1 **JARID2 facilitates transcriptional reprogramming in glioblastoma in response to standard**

2 **treatment**

3 Nora Rippaus[1], Alexander F-Bruns[1,¥], Georgette Tanner[1], Claire Taylor[1], Alastair Droop[2,¥], Matthew

4 A. Care[2], Joseph Wilkinson[1], Michael D. Jenkinson[3,4], Andrew Brodbelt[3], Aruna Chakrabarty[5],

5 Azzam Ismail[5], Susan Short[1,5] and Lucy F. Stead[1,§]

6

7 [1]Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, LS9 7TF, UK;

8 [2]Leeds Institute of Data Analytics, University of Leeds, LS2 9JT, UK; [3]Walton Centre NHS Trust,

9 Liverpool, L9 7LJ, UK; [4]Institute of Translational Medicine, University of Liverpool, L9 7LJ, UK;

10 [5]Leeds Teaching Hospitals NHS Trust, St James's University Hospital, Leeds, LS9 7TF, UK

11

12 [§]To whom correspondence should be addressed. Tel: +44 113 343 8410 Email:

13 l.f.stead@leeds.ac.uk.

14 [¥] Current address for AFB is Leeds Institute of Cardiovascular and Metabolic Medicine, University

15 of Leeds, Leeds, LS2 9JT, UK and for AD is Wellcome Sanger Institute, Wellcome Genome

16 Campus, Cambridge, CB10 1SA, UK

17

18

ABSTRACT

Background: Glioblastoma (GBM) is a fatal and incurable brain cancer with a dismal prognosis. In order to impact on this disease, we need to understand how infiltrating, non resectable tumour cells resist chemoradiation and facilitate disease recurrence. To this end, we generated or acquired bulk tumour RNA sequencing data from 45 paired primary and locally recurrent GBM tumours (split into original and validation cohorts) from patients that received standard treatment. We also generated DNA methylation profiles for 9 pairs and sequenced RNA from single cells isolated from a patient derived GBM spheroid model at different timepoints following *in vitro* chemoradiation.

Results: We have identified a set of genes with Jumonji and AT-Rich Interacting Domain 2 (JARID2) binding sites in their promoters that are universally dysregulated in post-standard treatment recurrent GBMs compared to the primary tumour. The direction of dysregulation is patient-dependent and not associated with differential promoter DNA methylation. Our *in vitro* experiments suggest that this dysregulation occurs dynamically following treatment as opposed to resulting from selection of cells with specific expression profiles.

Conclusion: JARID2 is an accessory protein to a chromatin remodeling complex, responsible for histone modifications observed during cell state transitions in both normal brain and GBM. We propose that JARID2 facilitates GBM recurrence following treatment by indirect transcriptional reprogramming of surviving cells in whichever manner is needed to reproduce the phenotypic heterogeneity required for tumour regrowth *in vivo*. The mechanism of this reprograming may present a therapeutic vulnerability for more effective treatment of GBM.

Keywords: Glioblastoma; High grade glioma; Recurrence; RNAseq; Cell plasticity; Chromatin remodelling; Longitudinal; Treatment resistance; Tumour adaption; Transcriptional reprogramming

BACKGROUND

Glioblastoma (GBM) is arguably one of the most challenging cancers to treat and is associated with very poor prognosis. This is in part because GBM cells infiltrate the surrounding normal brain making complete surgical removal impossible and, despite subsequent chemoradiation, the remaining cells resist treatment and facilitate tumour regrowth in almost 100% of cases. If we ever hope to more effectively treat GBM we must understand how and why unresected cells resist treatment and form a recurrent tumour. To this end, we and others have focused our attention on molecular profiling of paired primary and recurrent GBM tumours to specifically identify features which are expanded post-treatment and may offer insight into the properties of cells which survive, or the mechanisms that enable their continued proliferation[1-5]. As part of the Glioma Longitudinal AnalySiS consortium, we have analysed the genomes of more than 200 paired gliomas and determined that there is no clear evidence for therapy-driven selection of cells bearing specific resistance-conferring mutations (manuscript under review) in agreement with the work of Körber et al.[3, 6]. We have, therefore, focused our continued efforts herein on transcriptional features and the possibility of therapy-driven selection of GBM cell populations defined by expression profiles. Transcriptional heterogeneity is evident in GBM: expression profiles align with neurodevelopmental-like hierarchies that span genomic subclones and are functionally distinct, including with respect to treatment sensitivity *in vitro*[7, 8]. We generated or acquired RNAseq data from paired primary and recurrent GBMs from a cohort of 23 patients (our original cohort) that underwent standard treatment (debulking surgery followed by chemoradiation with the alkylating drug temozolomide) and had a local recurrence. We then acquired data from an additional 22 such patients (our validation cohort). Our analyses of these data, and of DNA methylation profiles from 9 pairs, consistently highlight the likely role of a Polycomb Repressive Complex 2 (PRC2) accessory protein called JARID2 (Jumonji and AT-Rich Interacting Domain 2) in the transcriptional changes observed after treatment in GBM, via histone modifications and chromatin remodeling. However, the direction of fold change is not consistent across patients. We then performed single cell RNAseq on a patient derived GBM model at different time points following administration of clinically relevant doses of chemoradiation *in vitro* and found that JARID2 associated transcriptional changes occur dynamically after treatment as opposed to resulting from selection of cells with a specific expression profile. JARID2's interaction with PRC2 is fundamental

73 to specifying neurodevelopmental cell lineages in response to environmental cues, and PRC2 has

74 been shown to be necessary for determining GBM cell phenotypes based on tumour

75 microenvironmental pressures, though the role of JARID2 in this has never been investigated[9-12].

76 We propose that GBM recurrence results from JARID2-associated transcriptional reprogramming,

77 via PRC2, of unresected cells in whichever direction enables recapitulation of the transcriptional

78 heterogeneity needed for continued tumour growth *in vivo*[13]. Therapeutic targeting of the

79 mechanism of such reprogramming may constitute a more effective treatment strategy than targeting

80 of the cell types that lie either side of the interconversions.

81

82 **RESULTS**

83 **Differential expression indicates a therapy-driven shift in neurodevelopmental genes**

84 Genes that were differentially expressed (DE) in the recurrent versus primary GBMs in our original

85 cohort (Supp.Table.1) were significantly enriched for those involved in cell development and lineage

86 determination, specifically in relation to neurodevelopment (Fig.1a and Supp.Table.2). Brain cell fate

87 is orchestrated by the combined actions of transcription factors (TFs) and chromatin remodeling

88 complexes, both of which have been implicated in establishing functionally heterogeneous

89 transcriptional hierarchies in gliomas[14, 15]. We therefore reasoned that specific DNA-binding factors

90 may coordinately regulate the genes we observe to be altered after treatment, potentially highlighting

91 certain cell types that survive. Candidate master transcriptional regulators can be identified from

92 expression data via gene set enrichment analysis (GSEA). However, we found that many

93 neurodevelopmental TFs were missing from publicly available gene sets, so we first developed a

94 more comprehensive DNA-binding factor gene set using ChIPseq data from the Gene Transcription

95 Regulation Database (see Methods)[16].

96 **Genes with JARID2 binding sites in their promoters (JBSgenes) are consistently and**

97 **significantly dysregulated in recurrent versus primary GBMs and stratify patients into two**

98 **response subtypes**

99 We performed per-patient GSEA, using our novel gene set, with genes pre-ranked by the magnitude

100 of fold change in expression between the primary and recurrent tumour. Genes with a Jumonji and

101 AT-Rich Interacting Domain 2 (JARID2) binding site in their promoters (JBSgenes) were the most

102 significantly, consistently and highly enriched within the genes changed after therapy across

103 patients: the normalized enrichment score (NES) for JARID2 was significant in all patients

104 (FDR<0.05) and gave the highest score in 91% (n=21/23) (Fig.1b). To determine whether JBSgenes

105 were consistently up-regulated or down-regulated after treatment, we repeated the analysis including

106 the direction of fold change in the gene ranking. We found that the JBSgenes were altered in a

107 consistent direction per patient but across patients the direction varied: in 30% (n=7/23) the JARID2

108 enrichment was being driven by down-regulation of JBSgenes (hereon referred to as D response

109 subtype) and in the remaining 70% (n=16/23) it through up-regulation (U response subtype) (Fig.1c).

110 We acquired data from an additional 22 paired primary and recurrent GBMs from patients who

111 underwent standard treatment and who had a local recurrence (the validation cohort) which
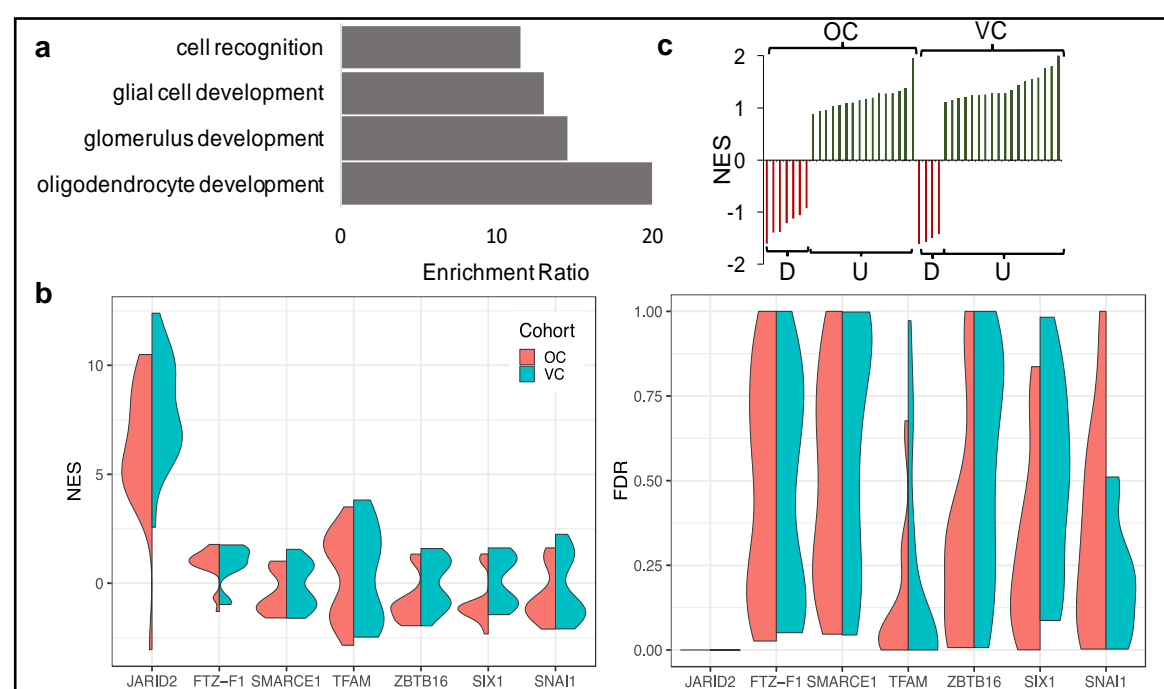
112

113 *Fig.1a) Biological processes enriched in the genes differentially expressed between matched*

114 *primary and recurrent GBMs (enrichment ratio and term count>10, p<0.0005);* ***b)*** *Per-patient*

115 *normalised enrichment scores (NES, left plot) and false discovery rates (FDR, right plot) for the top-*

116 *scoring promoter-binding factors associated with gene expression changes in recurrent vs primary*

117 *GBMs, highlighting the significance of JARID2;* ***c)*** *The NES for JARID2 for each patient (x-axis)*

118 *when direction of fold change is taken into account, shows that there are two response subtypes*

119 *based on whether genes are up (U) or down (D) regulated. OC: original cohort. VC: validation cohort.*

120   corroborated our findings (Figs.1b and c) with a similar ratio of D response subtype (n=4/22) and U

121   response subtype (n=18/22) patients (Fisher's exact test, p=0.52)[5].

122   **The same JBSgenes are dysregulated in each response subtype and their promoter DNA is**

123   **unmethylated in both primary and recurrent GBMs**

124   To investigate whether the JBSgenes driving the enrichment differed across individual patients or

125   between response subtypes, we quantified how often each gene was present in the leading edge of

126   the GSEA results across the 45 patients in the original and validation cohorts combined.
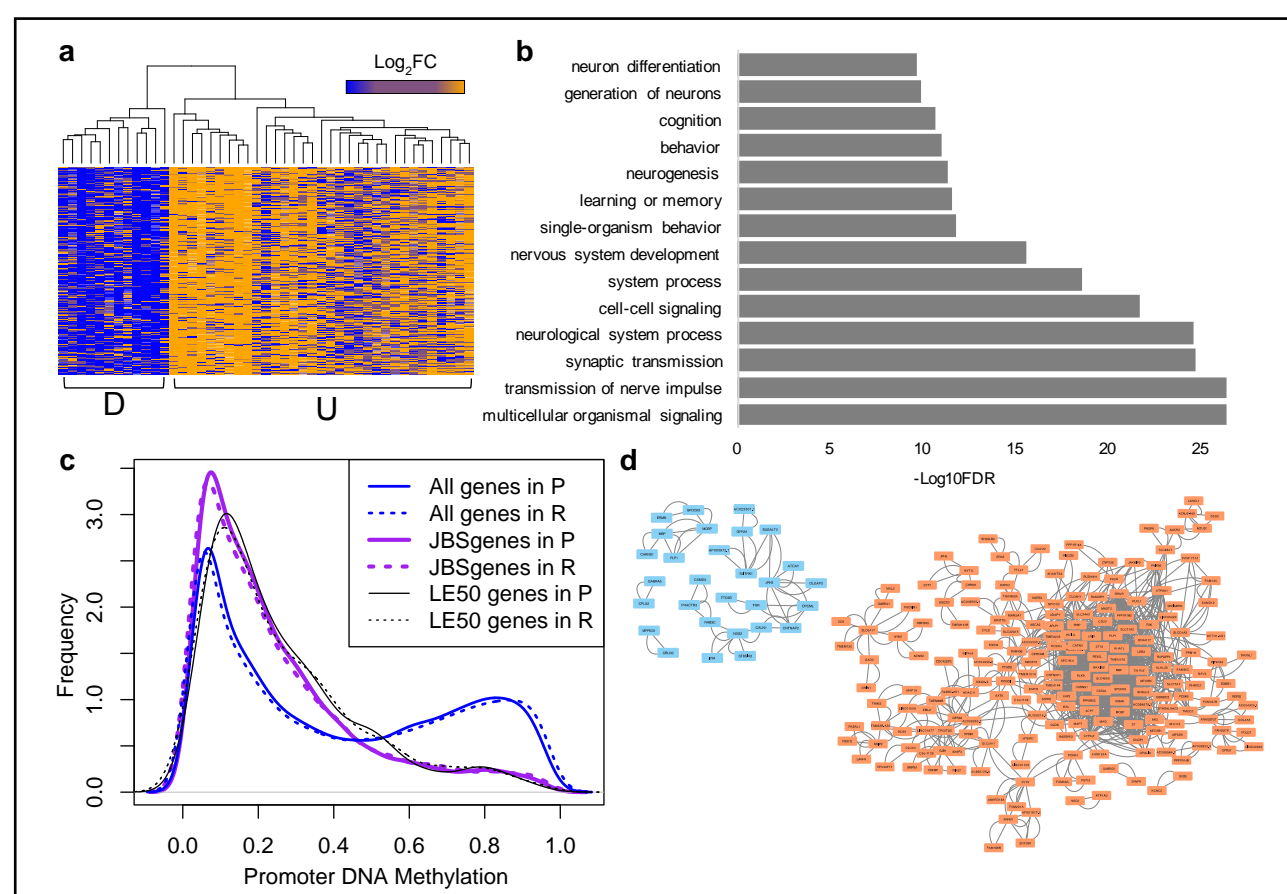

127

Fig.2a) Heatmap of fold change in expression after treatment per patient (columns) for the genes in the leading edge of the JARID2 GSEA results in more than 50% of patients across both cohorts (LE50 genes, rows). These same genes are upregulated in U response subtype patients as are downregulated in D response subtype patients; b) The biological processes (with <2000 terms) most enriched in LE50 genes; c) The distribution of average promoter DNA methylation for all genes, JARID2 binding site (JBS)genes and LE50 genes in primary (P) and recurrent (R) GBM samples; d) Networks of genes (nodes) for which expression is highly correlated (edges: R>|0.9|) with LE70 genes in the primary (blue, top left) and recurrent (orange, right) GBM samples.

136   335 genes were observed in the leading edge of more than 50% of patients (denoted LE50 genes)

137   and 43 genes in more than 70% (LE70 genes) (Supp.Table.3). The per-patient fold-changes of the

138   LE50 genes showed that these same JBSgenes drive the enrichment across patients irrespective of

139   response subtype i.e. the same genes are downregulated in D response patients as are upregulated

140   in U response patients (Fig.2a). The LE50 genes are enriched in functional annotations associated

141   with neurodevelopment and neuronal differentiation such as synaptic plasticity and interneuronal

142   communication (Fig.2b and Supp.Table.4). To investigate the DNA methylation status of the

143   promoters of these genes in primary and recurrent GBMs, in comparison to other genes, we

144   performed genome-wide methylation arrays on DNA from 9 pairs (see Supp.Table.1). No gene

145   promoters were differentially methylated between primary and recurrent samples (q>0.45 for all

146   genes). However, the distribution of promoter methylation across all genes was significantly different

147   to that of the JBSgenes and the LE50 genes in isolation, revealing that the DNA in the promoters of

148   the latter two is unmethylated in both the primary and recurrent GBMs (Fig.2c). This indicates that

149   the change in expression of these genes that we observe post-treatment is not driven by DNA

150   methylation.

151   **LE70 JBSgenes are more coordinately expressed in recurrent GBM**

152   We hypothesised that JARID2 is involved in the tighter co-regulation of leading edge JBSgenes in

153   GBM tumours after treatment, independent of whether that results in an increase or decrease in their

154   expression. To investigate this, we identified all genes for which expression is highly correlated

155   (R>|0.9|) in the primary GBMs and recurrent GBMs separately. We then determined the prevalence

156   of the LE70 genes and their connectivity in these two correlation networks. We found that the LE70

157   JBSgenes correlate with significantly more genes and with significantly more connectivity in

158   recurrent versus primary samples: 1% (29/2603) in the primary GBM network compared to 7%

159   (202/2855) in the recurrent GBM network (chi-squared; p=0 for both tests) (Fig.2d). This implies that

160   either cells in which these genes are co-regulated by JARID2 become more prevalent post-

161   treatment, or that their co-regulation by JARID2 occurs in response to treatment. To inspect this

162   further we designed an *in vitro* experiment to investigate the time course of JBSgene dysregulation

163   following treatment.

164 **Single cell analysis indicates that JARID2 associated gene co-regulation is an adaptive**

165 **response to therapy**

166 We cultured two plates of spheroids directly from a freshly resected primary GBM, in serum-free

167 conditions. We treated one plate with physiologically relevant single doses of TMZ (30μM) and

168 radiation (2Gy). We captured and sequenced RNA from single cells from spheroids one week post-

169 treatment when there was a significant deviation in the untreated vs treated spheroid growth curves,

170 and three weeks post-treatment when growth of the treated spheroids appeared to have recovered

171 (Fig.3). Our bespoke GSEA revealed that JBSgenes were significantly enriched amongst the genes

172 altered in treated versus untreated spheroids three-weeks post-treatment (FDR=0.18) but not one

173 week post-treatment (FDR=0.65). Furthermore, the genes that were DE (p<0.05) between treated

174 and untreated cells included significantly more LE50 genes at the three-week time point compared

175 to the one week time-point (chi-squared, p=0.007). These results suggest that the universal JBSgene

176 dysregulation that we observe in recurrent tumours is not caused by selection of a fixed

177 transcriptional profile, but rather transcriptional reprogramming following treatment.
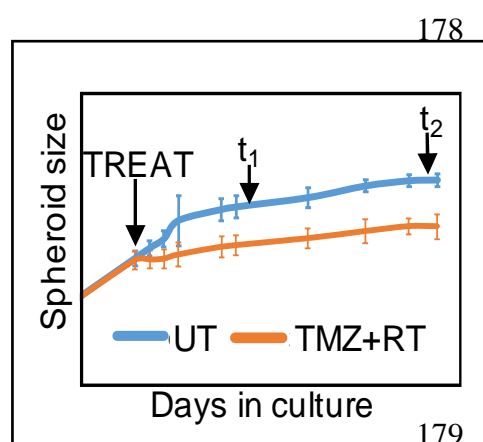
178

Fig.3. Growth curves for untreated (UT) and treated (TMZ+RT) patient-derived GBM spheroids. Time of treatment (TREAT) is indicated (arrow), in relation to single cell capture 1 week (t_1) and 3 weeks (t_2) post-treatment.

179

180 **JARID2 is involved in cell plasticity and implicated in interconversions between cell states in**

181 **glioma**

182 JARID2 is an accessory protein responsible for the genomic positioning of Polycomb Repressive

183 Complex 2 (PRC2)[12]. PRC2 is a chromatin remodeller that is indispensable for lineage

184 determination during neurogenesis[10]. It is responsible for the trimethylation of lysine 27 on histone

185 H3 (H3K27me3) which results in epigenetic silencing of the marked gene. PRC2 is directly implicated
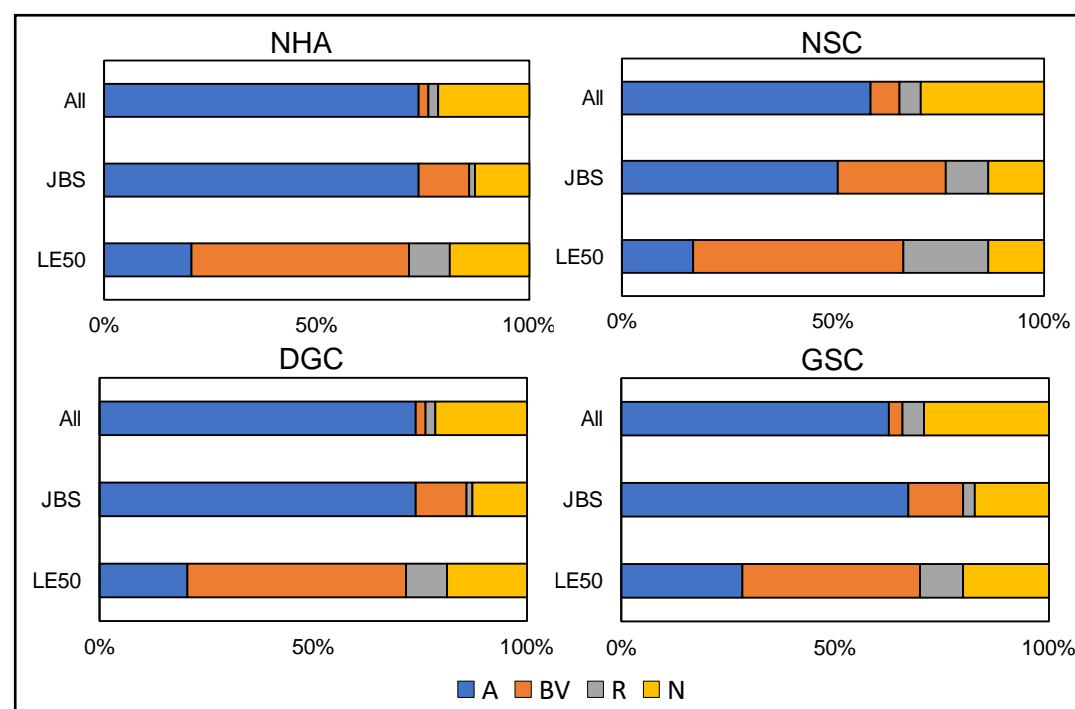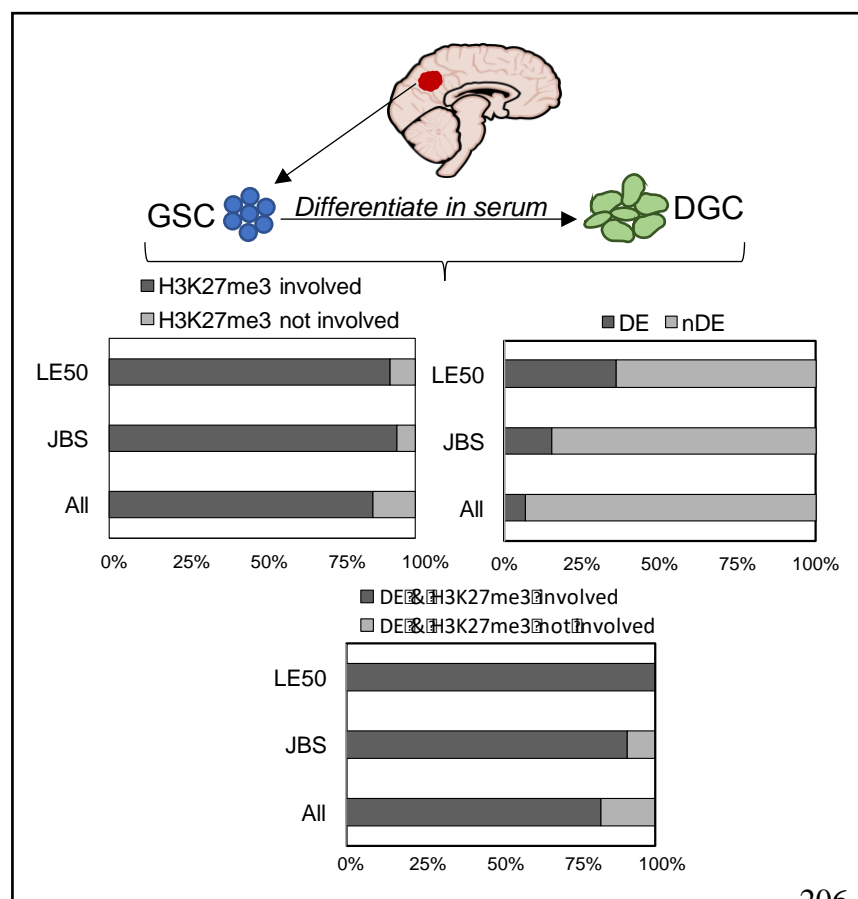
Fig.4 *Relative quantification of promoter status, with respect to specific histone marks, across all genes compared to JBSgenes and LE50 genes in normal human astrocytes (NHA), neural stem cells (NSC), differentiated glioma cells (DGC) and glioma stem cells (GSC). A=active=H3K4me3; R=repressed=H3K27me3; BV=bivalent=H3K27me3+H3K4me3; and N=neither mark.*

in cell plasticity in GBM by studies showing that its catalytic subunit is required to enable conversions between stem-like and differentiated cell types[11]. It is also indirectly implicated by the fact that the prevalence and location of H3K27me3 significantly differs between normal brain and glioma cells, and between GBM cells with different phenotypes[17]. These differences occur most frequently at bivalent promoters: those harbouring both the repressive H3K27me3 and an activating mark (H3K4me3) causing the gene to be silenced but primed for activation upon PRC2 disassociation and H3K27 demethylation. Developmental gene promoters are commonly bivalent in embryonic stem cells to enable subsequent rapid activation of specific lineage determination genes once cell fate is resolved, further highlighting PRC2's role in cell-type transitions[18].

        To determine whether JBSgenes, and specifically the LE50 genes, are implicated in cell type switching we mined published data on histone marks (H3K27me3 and H3K4me3) and gene expression in different normal and GBM cell types [17, 19]. Normal brain cell types were human neural stem cells (NSCs) and normal human astrocytes (NHAs). Different GBM 'cell types' pertains to the

205

Fig.5. *Glioma stem cell (GSC) cell lines can be derived from patient GBM tumours and then further cultured in serum to form differentiated glioma cells (DGC). The bar charts indicate the proportion of genes for which H3K27me3 is involved in any changes in promoter status (top left), or which were differentially expressed (DE, top right), or both (bottom), between matched GSC and DGC cell lines for all genes, JBSgenes or LE50 genes separately.*

206

207   ability to derive phenotypically distinct cell lines from the same patient GBM under different

208   conditions: those which enrich for glioma stem cells (GSCs, which can be considered analogous to

209   NSCs in the normal brain) and those which enrich for differentiated glioma cells (DGCs, somewhat

210   analogous to NHAs)[20]. We first quantified promoter status for each different cell types:

211   active=H3K4me3; repressed=H3K27me3; bivalent=H3K27me3+H3K4me3; and neither mark. In

212   support of the role of JARID2 as a PRC2 accessory protein, and increasing confidence that our novel

213   gene set has captured bona fide JBSgenes, we found that H3K27me3 was significantly enriched at

214   JBSgenes in all cell types investigated (Fig.4: middle compared to top barplot in each panel; grey

215   and orange shading pertains to promoters with the H3K27me3 mark), and that this was particularly

216   pronounced at bivalent promoters (Fig.4; orange shading). Moreover, the presence of H3K27me3,

217   again especially at bivalent promoters, was further significantly enriched within the LE50 subset of

218   JBSgenes (Fig.4 bottom compared to middle barplot in each panel). We then characterized changes

219   in promoter status between cells of different phenotype in the normal brain (NPC vs NHA) and GBM

220   (GSC vs DGC: Fig.5). We found, in agreement with the results from the original publication, that

221 changes involving H3K27me3 are more pronounced than any other. We further found that this is

222 significantly more evident in the promoters of JBSgenes: 85% of changes between NPC and NHA

223 involve H3K27me3 at JBSgene promoters compared to 76% at all gene promoters (chi- squared, p=

224 3.9×10^{-8}) and 93% of changes between GSC and DGC involve H3K27me3 at JBSgene promoters

225 compared to 78% at all gene promoters (Fig.5; chi-squared, p= 5.6×10^{-4}). This implicates JARID2 in

226 chromatin remodeling of gene promoters that differ between cell types in the both normal brain and

227 in GBM.

228 To determine whether this remodeling associates with gene expression changes we

229 identified genes that are DE (q<0.2) between GSC and DGC using RNAseq data from Patel et al.[19]

230 from matched lines derived from three patient samples. As shown in Fig.5, we found that significantly

231 more JBSgenes are DE (15% JBSgenes vs 5% of non-JBSgenes; chi-squared, p=0) and that, within

232 the JBSgenes, significantly more LE50 genes are DE (36% of LE50 genes vs 15% of remaining

233 JBSgenes; chi-squared, p=2.5×10^{-5}). We then overlaid the expression and histone mark data and

234 found, as also shown in Fig.5, that chromatin remodeling involving H3K27me3 is more pronounced

235 at DE JBSgenes (91% of changes at DE JBSgene promoters involve H3K27me3) and DE LE50

236 genes (100% of changes) compared with all DE genes (where 82% of promoter status changes

237 involve H3K27me3). Together, these data suggest that chromatin remodeling associated with

238 JARID2 affects the expression of genes that distinguish different cell types in the normal and human

239 brain.

240

241 **DISCUSSION**

242 We found that the genes differentially expressed between pairs of primary and locally recurrent GBM

243 tumours post standard treatment were enriched for those involved in brain cell development and

244 lineage determination, suggesting that GBM cell types defined within neurodevelopmental-like

245 transcriptional hierarchies may be associated with treatment resistance and tumour regrowth in

246 patients. Brain and GBM cell specification results from a combination of the concerted action of

247 transcription factors (TFs) and differential genome accessibility imposed by a variety of chromatin

248 remodeling molecules and complexes. We, therefore, applied an unbiased approach to investigate

249 whether any such DNA-binding factors were repeatedly implicated across patients in the master

250   regulation of genes for which we observe altered expression in recurrent versus primary GBM

251   tumours. We found that genes with Jumonji and AT-Rich Interacting Domain 2 (JARID2) binding

252   site(s) in their promoter (JBSgenes) are consistently and significantly dysregulated in recurrent GBM

253   tumours, in both our original and validation cohorts. JARID2 is indirectly responsible for eliciting the

254   programmes of epigenetic gene silencing required for cell lineage determination during

255   neurodevelopment[9]. It does so by docking Polycomb Repressive Complex 2 (PRC2) to specific

256   genomic loci where it trimethylates H3K27 to repress gene expression[12, 21]. As well as normal brain

257   cell delineation, PRC2 and the H3K27me3 mark have been specifically implicated in cell state

258   transitions in glioblastoma[11, 17, 22]. However, the involvement of JARID2 in this process, and the

259   genes in which expression is altered by this mechanism to dictate GBM cell phenotype have not

260   previously been elucidated. Our results highlight subsets of genes with JARID2 binding sites in their

261   promoters (JBSgenes) that are most commonly dysregulated (LE50 in more than 50% and LE70 in

262   more than 70% of patients) in GBM tumours after treatment. In support of the notion that expression

263   of these genes is regulated by JARID2-assoicated mechanisms in GBM, we have shown that the

264   DNA in their promoters is unmethylated in both primary and recurrent samples and that their

265   promoters are significantly more associated with the H3K27me3 mark than those of other genes in

266   GBM cell lines. The increasing importance of this coordinated regulation by JARID2 in GBM after

267   treatment is indicated by the larger and more connected networks of highly correlated LE70

268   JBSgenes in recurrent versus primary samples.

269   Together these results could suggest that specific GBM cell types, defined by transcriptional

270   profiles resulting from JARID2-associated epigenetic programming, resist treatment and expand

271   during tumour recurrence. It is widely thought, for example, that glioma stem cells specifically resist

272   treatment and are responsible for GBM regrowth[23, 24]. However, a confounding result in relation to

273   this interpretation is the fact that, whilst a specific subset of JBSgenes are universally dysregulated

274   in recurrent versus primary tumours, the direction of dysregulation is inconsistent; the same genes

275   are upregulated in the recurrence in ~70% of patients and down-regulated in the remaining 30%.

276   Furthermore, our *in vitro* work indicates that changes in expression of JBSgenes occurs dynamically

277   following treatment as opposed to resulting from an increased signal from expansion of a fixed

278   transcriptional profile. Our hypothesis is, therefore, that JARID2-associated chromatin remodeling is

279   not a treatment resistance mechanism per se, but a mechanism by which GBM tumours recover

280   from treatment to enable regrowth. In this way, the different directions of gene dysregulation are

281   owing to the need to recapitulate the GBM transcriptional heterogeneity, required for tumour growth

282   *in vivo*, from whichever cell types survived in that particular patient[13, 25, 26]. This is supported by our

283   findings that the JBSgenes dysregulated during treatment are a) significantly more likely to have

284   bivalent promoters (i.e. poised for activity relating to lineage decisions in response to environmental

285   queues) and b) significantly enriched amongst the genes differentially expressed between cells at

286   either end of the GBM transcriptional hierarchy (i.e. glioma stems cells and differentiated glioma

287   cells). Recent landmark findings also support our hypothesis, which posits that different GBM cell

288   types are able to resist treatment and that interconversions between cell types, as opposed to one-

289   way transitions down a differentiation pathway, are needed to enable tumour regrowth: i)

290   differentiated GBM cells form networks *in vivo* that enable them to better survive chemoradiation,

291   negating the idea that only stem-like cells are able to survive[27]; ii) glioma stem-cells, needed for

292   tumour regrowth owing to their proliferation capabilities, can form via non-hierarchical conversion of

293   differentiated cells in GBM[11, 14, 25].

294   Our hypothesis represents a paradigm shift, also recently suggested by Dirkse et al.[25], that

295   challenges the notion that effective treatment of GBM will be possible by therapeutically targeting

296   any one cell population, such as glioma stem cells. Instead, we propose that effective treatment will

297   only be possible by targeting the mechanisms of GBM cell plasticity resulting from transcriptional

298   reprogramming, which our results suggest are fundamentally linked with the role of JARID2. The

299   precise nature of this role, and its potential for therapeutic targeting, are the focus of our ongoing

300   work.

301

302   **CONCLUSION**

303   We have found that a subset of genes is universally dysregulated in patient GBMs following standard

304   treatment, likely because of epigenetic remodeling of their promoters via mechanisms involving

305   JARID2, as an adaptive response that facilitates tumour regrowth. The direction of this adaptive

306   response is, however, not constant across patients and may depend upon the cell state transitions

307   needed to recapitulate transcriptional heterogeneity in the recurrent tumour. This is the first time that

308 JARID2 has been implicated in GBM cell plasticity in association with tumour recurrence, and

309 highlights subsets of genes that may be involved in cell state transitions required for adaption of

310 GBM tumours to treatment.

311

312 **METHODS**

313 **Archival Samples and Profiling Data**

314 Four independent sources of paired patient GBM samples (surgical tissue from primary GBM and

315 subsequent recurrent samples) were used in this work. Samples were allocated to the original cohort

316 if they had undergone whole transcriptome RNA sequencing, and to the validation cohort if they had

317 undergone poly-A transcriptome sequencing. Clinical information and cohort assignment are given

318 in Supp.Table.1.

319 Stead Samples: 21 patients from four tissue banks (Leeds, Liverpool, Cambridge and Preston) with

320 tumour in paraffin blocks. Ethical approval was acquired (REC 13/SC/0509). RNA and DNA was

321 extracted from the same cells from neuropathologist annotated tumour regions (>60% cancer cells)

322 using appropriate Qiagen kits (Qiagen, Sussex, UK). Paired end strand-specific whole transcriptome

323 libraries were prepared for 16 pairs using the NEBNext Ultra Directional RNA Library Prep Kit for

324 Illumina (New England BioLabs, UK), following rRNA depletion with NEBNext rRNA Depletion Kit or

325 Ribo-Zero Gold. Libraries were sequenced on an Illumina HiSeq. DNA from 9 pairs (4 of which also

326 underwent RNAseq) was profiled using the Illumina Infinium Human Methylation 450K Bead Chip

327 array.

328 Rabadan Samples: Nine patients from Wang et al.[4] with transcriptome sequencing data (7 with

329 whole transcriptome data and 2 with poly-A transcriptome data) for paired tumours, downloadable

330 from the sequencing read archive (accession SRP074425).

331 Verhaak Samples: Four patients from Kim et al.[1] with poly-A transcriptome sequencing alignment

332 data acquired, and converted to raw fastq format, following application to the European Genome-

333 Phenome Archive (accession EGAS00001001033).

334 Nam Samples: 16 patients from Kim et al.[5] with poly-A transcriptome sequencing alignment data

335 acquired, and converted to raw fastq format, following application to the European Genome-

336 Phenome Archive (accession EGAD00001001424).

337 **Sequencing Data: Alignment, Differential Expression, Functional Enrichment and Correlation**

338 **Analysis**

339 RNAseq data was analysed as previously described except that reads were aligned to human

340 reference genome GRCh38, using the gencode.v27 genome annotation as a guide, using

341 STARv2.4.3a and functional enrichment analysis was done using WebGestalt [28-30].

342 **Gene set enrichment analysis (GSEA)**

343 We developed a novel gene set file for use in GSEA using the Gene Transcription Regulation

344 Database (GTRD v18.01), which contained the genomic binding locations of 682 human DNA-

345 binding factors from 4236 chromatin immunoprecipitation sequencing (ChIPseq) experiments[16]. A

346 gene was assigned to a DNA-binding factor's gene set if its promoter (transcription start site from

347 gencodev27 $\pm$1kbp) contained a binding site for that factor in $\geq$2 independent ChIPseq experiments.

348 We first performed pre-ranked GSEA, per patient, ordering genes by the magnitude of fold change

349 in expression $\log_2(|$recurrent FPKM +0.01/primary FPKM+0.01$|$) in classical mode. To indicate

350 directionality of dysregulation we then ranked genes by absolute fold changes ie. using $\log_2($recurrent

351 FPKM +0.01/primary FPKM+0.01) and weighted by magnitude[31].

352 **DNA methylation analysis**

353 The RnBeads package was used to import, quality check and preprocess IDAT files and then

354 perform pairwise differential methylation analysis. The combined and adjusted p-value

355 (comb.p.adj.fdr) in the promoter results file was used to determine significance. The average

356 methylation signal for each promoter (mean.mean) was extracted for both the primary and recurrent

357 samples and used to plot the distribution for different genes subsets in R using the density function.

358 **Patient-derived spheroids**

359 A patient presenting with a suspected GBM was consented for the use of their tissue in research

360 through the Leeds Multidisciplinary Research Tissue Bank (REC 15/YH/0080). GBM diagnosis was

361 confirmed intraoperatively by a neuropathologist who identified a tumour cell rich piece of tissue,

362 surplus to diagnosis, for transport to the laboratory in cold PBS for use in this work. Tissue was

363 washed in PBS and chopped in Accutase (Sigma-Aldrich, 500μL) before incubation at 37°C for 5

364 min. The sample was triturated and Neural Basal (NB) medium, consisting of Neurobasal Medium,

365 N2 and B27 supplements (ThermoFisher, 250mL, 1.25 and 2.5mL respectively), recombinant basic

366 fibroblast growth factor (bFGF), and epidermal growth factor (EGF) (R&D Systems 40ng/mL each),

367 was added to a total volume of 10 mL prior to spinning (1200 rpm at 5 min). The pellet was

368 resuspended in 5ml DNaseI then 1ml RBC lysis buffer (VWR International) with 1 min incubation at

369 room temp, addition of PBS to 10mL and further spinning following each resuspension. The pellet

370 was resuspended in 10mL PBS, filtered via a 70μm and 30μm strainer consecutively and counted.

371 Finally, cells were resuspended in NB-medium to a concentration of $2x10^4$cells/1mL with 200μL of

372 this cell suspension added into each well of an ultra-low-adherence plate and incubated at 37°C

373 5%CO2. 100μL of medium was replaced per well every 3 days. Cells were imaged regularly on the

374 EVOS Cell Imaging System (ThermoFisher) until they reached approximately 300μm in diameter. At

375 this point TMZ (Sigma) was concentrated in 100μL of NB media and used in a media replacement

376 for one plate of cells to give a final dose, per well, of 30μM; one hour later the same plate was

377 irradiated with 2Gy.

378 **Single cell capture and sequencing**

379 Single cells were captured from treated spheroids 1 week and 3 weeks post-treatment, and cells

380 from untreated spheroids the following day, by extracting and combining 8 spheroids per time point

381 and dissociating them via a PBS wash, Accutase incubation, and further PBS washing (as above)

382 to a concentration of 2.5×10^5 cells/mL. Cells were diluted in C1 cell suspension reagent at a ratio

383 of 3:2, respectively. Single cells were captured on a medium (10–17 μM) C1 Single-Cell Auto Prep

384 IFC for mRNAseq, lysed and underwent on-chip cDNA amplification via SMART Seq2 according to

385 the manufacturer's instructions using the Fluidigm C1 Single-Cell Auto Prep System and protocols

386 on Script Hub™ (Fluidigm)[32]. cDNA was quantified using the High Sensitivity DNA Assay on the

387 Agilent 2100 Bioanalyser and paired end Nextera XT (Illumina) libraries were made and sequenced,

388 with multiplexing using indexes provided by Dr Iain Macaulay, on an Illumina HiSeq. RNA

389 sequencing data was processed and expression was quantified as per the bulk tissue. GSEA

390 analysis was performed twice using expression (in fragments per kilobase per million mapped)

391 datasets to identify differences between untreated and treated cells at the 1 week timepoint and

392 three week timepoint separately. Differentially expressed genes were identified using Seurat[33].

393 **Cell type expression and histone modification status**

394   Data on gene promoter histone status in different normal and GBM cell types were extracted from

395   the supplementary material (Table S2) from Rheinbay et al.[17]. This included data from three patient

396   GBM tumours that had been used to derive glioma stem cells (GSCs) which were subsequently

397   cultured in serum to produce differentiated glioma cells (DGC). We required that all three samples

398   had been assigned the same status to be included in our analyses. Raw sequencing data was

399   downloaded on a further three GSC and DGC pairs from Patel et al.[19] via the Gene Expression

400   Omnibus (accession GSE57872). These data were processed and aligned, and pairwise differential

401   expression analysis was performed, exactly as for the paired primary and recurrent RNAseq data.

402   Quantification and assessment of the significance of overlap in genes with different promoter states

403   and/or differential expression with the JBSgenes or LE50 genes was done in the R statistical

404   package.

405

406   **LIST OF ABBREVIATIONS**

407   ChIP- Chromatin immunoprecipitation

408   DE – Differentially expressed

409   DGC – Differentiated glioma cell

410   GBM – Glioblastoma

411   GSC – Glioma stem cell

412   GSEA – Gene set enrichment analysis

413   H3K4me3 - Trimethylated histone 3 at lysine 4

414   H3K27me3 – Trimethylated histone 3 at lysine 27

415   JARID2 - Jumonji And AT-Rich Interaction Domain Containing 2

416   JBSgenes – JARID2 binding site genes

417   LE - Leading edge

418   NES – Normalized enrichment score

419   PRC2 – Polycomb Repressive Complex 2

420   RNAseq – RNA sequencing

421   TF – Transcription factor

422

DECLARATIONS

Ethics approval and consent to participate

All patients whose data was used in this study consented to the use of their tissue in research. The data generated for this study was used in accordance with ethical approval acquired from NHS NRES Committee South Central - Oxford A (REC 13/SC/0509).

Consent for publication

Not applicable

Availability of data and materials

The RNAseq datasets analysed during the current study are available in the following repositories: Rabadan samples are downloadable from the sequencing read archive (accession SRP074425); Verhaak and Nam samples are available following application to the European Genome-Phenome Archive (accession numbers EGAS00001001033 and EGAD00001001424, respectively). The RNAseq and DNA methylation datasets generated during the current study are not publicly available as data transfer agreements are required to ensure compliance with ethical approvals but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

This work was funded by grants from Leeds Charitable Foundation, now Leeds Cares (9R11/14-11 to LFS); Yorkshire Cancer Research, Brain Tumour Research and Support and Ellie's Fund (LPP072 to LFS); and personal fellowships awarded to LFS (funded by The Wellcome Trust and the University of Leeds) and AD (funded by UKRI, MR/S00386X/1). The funding bodies had no role in the design of the study, the collection, analysis, or interpretation of data or in writing the manuscript.

Authors' contributions

LFS devised the project. LFS and SS acquired funding. MDJ and AB sourced samples and provided clinical annotation. NR and AFB processed samples following annotation and diagnostic confirmation from AC and AI. LFS, GT, AD, MC and JW performed data analysis and LFS interpreted

451   it. NR performed experimental work and NR and CT captured and sequenced single cells. LFS wrote

452   the manuscript, which was reviewed and approved by all authors.

453   ***Acknowledgements***

465

REFERENCES

467   1.   Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, Hu
468   J, et al: **Whole-genome and multisector exome sequencing of primary and post-
469   treatment glioblastoma reveals patterns of tumor evolution.** *Genome Res* 2015, **25:**316-
470   327.
471   2.   Droop A, Bruns A, Tanner G, Rippaus N, Morton R, Harrison S, King H, Ashton K, Syed K,
472   Jenkinson MD, et al: **How to analyse the spatiotemporal tumour samples needed to
473   investigate cancer evolution: A case study using paired primary and recurrent
474   glioblastoma.** *Int J Cancer* 2017.
475   3.   Körber V, Yang J, Barah P, Wu Y, Stichel D, Gu Z, Fletcher MNC, Jones D, Hentschel B,
476   Lamszus K, et al: **Evolutionary Trajectories of IDHWT Glioblastomas Reveal a Common
477   Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis.** *Cancer Cell* 2019,
478   **35:**692-704.e612.
479   4.   Wang J, Cazzato E, Ladewig E, Frattini V, Rosenbloom DIS, Zairis S, Abate F, Liu Z, Elliott O,
480   Shin Y-J, et al: **Clonal evolution of glioblastoma under therapy.** *Nat Genet* 2016, **48:**768-
481   776.
482   5.   Kim J, Lee I-H, Cho Hee J, Park C-K, Jung Y-S, Kim Y, Nam So H, Kim Byung S, Johnson
483   Mark D, Kong D-S, et al: **Spatiotemporal Evolution of the Primary Glioblastoma Genome.**
484   *Cancer Cell* 2015, **28:**318-328.
485   6.   Aldape K, Amin SB, Ashley DM, Barnholtz-Sloan JS, Bates AJ, Beroukhim R, Bock C, Brat DJ,
486   Claus EB, Costello JF, et al: **Glioma Through the Looking GLASS: Molecular Evolution of
487   Diffuse Gliomas and the Glioma Longitudinal AnalySiS Consortium.** *Neuro-Oncology*
488   2018**:**noy020-noy020.

489   7.   Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, Bayani J, Pressey JC, Lionel AC,
490   Clarke ID, et al: **Single cell-derived clonal analysis of human glioblastoma links functional**
491   **and genomic heterogeneity.** *Proc Natl Acad Sci U S A* 2015, **112:**851-856.
492   8.   Lan X, Jorg DJ, Cavalli FMG, Richards LM, Nguyen LV, Vanner RJ, Guilhamon P, Lee L,
493   Kushida MM, Pellacani D, et al: **Fate mapping of human glioblastoma reveals an invariant**
494   **stem cell hierarchy.** *Nature* 2017, **549:**227-232.
495   9.   Al-Raawi D, Jones R, Wijesinghe S, Halsall J, Petric M, Roberts S, Hotchin NA, Kanhere A: **A**
496   **novel form of JARID2 is required for differentiation in lineage-committed cells.** *The EMBO*
497   *Journal* 2018:e98449.
498   10. Corley M, Kroll KL: **The roles and regulation of Polycomb complexes in neural**
499   **development.** *Cell and tissue research* 2015, **359:**65-85.
500   11. Natsume A, Ito M, Katsushima K, Ohka F, Hatanaka A, Shinjo K, Sato S, Takahashi S,
501   Ishikawa Y, Takeuchi I, et al: **Chromatin Regulator PRC2 Is a Key Regulator of Epigenetic**
502   **Plasticity in Glioblastoma.** *Cancer Research* 2013, **73:**4559.
503   12. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan G-C, Lee Y, Orkin SH:
504   **Jumonji Modulates Polycomb Activity and Self-Renewal versus Differentiation of Stem**
505   **Cells.** *Cell* 2009, **139:**1303-1314.
506   13. Wang X, Prager BC, Wu Q, Kim LJY, Gimple RC, Shi Y, Yang K, Morton AR, Zhou W, Zhu Z, et
507   al: **Reciprocal Signaling between Glioblastoma Stem Cells and Differentiated Tumor Cells**
508   **Promotes Malignant Progression.** *Cell Stem Cell* 2018, **22:**514-528.e515.
509   14. Suvà Mario L, Rheinbay E, Gillespie Shawn M, Patel Anoop P, Wakimoto H, Rabkin
510   Samuel D, Riggi N, Chi Andrew S, Cahill Daniel P, Nahed Brian V, et al: **Reconstructing and**
511   **Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells.** *Cell*
512   2014, **157:**580-594.
513   15. Podobinska M, Szablowska-Gadomska I, Augustyniak J, Sandvig I, Sandvig A, Buzanska L:
514   **Epigenetic Modulation of Stem Cells in Neurodevelopment: The Role of Methylation and**
515   **Acetylation.** *Frontiers in Cellular Neuroscience* 2017, **11:**23.
516   16. Yevshin I, Sharipov R, Valeev T, Kel A, Kolpakov F: **GTRD: a database of transcription factor**
517   **binding sites identified by ChIP-seq experiments.** *Nucleic Acids Research* 2017, **45:**D61-
518   D67.
519   17. Rheinbay E, Suvà Mario L, Gillespie Shawn M, Wakimoto H, Patel Anoop P, Shahid M, Oksuz
520   O, Rabkin Samuel D, Martuza Robert L, Rivera Miguel N, et al: **An Aberrant Transcription**
521   **Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma.**
522   *Cell Reports* 2013, **3:**1567-1579.
523   18. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig
524   M, Plath K, et al: **A Bivalent Chromatin Structure Marks Key Developmental Genes in**
525   **Embryonic Stem Cells.** *Cell* 2006, **125:**315-326.
526   19. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV,
527   Curry WT, Martuza RL, et al: **Single-cell RNA-seq highlights intratumoral heterogeneity in**
528   **primary glioblastoma.** *Science* 2014, **344:**1396.
529   20. Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, Pastorino S, Purow BW, Christopher
530   N, Zhang W, et al: **Tumor stem cells derived from glioblastomas cultured in bFGF and EGF**
531   **more closely mirror the phenotype and genotype of primary tumors than do serum-**
532   **cultured cell lines.** *Cancer Cell* 2006, **9:**391-403.
533   21. Kaneko S, Bonasio R, Saldaña-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg
534   D: **Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to**
535   **Chromatin.** *Molecular Cell* 2014, **53:**290-300.

22. Liau BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, Venteicher AS, Hebert CH, Carey CD, Rodig SJ, et al: **Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance.** *Cell Stem Cell* 2017, **20:**233-246.e237.

23. Chen J, Li Y, Yu T-S, McKay RM, Burns DK, Kernie SG, Parada LF: **A restricted cell population propagates glioblastoma growth after chemotherapy.** *Nature* 2012, **488:**522.

24. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, Henkelman RM, Cusimano MD, Dirks PB: **Identification of human brain tumour initiating cells.** *Nature* 2004, **432:**396-401.

25. Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, Brons NHC, Leite S, Sauvageot N, Sarkisjan D, et al: **Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment.** *Nature Communications* 2019, **10:**1787.

26. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li Y, et al: **Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment.** *Cancer Cell* 2017, **32:**42-56.e46.

27. Osswald M, Jung E, Sahm F, Solecki G, Venkataramani V, Blaes J, Weil S, Horstmann H, Wiestler B, Syed M, et al: **Brain tumour cells interconnect to a functional and resistant network.** *Nature* 2015, **528:**93.

28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2012, **29:**15-21.

29. Conway C, Graham JL, Chengot P, Daly C, Chalkley R, Ross L, Droop A, Rabbitts P, Stead LF: **Elucidating drivers of oral epithelial dysplasia formation and malignant transformation to cancer using RNAseq.** *Oncotarget; Vol 6, No 37* 2015.

30. Wang J, Duncan D, Shi Z, Zhang B: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.** *Nucleic Acids Research* 2013, **41:**W77-W83.

31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences* 2005, **102:**15545.

32. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R: **Full-length RNA-seq from single cells using Smart-seq2.** *Nature Protocols* 2014, **9:**171.

33. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell transcriptomic data across different conditions, technologies, and species.** *Nature Biotechnology* 2018, **36:**411.