. d) Results of hierarchical clustering of each of the four clones matched to G551 with the 24 GSCs. The G551 line clusters with C1, while 1 of the four matched clones clusters within C2.

Figure 4 Serotonin signalling as a common target among GSC subtypes. a) Drug screen of 165 compounds on GSC populations. b) Binding targets of SB224289 and GBR-12909. Edge width represents binding affinity between the drug and its targets, while colour represents the evidence source for the interaction; green= experimental evidence, others = predicted interactions from text-mining, co-expression, etc. c) Expression of neurotransmitter-related genes across the 27 GSC populations. d) Dose response curve for perphenazine in C1 lines. e-f) Limiting dilution assay with DMSO or perphenazine (5 µM) treatment in C1 lines. g) Dose response curve for perphenazine in C2 lines. h-i) Limiting dilution assay with DMSO or perphenazine (5 µM) treatment in C2 lines. j) Dose response curve for perphenazine in C3 lines. k-l) Limiting dilution assay with DMSO or perphenazine (5 µM) treatment in C3 lines.

ONLINE METHODS

Patient samples and cell culture

All tissue samples were obtained following informed consent from patients, and all experimental procedures were performed in accordance with the Research Ethics Board at The Hospital for Sick Children (Toronto, Canada). Approval to pathological data was obtained from the respective institutional review boards. Primary tissue samples were dissociated in artificial cerebrospinal fluid followed by treatment with enzyme cocktail at 37°C. GSC lines were grown as adherent monolayer cultures in serum-free medium as previously described[26]. Briefly, cells were grown adherently on culture plates coated with poly-L-ornithine and laminin. Serum-free NS cell self-renewal media (NS media) consisted of Neurocult NS-A Basal media, supplemented with 2 mmol/L L-glutamine, N2 and B27 supplements, 75 µg/mL bovine serum albumin, 10 ng/mL recombinant human EGF (rhEGF), 10 ng/mL basic fibroblast growth factor (bFGF), and 2 µg/mL heparin. Single cell-derived clonal populations were generated by staining for cell surface markers CD15 and human specific CD133/1 followed by FACS live sorting. Single cells from four populations (CD negative, CD15 positive, CD133 positive CD15/CD133 double positive) were collected and expanded in serum-free conditions[8].

Chemical Screen and Secondary Screen/Dose Response Curve

Cells were screened with the LOPAC library (Sigma) at the High-Throughput Screening Division, formerly known as the SMART Laboratory, at the Lunenfeld-Tanenbaum Research Institute. Cells were seeded in laminin coated 384-well plates at a density of 2000 cells per well and chemicals were added at a concentration of 5 M and incubated for five days at 37°C. Cell viability was assessed by measuring Alamar Blue incorporation as per the manufacturer's protocol (Invitrogen). Percent growth inhibition was calculated relative to DMSO treated control wells. The potency

of hits from the primary screen were re-tested at 1 M and 5 M or in a 9-point 2-fold dilution series ranging from 50 M-0.2 M concentrations.

In Vitro Limiting Dilution Assay

Cells were plated in serial dilutions on non-adherent 96-well plates and in six biological replicates under NS conditions. Serial dilutions ranged from 2000 cells to 3 cells per well. After 7 and 14 days of plating with chemical or vehicle, each well was scored for the presence or absence of neurosphere formation. Data was plotted and tested for inequality in frequency between multiple groups and tested for adequacy of the single-hit model using Extreme Limiting Dilution Analysis (ELDA) software.

ATAC-seq

ATAC-seq was used to profile the accessible chromatin landscape of 27 GSC lines and 17 clonal lines. 50,000 cells were processed from each sample as previously described[27]. The resulting libraries were sequenced with 50 bp single-end reads which were mapped to hg19. Reads were filtered to remove duplicates, unmapped or poor quality (Q <30) reads, mitochondrial reads, chrY reads, and those overlapping the ENCODE blacklist. Following alignment, accessible chromatin regions/peaks were called using MACS2. Default parameters were used except for the following: --keep-dup all -B --nomodel --SPMR -q 0.05 --slocal 6250 --llocal 6250. The signal intensity was calculated as the fold enrichment of the signal per million reads in a sample over a modelled local background using the bdgcmp function in MACS2.

A given chromatin region was considered exclusive to one of the clusters if it was called as a peak in any of the cluster's samples using a q-value filter of 0.05 and was

not called as a peak in any of the other samples using a q-value filter of 0.2, in order to ensure stringency of exclusivity.

The ATAC-seq saturation analysis was performed by randomizing the order of samples, and successively calculating the number of additional peaks discovered with the addition of each new sample. This process was repeated 10,000 times and averaged. A self-starting non-linear regression model was then fitted to the data to estimate the level of saturation reached.

In the clonal analysis, we mapped the signal of all clones to the full catalogue of peaks identified in the 27 GSC lines. For each clustering, the signal matrix for the 24 GSCs and the 1 clonal population was quantile normalised before clustering.

Data have been deposited at GEO (GSE109399).

DNA Methylation arrays

Bisulphite conversion of the DNA for methylation profiling was performed using the EZ DNA Methylation kit (Zymo Research) on 500 ng genomic DNA from all 27 samples. Conversion efficiency was quantitatively assessed by quantitative PCR (qPCR). The Illumina Infinium MethylationEPIC BeadChips were processed as per manufacturer's recommendations. The R package ChAMP v2.6.4[28] was used to process and analyse the data.

Data have been deposited at GEO (GSE109399).

RNA-seq

RNA was extracted from GSC lines using the Qiagen RNeasy Plus kit. RNA sample quality was measured by Qubit (Life Technologies) for concentration and by Agilent Bioanalyzer for RNA integrity. All samples had RIN above 9. Libraries were prepared

using the TruSeq Stranded mRNA kit (Illumina). Two hundred nanograms from each sample were purified for polyA tail containing mRNA molecules using poly-T oligo attached magnetic beads, then fragmented post-purification. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. This is followed by second strand cDNA synthesis using RNase H and DNA Polymerase I. A single "A" base was added and adapter ligated followed by purification and enrichment with PCR to create cDNA libraries. Final cDNA libraries were verified by the Agilent Bioanalyzer for size and concentration quantified by qPCR. All libraries were pooled to a final concentration of 1.8nM, clustered and sequenced on the Illumina NextSeq500 as a pair-end 75 cycle sequencing run using v2 reagents to achieve a minimum of ~40 million reads per sample. Reads were aligned to hg19 using the STAR aligner v2.4.2a [29] and transcripts were quantified using RSEM v1.2.21[30].

Data are being deposited at EGA.

Motif Enrichment

Regions exclusively accessible in one of the GSC subtypes and not the others were used as input sequences for the motif enrichment, while the full ATAC-seq catalogue served as the background set when running HOMER v4.7 to detect enrichments of transcription factor binding motifs. Enriched motifs were then grouped into families based on similarities in DNA-binding domains using the CIS-BP database[15]. Each family was assigned the fold-enrichment value of the most enriched motif within the family.

The transcription factors whose motifs were found enriched in C1-exclusive accessible regions, were run together through GSEA[31], and the gene set corresponding to genes potentially regulated by SP1 was identified as significantly enriched (GSEA gene set GGGCGGR_SP1_Q6).

Gene essentiality screen

Illumina sequencing reads from genome-wide TKOv1 CRISPR screens in patient-derived GSCs were mapped using MAGECK[32] and analysed using the BAGEL algorithm with version 2 reference core essential genes/non-essential genes[33,34]. Resultant raw Bayes Factor (BF) statistics were used to determine essentiality of transcription factor genes using a minimum BF of 3 and a 5% FDR cut-off.

Similarity Network Fusion

The Similar Network Fusion (SNF) method was run the 27 cell lines using gene expression, DNA methylation and ATAC-seq data. The SNF method does not require any prior feature selection so we used the full matrix of gene expression (20753 genes), the full matrix of methylation data (Beta values, 629309 probes) and the ATAC-seq peaks matrix (255890 peaks). We used the SNFtool R package (v2.2.0) with the parameters K = 10, alpha = 0.4, T = 20, as determined through empirical testing. Spectral clustering implemented in the SNFtool package was run on the SNF fused similarity matrix to obtain the groups corresponding to k=2 to 12.

We identified the top associated genes and methylation probes and ATAC-seq peaks that have the largest agreement with the final fused network structure. To do so we computed the Normalized Mutual Information (NMI) score (as part of the SNFtool package) for each feature (i.e each gene, methylation probe and ATAC-seq

peak). For each feature, we constructed a patient network based on the feature alone and subsequently used spectral clustering. We then compared the result of the resultant clustering to the one obtained from the whole fused similarity matrix by computing the NMI score as previously described[12]. As mentioned in this paper, a score of 1 indicates the strongest feature and shows that the network of patients based on the given feature leads to the same groups as the fused network. A score of 0 means that there is no agreement between the groups that can be derived from the feature and the fused network groups.

Figure 1

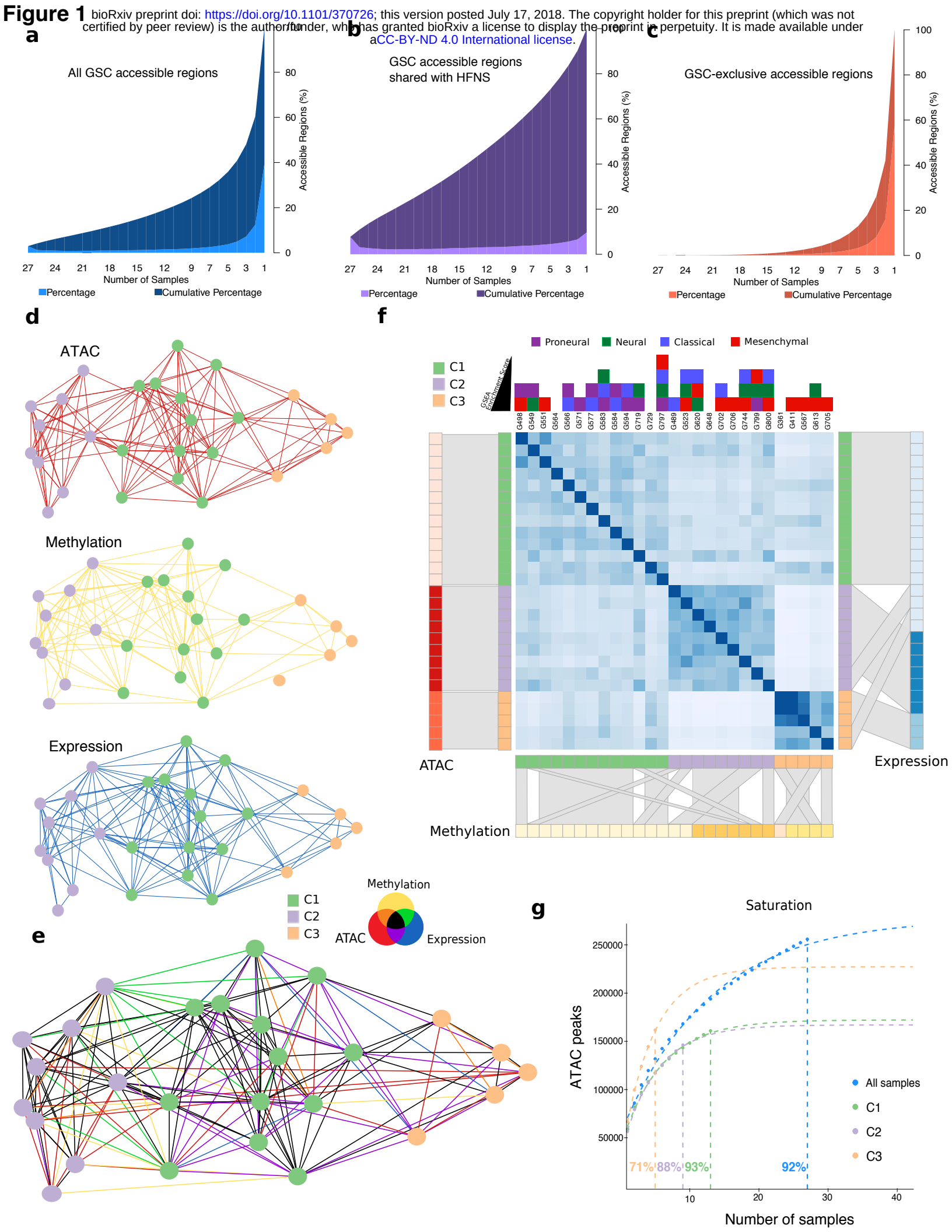

Figure 2

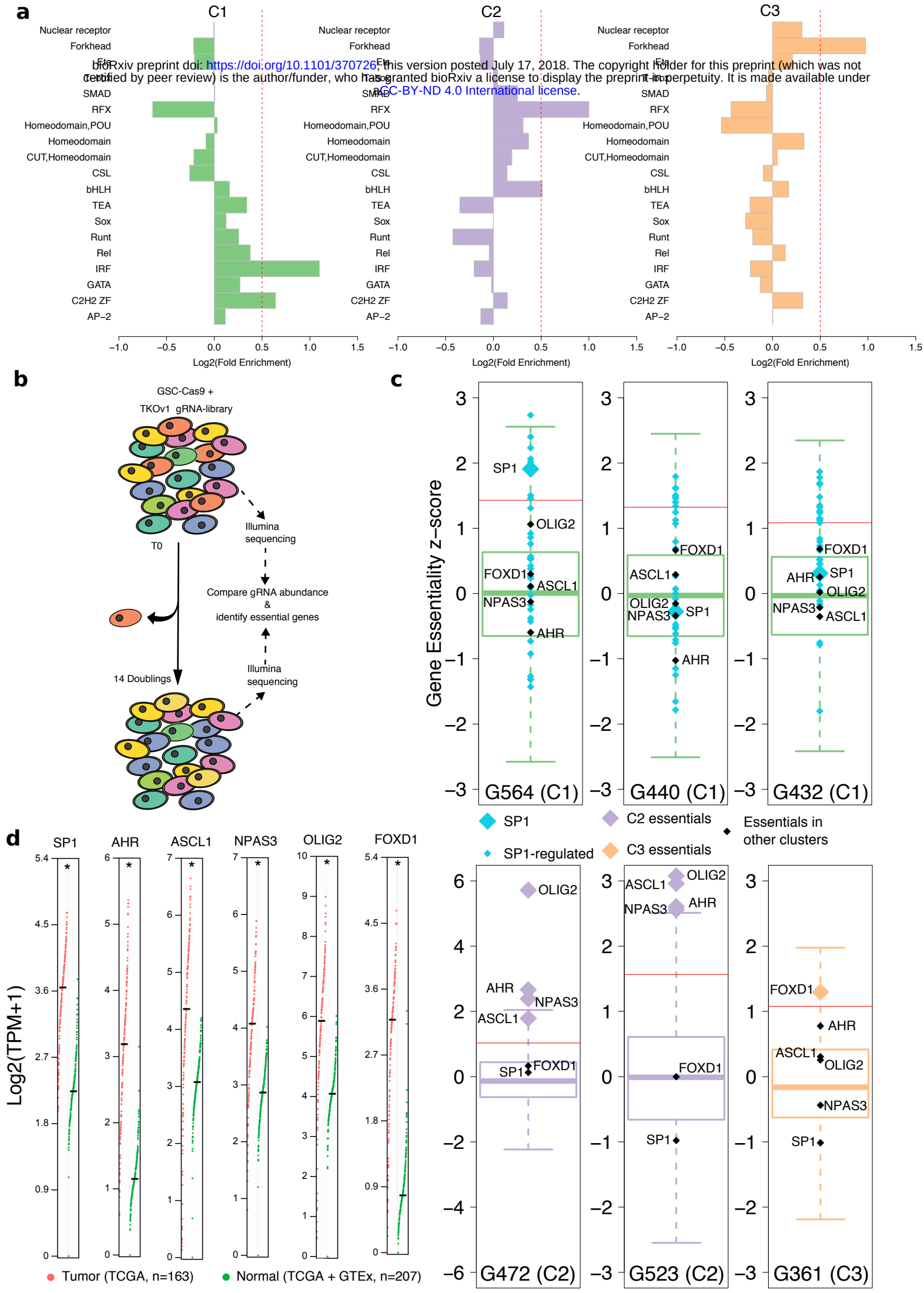

Figure 3

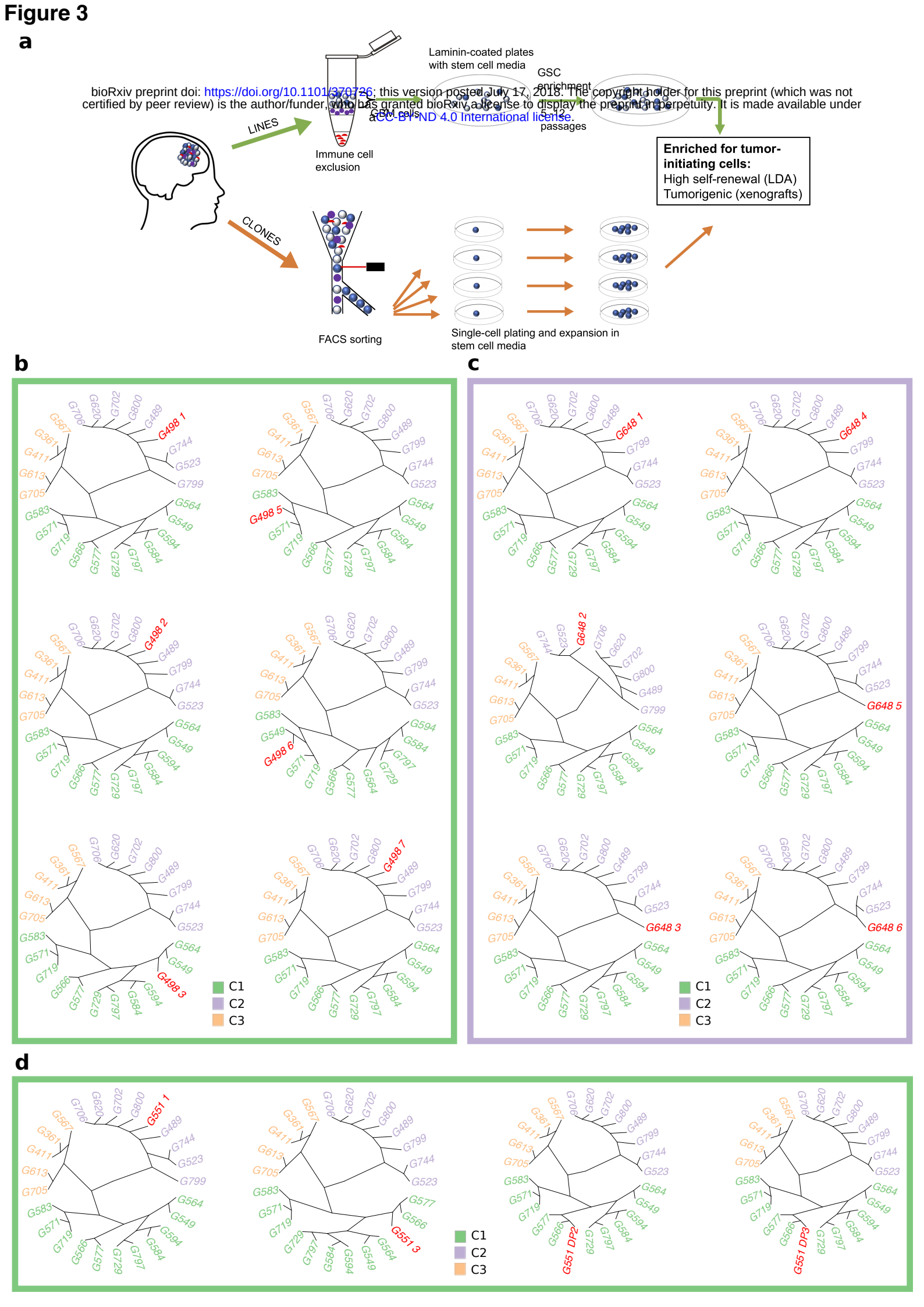

Figure 4