# HHS Public Access

Author manuscript

*Nat Genet*. Author manuscript; available in PMC 2016 November 17.

# Recurrent somatic mutations in *POLR2A* define a distinct subset of meningiomas

**Victoria E Clark**[1,2,3], **Akdes Serin Harmancı**[1,2], **Hanwen Bai**[1,3,14], **Mark W Youngblood**[1,2,3,14], **Tong Ihn Lee**[4], **Jacob F Baranoski**[1,2,3], **A Gulhan Ercan-Sencicek**[2,5], **Brian J Abraham**[4], **Abraham S Weintraub**[4], **Denes Hnisz**[4], **Matthias Simon**[6], **Boris Krischek**[7], **E Zeynep Erson-Omay**[1,2], **Octavian Henegariu**[1,2,3,5,8], **Geneive Carrión-Grant**[1,2], **Ketu Mishra-Gorur**[1,2,3,5,8], **Daniel Durán**[1,2,3], **Johanna E Goldmann**[4], **Johannes Schramm**[9], **Roland Goldbrunner**[7], **Joseph M Piepmeier**[2], **Alexander O Vortmeyer**[10], **Jennifer Moliterno Günel**[1,2], **Kaya Bilgüvar**[1,3,11], **Katsuhito Yasuno**[1,2], **Richard A Young**[4,12], and **Murat Günel**[1,2,3,5,8,13]

[1]Yale Program in Brain Tumor Research, Yale School of Medicine, New Haven, Connecticut, USA

[2]Department of Neurosurgery, Yale School of Medicine, New Haven, Connecticut, USA

[3]Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA

[4]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA

[5]Yale Program on Neurogenetics, Yale School of Medicine, New Haven, Connecticut, USA

[6]Department of Neurosurgery, University of Bonn Medical School, Bonn, Germany

[7]Department of General Neurosurgery, University Hospital of Cologne, Cologne, Germany

Correspondence should be addressed to M.G. (murat.gunel@yale.edu).
[14]These authors contributed equally to this work

**Contributions**
V.E.C. performed whole-exome sequencing analysis, molecular inversion probe (MIP) sequencing and analysis, amplicon sequencing and analysis, chr22 copy-number assessment via qPCR, structural analysis of mutations, H3K27ac ChIP-seq of *POLR2A* mutant tumors, *TRAF7* splice mutation characterization, clinical correlations, and sample selection. V.E.C., M.W.Y., J.F.B., and D.D. performed Sanger confirmation of detected mutations for the 775-meningioma cohort and sample preparation. A.S.H. performed expression microarray analysis, CNV analysis, and RNA-seq variant calling. B.J.A. called super-enhancers on H3K27ac meningioma ChIP-seq samples. A.S.H., H.B., and V.E.C. performed differential super-enhancer binding analysis. H.B. performed predication analysis of microarray on meningioma samples. V.E.C. and M.W.Y. performed RNA-seq analysis. K.Y. created the bioinformatics pipelines used for DNA-seq experiments (whole-exome, MIPs, and amplicon). O.H. performed KLF4 luciferase experiments. T.I.L., A.G.E.-S., A.S.W., D.H., M.S., B.K., E.Z.E.-O., G.C.-G., K.M.-G., J.E.G., J.S., R.G., J.M.P., A.O.V., J.M.G., K.B., R.A.Y., and M.G. assisted with sample preparation and/or experimental design. V.E.C., M.W.Y., and M.G. wrote the manuscript, and A.S.H., H.B., T.I.L. and R.A.Y. edited the manuscript. M.G. designed and oversaw the project.

[8]Department of Neuroscience, Yale School of Medicine, New Haven, Connecticut, USA

[9]Medical Faculty, University of Bonn Medical School, Bonn, Germany

[10]Department of Pathology, Yale School of Medicine, New Haven, Connecticut, USA

[11]Yale Center for Genome Analysis, Yale School of Medicine, Orange, Connecticut, USA

[12]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[13]Yale Comprehensive Cancer Center, Yale School of Medicine, New Haven, Connecticut, USA

## Abstract

RNA polymerase II mediates the transcription of all protein-coding genes in eukaryotic cells, a process that is fundamental to life. Genomic mutations altering this enzyme have not previously been linked to any pathology in humans, which is a testament to its indispensable role in cell biology. On the basis of a combination of next-generation genomic analyses of 775 meningiomas, we report that recurrent somatic p.Gln403Lys or p.Leu438_His439del mutations in *POLR2A*, which encodes the catalytic subunit of RNA polymerase II (ref. 1), hijack this essential enzyme and drive neoplasia. *POLR2A* mutant tumors show dysregulation of key meningeal identity genes[2, 3], including *WNT6* and *ZIC1/ZIC4*. In addition to mutations in *POLR2A*, *NF2*, *SMARCB1, TRAF7, KLF4, AKT1, PIK3CA*, and *SMO*[4, 5, 6, 7, 8], we also report somatic mutations in *AKT3*, *PIK3R1*, *PRKAR1A*, and *SUFU* in meningiomas. Our results identify a role for essential transcriptional machinery in driving tumorigenesis and define mutually exclusive meningioma subgroups with distinct clinical and pathological features.

Transcriptional regulation is a precisely orchestrated process that involves a diverse collection of molecules ranging from large-scale chromatin remodelers to gene-specific modifiers. Alterations in components of transcriptional regulation are particularly prevalent in neoplasia; such alterations include histone H3 mutations in pediatric glioblastoma[9], mediator subunit mutations in uterine leiomyosarcoma[10], and mutations of specific transcription factors in multiple cancer types[11]. Transcriptional machinery cooperatively acts to attract the multimeric enzyme RNA polymerase II (Pol II) to selective genomic sites, ultimately driving transcription of all mRNAs and most micro- and small nuclear RNAs. In this capacity, Pol II coordinates virtually every component of eukaryotic transcription, from pre-initiation to splicing and other post-transcriptional modifications. Loss of Pol II activity is incompatible with life, such that the ingestion of the potent Pol II inhibitor α-amanitin is fatal[12]. Given the essential biological functions of Pol II, it is not surprising that no mutations affecting any of its components have ever been reported in human disease.

Meningiomas, which arise from the membranes surrounding the brain and spinal cord, are the most common intracranial tumors[13]. We and others have previously identified somatic mutations in seven genes (*NF2* with co-occurring mutations in *SMARCB1*, *TRAF7* with a recurrent mutation in *KLF4* or mutations affecting PI3K signaling through *AKT1* or *PIK3CA* mutations, and *SMO*) and/or loss of chromosome 22 (chr22) to define mutually exclusive subgroups[4, 5, 6, 7, 8]. To discover the genetic basis of the remaining mutation-unknown tumors, we used iterative, integrated genomic approaches on a set of 775

meningiomas, including 485 primary, previously unradiated grade I tumors (Supplementary Table 1).

Focusing on samples that lacked known meningioma driver genes, we discovered recurrent mutations in *POLR2A* (polymerase (RNA) II (DNA directed) polypeptide A, 220 kDa), which encodes RPB1, the largest and catalytic subunit of Pol II. Recurrent *POLR2A* p.Gln403Lys ($n = 19$) or p.Leu438_His439del ($n = 4$) mutations accounted for approximately 6% of all benign cases. *POLR2A* mutations affected exon 7 and localized near each other in the highly conserved dock domain, which mediates interactions between Pol II and TFIIB during formation of the pre-initiation complex (Fig. 1a,b). These alterations were confirmed as somatic in all samples with available blood pairing and were mutually exclusive with previously established drivers[4] ($P = 3.02 \times 10^{-10}$, one-sided Fisher's exact test) (Fig. 1c,d). Whole-genome sequencing of three *POLR2A* meningioma–blood pairs did not identify recurrent chromosomal translocations (Fig. 1e, Supplementary Fig. 1, and Supplementary Table 2). Indeed, similar to other benign meningiomas, most *POLR2A* mutant tumors (88.2%) were genomically stable without detectable large-scale chromosomal amplifications or deletions (average percentage of the genome altered: 1.2%, 4.7%, and 0.6% in *POLR2A*, *NF2*, and non-*NF2* mutant meningiomas, respectively) (Fig. 1f and Supplementary Table 1).

*POLR2A* mutant meningiomas had a small number of rare, protein-altering somatic mutations (mean, 0.17 per Mb of sequencing; range, 0.03–0.30); this number was not significantly different from that in tumors harboring mutations in established meningioma genes (mean, 0.23 per Mb of sequencing; $P = 0.08$, *POLR2A* versus others for deleterious somatic mutations; $P = 0.82$ for all somatic mutations; two-sided Student's *t*-test) (Fig. 1g, Supplementary Table 3, and Supplementary Fig. 2). Furthermore, *POLR2A* mutations and mutations in other meningioma driver genes were found to be mutually exclusive, and no additional genes were found to be co-mutant in more than two exome-sequenced *POLR2A*-mutant samples (Supplementary Tables 1 and 3). Clinically, *POLR2A*-mutant tumors harbored distinct characteristics, including meningothelial histology ($P = 3.62 \times 10^{-4}$, Fisher's exact test) and a tendency to originate from the tuberculum sellae region ($P = 1.51 \times 10^{-5}$, Fisher's exact test), where the pituitary gland is located (Fig. 1h and Supplementary Table 4). *POLR2A* mutations were exclusively detected in WHO grade I (benign) meningiomas, with no events found in 159 grade II or grade III (atypical or malignant) meningiomas screened for these variants. The homogeneous clinical features and genomic stability of this group suggests that recurrent *POLR2A* mutations classify a unique subset of benign, but not atypical or malignant, meningiomas.

Among the remaining mutation-unknown samples in our cohort, we also identified somatic alterations in several other genes whose role in meningioma was not fully appreciated. Among *NF2*-mutant meningiomas, we found that a surprisingly large number of samples (6.4%) harbored recurrent p.Arg383Gln or p.Arg386His missense mutations affecting *SMARCB1* (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1) ($n = 2$ or 14, respectively, out of 250 *NF2*-mutant meningiomas screened). Although our group and others have noted somatic alterations in *SMARCB1* in meningiomas[4, 7], their recurrent nature, frequency, and co-occurrence with somatic *NF2*

mutations were not previously appreciated. Interestingly, most of the *NF2/SMARCB1* mutant samples originated from the midline dura (falx) of the anterior convexity (Supplementary Tables 1 and 5), a region in which *NF2* tumors lacking *SMARCB1* mutations are rarely found (70% versus 16%, respectively; $P = 1.75 \times 10^{-4}$, two-sided Fisher's exact test; odds ratio = 13.08; 95% confidence interval = 2.83–83.09 for *SMARCB1/NF2* versus *NF2* localizing to anterior falx).

In the *TRAF7* meningioma subgroup, we identified an intronic hot spot that harbored somatic mutations affecting the interval just upstream of the first WD40 repeat domain ($n = 12$) (Fig. 2a,b). Mutations in this region alter the splicing of exons 10 and 11, and result in the inclusion of a 28-amino-acid in-frame insertion and a Ser-to-Cys missense mutation (Fig. 2c and Supplementary Fig. 3). We also discovered co-mutations occurring with *TRAF7* that activate PI3K signaling, including somatic hot-spot mutations in *PIK3CA* (PI3K catalytic subunit, alpha) that were previously noted by our group and others[4, 8] (Fig. 2d), in *PIK3R1* (PI3K regulatory subunit, alpha), and in *AKT3* (v-akt murine thymoma viral oncogene homolog 3). The somatic AKT3 p.Glu17Lys variant observed in one whole-exome-sequenced meningioma was previously characterized to activate AKT signaling in melanoma[14] (Supplementary Table 1). These mutations were mutually exclusive with each other and also with the previously described AKT1 p.Glu17Lys recurrent mutation. Overall, among primary and benign meningiomas, which localized primarily to the anterior skull base or convexity regions and were screened for mutations in *PIK3CA*, mutations in *PIK3R1*, and the recurrent AKT1 p.Glu17Lys variant ($n = 333$), we identified PI3K pathway mutations in 19.22% of tumors ($n = 64$; 42 with AKT1 p.Glu17Lys mutations, 20 with *PIK3CA*, and 2 with *PIK3R1*) (Supplementary Table 1).

We also identified new alterations in genes associated with Sonic hedgehog (SHH) signaling, including loss-of-function mutations in *SUFU* (suppressor of fused homolog (*Drosophila*)) ($n = 5$) and a recurrent p.Ala17Asp missense mutation in *PRKAR1A* (protein kinase, CAMP-dependent, regulatory, type I, alpha) ($n = 3$) (Supplementary Fig. 4). *SUFU* is a known tumor suppressor gene and acts downstream of *SMO* to inhibit SHH signaling. A germline *SUFU* mutation (p.Arg123Cys) with somatic loss of the wild-type allele has previously been described in a family with multiple meningiomas[15], but somatic *SUFU* mutations have not been reported in sporadic meningiomas. Similarly, *PRKAR1A* mutations have not previously been reported in meningiomas but have been identified in multiple neoplasia syndrome Carney complex[16]. Inactivation of *PRKAR1A* has been shown to result in constitutive PKA activation, with ultimate effects on the balance of SHH signaling[17].

Somatic mutations in *NF2*, *TRAF7* (co-occurring with either the KLF4 p.Lys409Gln or PI3K pathway mutations), genes in the Hedgehog pathway, and *POLR2A* were all mutually exclusive, defining distinct meningioma subgroups (Fig. 2e). With these subtypes identified (Fig. 2f), we next aimed to gain insight into the mechanisms by which these driver genes lead to the formation of meningiomas. We initially focused on *POLR2A* mutant tumors. The Pol II holoenzyme mediates multiple essential cellular processes, such that mutations affecting its subunits can potentially drive neoplasia through a range of biological mechanisms. The dock domain of RPB1 is necessary for bridging among Pol II, TFIIB, and TFIID (which contains the TATA-box binding protein) during transcription initiation. The

fidelity of this interaction ensures that Pol II's catalytic subunit is correctly positioned at the DNA promoter, and defects could conceivably affect splicing, transcript accuracy, DNA repair, gene expression, or other essential processes. RNA sequencing of *POLR2A* meningiomas did not identify significant changes in splicing, differential promoter usage, or transcription errors (Supplementary Tables 6 and 7 and Supplementary Fig. 2). We also did not identify a significant difference in the total number of somatic coding and noncoding mutations between *POLR2A*-mutant tumors and others ($P = 0.82$; Fig. 1f and Supplementary Fig. 2), suggesting that defects in DNA repair are unlikely to underlie these tumors. We next used CRISPR–Cas9 gene editing to create knock-in models of the two *POLR2A* dock domain mutations in mouse embryonic stem cells and again did not identify any significant difference in splicing or differential promoter use (Supplementary Tables 8 and 9) or any evidence supporting global transcriptional amplification[18] (Supplementary Fig. 5).

We next carried out expression profiling across a large cohort of meningiomas ($n = 79$). We identified distinct clusters that were almost completely consistent with our established subgroups, suggesting that the transcriptional architecture of each tumor was driven largely by the underlying driver mutation (Fig. 3 and Supplementary Fig. 6). We identified subsets of genes and pathways that best characterized each subgroup, and we found that WNT signaling and differential expression of key meningeal transcription factors were associated with several meningioma drivers (with the use of prediction analysis of microarray[19]; see Online Methods) (Fig. 3a,c and Supplementary Table 10).

We next correlated the gene expression patterns with active super-enhancers, which are broad genomic regions that harbor dense clustering of histone 3 lysine 27 acetylation (H3K27ac) signals[20, 21, 22]. We identified 31 super-enhancers with concordant changes in gene expression that were more active in all meningioma subgroups ($n = 20$ in total) as compared to dura ($n = 2$), including the transcriptional regulators *FOXP1*, *FOXP4*, *HDAC4*, and *KLF2* (Supplementary Table 11 and Supplementary Fig. 7).

We also identified meningioma subtype-specific super-enhancers (Fig. 4 and Supplementary Table 12). In the *POLR2A* subgroup, a super-enhancer covering the *WNT6* and *WNT10A* (wingless-type MMTV integration site family, members 6 and 10A) gene locus was differentially activated relative to other groups (false discovery rate (FDR) = $1.32 \times 10^{-3}$; $\log_2$(fold change) = 1.58), with *WNT6* also being significantly overexpressed (FDR = $1.97 \times 10^{-11}$; $\log_2$(fold change) = 2.00) (Fig. 4b,e). In contrast, a super-enhancer encompassing *ZIC1* (Zic family member 1), a known meningeal identity gene[2], and *ZIC4* (Zic family member 4) was lost in *POLR2A*-mutant meningiomas (FDR = $5.45 \times 10^{-3}$; $\log_2$(fold change) = −1.72), and both genes were significantly downregulated (Fig. 4b,f) (FDR = $6.49 \times 10^{-5}$ and $3.92 \times 10^{-8}$, respectively; $\log_2$(fold change) = −0.44 and −0.92, respectively). These results are informative in light of the known roles of WNT and ZIC signaling in the growth and differentiation of meningeal progenitors[2, 3] (Fig. 4g).

Similarly, in the *NF2*/*SMARCB1* mutant subgroup, the *BCL2* (B cell CLL/lymphoma 2) oncogene and the developmentally important WNT-signaling regulators *DAAM2* (dishevelled associated activator of morphogenesis 2), *NKD1* (naked cuticle homolog 1),

*FZD4* (frizzled class receptor 4), and *SFRP2* (secreted frizzled-related protein 2) were expressed in significantly higher amounts and had more highly bound super-enhancers (Fig. 4c, Supplementary Fig. 8, and Supplementary Table 13).

For the *KLF4*/*TRAF7* subgroup, given that wild-type KLF4 has a role in super-enhancer formation in embryonic stem cells[20], we hypothesized that mutant KLF4 might directly alter super-enhancer binding. Indeed, we found that the KLF4 p.Lys409Gln mutant bound to a previously unreported consensus site (Fig. 4h) that was significantly enriched in the *KLF4*/*TRAF7*-group-specific super-enhancers as compared to background ($P = 0.001$), leading to activation of downstream gene expression (Fig. 4i). The developmentally important transcription-factor-encoding gene *GRHL3* (grainyhead-like 3 (*Drosophila*)), which is fundamental in neural tube development[23] (FDR = $1.37 \times 10^{-10}$, $\log_2$(fold change) = 1.09), had higher super-enhancer binding and expression in *KLF4*/*TRAF7* co-mutant tumors compared with the other four meningioma subgroups (Fig. 4d and Supplementary Fig. 8).

Comprehensive genomic analysis of benign meningioma has yielded fascinating insights about the biology of this disease. We have discovered a precarious balance that exists in fundamental transcriptional machinery, whereby a single mutation can transform physiologic activity into a driver of neoplasia. Using gene expression and super-enhancer analysis, we found that alterations in *POLR2A* result in dysregulation of key identity genes in meningeal development[2, 3]. Indeed, a pan-analysis of meningioma drivers showed recurrent involvement of embryonic signaling mechanisms that are fundamental to formation of the meninges. This interesting finding raises the possibility that benign meningioma may arise from early progenitor cells as a result of developmental dysregulation or, alternatively, induced dedifferentiation of mature cells into a stem-like state. Further experiments are needed to answer these questions.

## Methods

### Clinical materials

Institutional review board approvals for genetic studies, along with written consent from all study subjects, were obtained at the participating institutions. Formal pathology reports along with pre-operative magnetic resonance imaging studies were collected from participants.

### Selective tissue dissection

For each frozen specimen submitted for whole-exome sequencing, sections were re-reviewed to confirm the diagnosis and assess the adequacy of the frozen tissue for experimental analysis. On hematoxylin-and-eosin-stained sections from frozen tissue blocks, areas of interest were identified and microscopically dissected to ensure that each sample consisted of >70% tumor cells; unwanted regions such as inflammatory and necrotic areas were excluded. Tumors in the replication cohorts did not undergo selective tissue dissection. DNA/RNA was prepared using the All prep DNA/RNA mini kit (Qiagen, 80004) with the assistance of a QIAcube (Qiagen, 9001292).

## Whole-genome genotyping

The Illumina platform was used for whole-genome genotyping and analysis of the samples ($n$ = 94). Human OmniExpress-12v1.0 BeadChips containing 733,202 markers were used according to the manufacturer's protocol (Illumina, WG-315-1101). All samples had an overall genotype call rate greater than 98%. We detected copy-number alterations by comparing the normalized signal intensity between tumor and matched blood or tumor and the average of all blood samples. Segmentation was performed on log intensity (R) ratios using the DNA Copy algorithm[26]. Large-scale chromosomal deletion or amplification was defined as affecting more than one-third of the chromosomal arm with an accompanying log ratio of signal intensities < −0.1 or > 0.1 and B allele frequencies at heterozygous sites deviating from 0.5 by at least 0.05 units. Large-scale copy-neutral loss of heterozygosity was defined similarly, with the exception of the log ratio of signal intensities being between −0.1 and 0.1.

## Exome capture and sequencing

Nimblegen/Roche human solution-capture exome array (Roche NimbleGen, 05860504001) was used to capture the exomes of blood and tumor samples according to the manufacturer's protocol with modifications[27]. We sequenced the library on Illumina HiSeq instruments using paired-end 74-bp reads by multiplexing two tumor samples or three blood samples per lane.

## Exome sequence analysis

We analyzed sequence reads that passed the Illumina quality filter. Before alignment, we trimmed the low-quality 3′ end of the reads and filtered reads with low-complexity sequences and overall low base qualities (FASTX-Toolkit). We trimmed PCR primer-contaminated sequence segments (cutadapt version 0.9.5) and kept sequences for further analysis only if both reads in a pair had more than 35 bases remaining after the above trimming and filtering quality measures. We aligned reads to the human genome reference sequence (version GRCh37, the same as used in phase 1 of the 1000 Genomes Project) using Stampy (version 1.0.16)[28] in a hybrid mode with BWA (version 0.5.9-r16)[29]. PCR duplicates were flagged using the MarkDuplicates algorithm from Picard and excluded from further analysis as previously described[30]. Metrics describing the alignment quality, such as the fraction of unique read pairs, the fraction of aligned reads, and depth of coverage, were collected for each sample with the CollectAlignmentSummaryMetrics and CalculateHsMetrics utilities of Picard (Supplementary Table 14). We performed multi-sequence local realignment around putative and known insertion/deletion (indel) sites, where the reads from tumor and matched blood samples were analyzed at the same time. This was followed by the base quality score recalibration using the Genome Analysis Toolkit (GATK, version 1.5–20)[30]. We detected variant sites (point mutations and small indels) for tumor and matched blood pairs using the UnifiedGenotyper algorithm from GATK[30]. Variants were concurrently called for unpaired tumor samples using UnifiedGenotyper. To identify somatic events, we used a genotype-likelihood-based somatic score proposed by Li[31]. We used the same phred-scaled score as implemented in bcf tools[32]. We considered a mutation to be somatic if the matched blood sample was homozygous reference while the tumor sample

was heterozygous non-reference and if the somatic score was 50. We filtered variant sites by assessing the allelic and overall depth of coverage, the strand, mapping quality and read position biases between reference and non-reference allele supporting reads, the consistency between the number of alleles and that of haplotypes, the root mean square of the mapping quality, the coverage-adjusted variant quality, and the number of mapping-quality-zero reads. For insertions, we used the maximum score of flanking bases, and for deletions, we used the maximum score of deleted bases. Finally, we annotated variant alleles using the Ensembl database (version 66) with the help of the Variant Effect Predictor (v2.4) tool. From these functional annotations, we selected the most deleterious consequence for each variant site and considered the variant allele to be deleterious if its consequence was annotated as stop gained, complex change in transcript, frameshift, splice acceptor/donor sites, stop lost, non synonymous codon predicted to be deleterious/damaging, or in-frame codon loss/gain. We initially considered genes that were mutated in more than one tumor and prioritized genes on the basis of the ratio of deleterious to non-deleterious mutations observed. The co-occurrence and mutual exclusivity of genes were assessed using one-sided Fisher's exact test[33]. To leverage whole-exome sequencing data to detect large-scale chromosomal amplifications or deletions, we used the exomeCNV algorithm[34].

### Whole-genome sequencing and breakpoint analysis

After quality control and quantification of genomic DNA, whole-genome sequencing libraries were prepared using the Illumina TruSeq DNA PCR-Free sample preparation kit (Illumina, FC-121-3001) according to the manufacturer's protocol. Sequencing was performed on the Illumina Hiseq 4000 platform using paired-end chemistry at a read length of 100 bp. Sequencing reads were processed using GATK Best Practices. Break dancer[35] was used to call breakpoints, with multi-sample calling mode used to assess the somatic status of breakpoints. Intrachromosomal breakpoints were filtered if (i) breakpoints were less than 200 bp apart, (ii) there were any supporting reads from the matching blood, (iii) there were fewer than eight supporting reads in the tumor, or (iv) breakpoints had a quality score of <40. Breakpoints were then annotated using ANNOVAR for hg19 refSeq genes, cyto-band information, segmental duplications, and repeat masked regions and were subsequently filtered if there was overlap with segmental duplications or any repeat elements. Breakpoints with >25 reads are shown in Supplementary Table 2. Breakpoints with >25 supporting reads were plotted in Circos plots (Supplementary Fig. 1).

### Custom molecular inversion probe (MIP) sequencing and analysis

Targeted sequencing of exomic regions and exon–intron boundaries of *NF2*, *SMARCB1*, *TRAF7*, *PIK3CA*, *PIK3R1*, *PRKAR1A*, *SMO*, and *SUFU* plus the recurrent variants AKT1 p.Glu17Lys and KLF4 p.Lys409Gln was performed using MIPs. Probes were designed and libraries were generated as previously described, with slight modifications[36]. We carried out multiplex capture of targeted sequences using 150 ng of genomic DNA for each sample through probe hybridization, gap filling and ligation, and exonuclease digestion. The captured DNA was amplified using barcoded primers. Libraries were cleaned up with AMPure XP beads (Beckman Coulter, A63880), quantified by picogreen, and pooled with 80 meningiomas per lane. Pooled libraries were sequenced on the HiSeq 2500 system using paired-end 101-bp reads according to the manufacturer's protocol. Sequence reads were

processed like the whole-exome data as described previously, but without duplication removal, indel realignment, or base quality recalibration[4]. Variants were called using GATK's UnifiedGenotyper.

### Custom amplicon sequencing and analysis

Libraries consisting of the coding exons from *TRAF7*, *NF2*, *SMO*, and the recurrent mutations for AKT1 p.Glu17Lys and KLF4 p.Lys409Gln were created using the TargetRich custom amplicon kit (Kailos Genetics). We pooled and sequenced 48 samples on a version 2 Illumina MiSeq using paired-end 150-bp reads. Sequence reads were processed like the whole-exome data, but without duplication removal and base quality recalibration. Variants were called for all samples using GATK's UnifiedGenotyper.

### Sanger sequencing

Coding variants detected by whole-exome sequencing or targeted next-generation sequencing were confirmed by Sanger sequencing using standard protocols[27]. In addition to whole-exome-sequenced tumors ($n = 107$), we screened *POLR2A* exon 7, containing the recurrent dock domain mutations, by Sanger sequencing in 452 tumors. Amplicons were cycle-sequenced on ABI's 9800 Fast Thermocyclers and analyzed as previously described[4].

### Chromosome 22 quantification by qPCR

We assessed chr22 loss by quantitative real-time PCR (qRT-PCR) using Fast SYBR Green Master Mix (Thermo Fisher, 438561). For each sample, two exons in *NF2* were used for quantification and normalized against primers on chromosomes 11 and 16. Samples and controls were run in triplicate. Dissociation curves were generated to ensure primer specificity. To determine the threshold cycle, we used female reference DNA (Promega, G1521), diluted at 22 ng/μl, for four serial dilutions from 1/4-fold to 1/256-fold. Each qPCR run was compared to three samples: a commercially available reference female DNA, DNA from a whole-genome-genotyped meningioma sample with chr22 loss (MN-290), and DNA from a whole-genome-genotyped meningioma sample with intact chr22 (MN-1). We considered a ratio of <0.7 as indicative of a loss and a ratio of >1.3 as indicative of a gain.

### Structural analysis

Figures were created using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.1, Schrödinger, LLC). We used the following PDB IDs: PDB ID 2WBU for KLF4; PDB ID 2OVQ for FBXW7; and PDB ID 1R5U for yeast RPB1. The Clustal Omega algorithm was used for alignment of TRAF7 WD40 repeat domains (Fig. 2b).

### RNA-seq and analysis

Meningioma mRNA was extracted using the Allprep DNA/RNA Mini Kit. RNA integrity number (RIN) was assessed using an Agilent Bioanalyzer 2100, and samples with RIN < 5.7 were excluded. Ribozero depletion was performed, and samples were loaded four per HiSeq 2500 lane (paired-end, 75-bp sequencing). RNA-seq reads passing the Illumina quality filter underwent read pre-processing as described in the section "Exome sequence analysis" and were analyzed using the Tuxedo suite of tools[37, 38]. Briefly, reads were aligned using

TopHat2, transcripts were assembled using Cufflinks, a merged assembly file was created using Cuffmerge, abundance calculations were carried out using Cuffquant, differential expression was analyzed using Cuffdiff, and CummeRbund was used to visualize the Cuffdiff output. To visualize RNA-seq reads spanning *TRAF7* intron 10, we visualized aligned BAM files using the Integrative Genomics Viewer[39].

### Microarray data analysis

We used the Illumina HumanHT12.v4 BeadChip array (Illumina, BD-103-0204) on samples collected from ten different batches. Data were normalized using normal–exponential convolution model–based background correction and quantile normalization using the limma R package[40]. We normalized all the batches at once after excluding probes with low quality. Samples estimated to have a zero proportion of expressed probes, mean signal intensity less than 5.54 or an RIN value less than 3, were excluded. Hierarchical clustering of the gene expression data showed batch and chip effects in the data. The batch and chip effects were removed using ComBat in sva R package[41]. We performed unsupervised hierarchical clustering based on a Euclidean distance metric and average linking clustering on the probes. Differentially expressed genes were identified using an empirical Bayesian method, ebayes, implemented in limma. A log-odds score of 2 and an absolute log fold change of 1 were used to define the threshold. MicroRNA was not analyzed.

### CRISPR–Cas9 gene editing of v6.5 mouse embryonic stem cells

To model the *POLR2A* missense mutations, we established 20-nt CRISPR guide sequences with BbsI overhang. For creation of the repair template, gBlocks (Integrated DNA Technologies) were designed to carry the desired mutation with 750 bp of flanking wild-type genomic sequence plus the addition of restriction sites (SacI and SphI) to the ends of the gBlocks. To perform targeted gene editing, we transiently transfected v6.5 mouse embryonic stem cells (mESCs) using Xfect; GFP$^+$ cells were sorted by flow cytometry at 26 h post-transfection, and individual colonies were picked, trypsinized, cultured, and sequenced to confirm successful generation of a knock-in genotype. *POLR2A* dock domain mutations were validated by Sanger sequencing.

### FPKM calculation of RNA-seq data from *POLR2A* mutant and wild-type v6.5 mESCs

RNA was extracted from an equal number of cells harvested in triplicate from three independent cell lines: two independent *POLR2A*$^{Q403K}$ lines (A7 and B10) generated by CRISPR–Cas9 gene editing, and wild-type v6.5 mESCs[42]. Briefly, total RNA was purified from equal numbers of cells using the mirVANA miRNA isolation kit (Ambion, AM1560), spiked in with External RNA Controls Consortium ExFold Mix #1 RNA spike-in controls, digested with DNA-free DNase I, and processed for RNA-seq as described above. The RNA-seq data were analyzed as described above; then the FPKM (fragments per kilobase of transcript per million mapped reads) values were re-normalized using the spike-in controls with a LOESS regression. Transcripts with an FPKM value of 1 or greater in wild-type cells were selected, and the density of transcripts with a given FPKM value was plotted.

## Meningioma gene expression subgroups

'Partitioning around medoids' clustering was performed on the top 1,000 most variably expressed genes to identify robust meningioma clusters, using the 'pam' function of the 'cluster' package in R. Sample silhouette widths were also calculated using the 'silhouette' function. To determine the optimal number of gene expression clusters, clustering was repeated for 2–15 total clusters. When five clusters were formed, the mean sample silhouette width was the highest, and therefore we chose this number of clusters for downstream analysis. To identify the 'core' samples of each cluster, we further excluded samples with negative silhouette widths, retaining 79 tumor samples[43, 44, 45]. In addition, Fisher's exact tests were used to test the associations between gene expression clusters and mutational profiles. Cluster names were determined accordingly. Lastly, to assess clustering robustness, we used four additional clustering methods on the same set of 79 samples, including consensus clustering (ConsensusClusterPlus package), non-negative matrix factorization (NMF package), hierarchical clustering (cluster package), and principal component analysis (cmdscale function)[46, 47]. All five clustering algorithms consistently identified the five meningioma subgroups.

## Meningioma signature gene identification

First, we carried out significance analysis of microarray to identify significantly differentially expressed genes among the five meningioma subgroups, using the samr package in R[44, 45, 48]. Second, we carried out prediction analysis of microarray (PAM) to discover gene signatures specific to each of the meningioma subgroups, using the pamr package in R[19, 44, 45, 49]. The PAM scores reflected positive or negative correlations between gene expression and each subgroup[19]. A total of 5,641 signature genes were selected (median FDR = 0.0042; 95th percentile FDR = 0.013). In addition, each signature gene was assigned to the subgroup with the highest absolute PAM score. Lastly, functional enrichment of subgroup-specific genes was performed using DAVID[50].

## H3K27ac ChIP-seq

Chromatin immunoprecipitation followed by sequencing (ChIP-seq) was performed as previously described, with minor adjustments[51]. H3K27ac ChIP and input samples were sequenced for each tumor ($1 \times 75$ bp, HiSeq 2000). Reads were aligned uniquely with Bowtie (0.12.7)[52] to the human genome (hg19), and regions of enrichment were identified with MACS (v1.4).

## Meningioma super-enhancer analysis

H3K27ac ChIP and input samples were sequenced for each tumor ($1 \times 75$ bp, HiSeq 2000). Reads were aligned uniquely with Bowtie (0.12.7)[52] to the human genome (hg19), and regions of enrichment were identified with MACS (v1.4)[53]. We filtered reads by removing those that contacted RNA repeats from RepeatMasker by at least 1 bp using BEDtools. Super-enhancers were identified as described by Chipumoro *et al.*[54] using these reads. Super-enhancers were separated from typical enhancers using ROSE2 with parameters -s (stitching) 12500, -t (promoter exclusion zone) 2000. We used DiffBind to estimate the

significance of super-enhancer read changes (adjusted $P$ value threshold = 0.05) between meningioma subtypes[55] (Supplementary Table 12).

### Super-enhancer and gene expression concordance analyses

We undertook several complementary approaches to calculate super-enhancer and gene expression concordance. We initially considered the union of super-enhancers from all meningioma and dura samples, merging overlapping super-enhancer regions and thereby creating consensus super-enhancer sets. We then identified super-enhancers common to all meningioma subgroups. To detect meningioma-specific super-enhancers, we excluded super-enhancers that were present in all dura samples. Super-enhancer regions were annotated using the GREAT Annotation tool. Finally, we evaluated the concordance between super-enhancers and gene expression changes. We identified meningioma-specific super-enhancer-driven genes as those with meningioma-specific super-enhancers and increased expression in meningiomas as compared with dura samples (differential expression log-odds score > 2) (Supplementary Table 11).

In parallel (Supplementary Table 13), for each meningioma subgroup or dura sample, overlapping super-enhancers were merged. Super-enhancers identified in at least two samples of the same subgroup were regarded as a 'consensus' subset and were retained for subsequent analysis. In addition, consensus super-enhancers in meningiomas that overlapped with those of dura were excluded, leaving a set of meningioma-specific super-enhancers. Subgroup-specific super-enhancers were further identified using the same approach as described above.

### KLF4 ChIP

Lipofectamine 2000 was used to transfect 293T cells with wild-type KLF4, mutant KLF4, or empty vector, according to standard manufacturer protocols. At 48 h post-transfection, the cells were harvested for ChIP. Briefly, $5 \times 10^7$ cells were cross-linked with 1% formaldehyde, treated with glycine to halt fixation, and washed with PBS. Nuclei were harvested and resuspended in nuclear lysis buffer containing 0.1% SDS. Chromatin was sheared by sonication with a Bioruptor Standard sonication device with a 1.5-ml tube holder (30 min total, 30 s pulses, 30 s rest). Soluble chromatin was incubated with protein G Dynabeads (100 μl) coated with 10 μg of anti-KLF4 (AF3640, R&D Systems) overnight at 4 °C. To control for enrichment, we held back 10% of the chromatin for input. The chromatin was precipitated magnetically, washed, and eluted from the antibody-coated beads. Cross-links were reversed by overnight incubation at 65 °C, after which the DNA was purified and quantified using PicoGreen. Enrichment was verified by ChIP-qPCR using relative (  Ct) quantification. Libraries were prepared for next-generation sequencing on the Illumina platform and sequenced (1 × 75 bp, HiSeq 2500).

### KLF4 ChIP-seq data analysis

For ChIP-seq data analysis, Cutadapt was used to remove sequencing adaptors, FASTX-Toolkit was used for read quality trimming, bowtie1 using the m = 1 option was used to align uniquely mapping reads, PICARD was used to mark and remove PCR duplicates, SPP/PhantomPeakQualTools was used for cross-correlation analysis, MACS2 was used for peak

calling, GREAT was used for peak annotation, and MEME-ChIP was used for motif analysis.

### Enrichment of novel motif in KLF4-specific super-enhancer active genes

We first searched the novel KLF4 consensus motif against the human genome (hg19) using the TOMTOM tool[56]. Then we generated the background KLF4 consensus motif regions based on $n = 1,000$ permutations. We recalculated the overlapping regions between the permuted regions and KLF4-specific super-enhancer active genes. The permutation $P$ value was calculated as the proportion of overlapping regions that was more than what we observed. We also performed a similar permutation test to assess the enrichment between the KLF4 consensus motif regions and other group-specific active super-enhancers.

## Acknowledgments

## References

1. Wintzerith M, Acker J, Vicaire S, Vigneron M, Kedinger C. Complete sequence of the human RNA polymerase II largest subunit. Nucleic Acids Res. 1992; 20:910. [PubMed: 1542581]

2. Inoue T, Ogawa M, Mikoshiba K, Aruga J. Zic deficiency in the cortical marginal zone and meninges results in cortical lamination defects resembling those in type II lissencephaly. J Neurosci. 2008; 28:4712–4725. [PubMed: 18448648]

3. Choe Y, Zarbalis KS, Pleasure SJ. Neural crest-derived mesenchymal cells require Wnt signaling for their development and drive invagination of the telencephalic midline. PLoS One. 2014; 9:e86025. [PubMed: 24516524]

4. Clark VE, et al. Genomic analysis of non-*NF2* meningiomas reveals mutations in *TRAF7, KLF4, AKT1*, and *SMO*. Science. 2013; 339:1077–1080. [PubMed: 23348505]

5. Brastianos PK, et al. Genomic sequencing of meningiomas identifies oncogenic *SMO* and *AKT1* mutations. Nat Genet. 2013; 45:285–289. [PubMed: 23334667]

6. Reuss DE, et al. Secretory meningiomas are defined by combined *KLF4* K409Q and *TRAF7* mutations. Acta Neuropathol. 2013; 125:351–358. [PubMed: 23404370]

7. Schmitz U, et al. INI1 mutations in meningiomas at a potential hotspot in exon 9. Br J Cancer. 2001; 84:199–201. [PubMed: 11161377]

8. Abedalthagafi M, et al. Oncogenic PI3K mutations are as common as *AKT1* and *SMO* mutations in meningioma. Neuro-oncol. 2016; 18:649–655. [PubMed: 26826201]

9. Schwartzentruber J, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012; 482:226–231. [PubMed: 22286061]

10. Kämpjärvi K, et al. Somatic *MED12* mutations in uterine leiomyosarcoma and colorectal cancer. Br J Cancer. 2012; 107:1761–1765. [PubMed: 23132392]

11. Bhagwat AS, Vakoc CR. Targeting transcription factors in cancer. Trends Cancer. 2015; 1:53–65. [PubMed: 26645049]

12. Santi L, et al. Acute liver failure caused by *Amanita phalloides* poisoning. Int J Hepatol. 2012; 2012:487480. [PubMed: 22811920]

13. Ostrom QT, et al. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006–2010. Neuro-oncol. 2013; 15:ii1–ii56. [PubMed: 24137015]

14. Davies MA, et al. A novel *AKT3* mutation in melanoma tumours and cell lines. Br J Cancer. 2008; 99:1265–1268. [PubMed: 18813315]

15. Aavikko M, et al. Loss of SUFU function in familial multiple meningioma. Am J Hum Genet. 2012; 91:520–526. [PubMed: 22958902]

16. Kirschner LS, et al. Mutations of the gene encoding the protein kinase A type I-á regulatory subunit in patients with the Carney complex. Nat Genet. 2000; 26:89–92. [PubMed: 10973256]

17. Kotani T. Protein kinase A activity and Hedgehog signaling pathway. Vitam Horm. 2012; 88:273–291. [PubMed: 22391308]

18. Lin CY, et al. Transcriptional amplification in tumor cells with elevated c-Myc. Cell. 2012; 151:56–67. [PubMed: 23021215]

19. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci USA. 2002; 99:6567–6572. [PubMed: 12011421]

20. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. [PubMed: 23582322]

21. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. [PubMed: 24119843]

22. Lovén J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013; 153:320–334. [PubMed: 23582323]

23. Rifat Y, et al. Regional neural tube closure defined by the Grainy head-like transcription factors. Dev Biol. 2010; 345:237–245. [PubMed: 20654612]

24. Wu XH, Wang Y, Zhuo Z, Jiang F, Wu YD. Identifying the hotspots on the top faces of WD40-repeat proteins from their primary sequences by â-bulges and DHSW tetrads. PLoS One. 2012; 7:e43005. [PubMed: 22916195]

25. Gymnopoulos M, Elsliger MA, Vogt PK. Rare cancer-specific mutations in *PIK3CA* show gain of function. Proc Natl Acad Sci USA. 2007; 104:5569–5574. [PubMed: 17376864]

26. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004; 5:557–572. [PubMed: 15475419]

27. Bilgüvar K, et al. Whole-exome sequencing identifies recessive *WDR62* mutations in severe brain malformations. Nature. 2010; 467:207–210. [PubMed: 20729831]

28. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011; 21:936–939. [PubMed: 20980556]

29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]

30. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]

31. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011; 27:2987–2993. [PubMed: 21903627]

32. Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

33. Cui Q. A network of cancer genes with co-occurring and anti-co-occurring mutations. PLoS One. 2010; 5:e13180. [PubMed: 20957180]

34. Sathirapongsasuti JF, et al. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. Bioinformatics. 2011; 27:2648–2654. [PubMed: 21828086]

35. Chen K, et al. Break Dancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 2009; 6:677–681. [PubMed: 19668202]

36. O'Roak BJ, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012; 338:1619–1622. [PubMed: 23160955]

37. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]

38. Trapnell C, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012; 7:562–578. [PubMed: 22383036]

39. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–192. [PubMed: 22517427]

40. Smyth, GK. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Gentleman, R.; Carey, VJ.; Huber, W.; Irizarry, RA.; Dudoit, S., editors. Springer; 2005. p. 397-420.

41. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012; 28:882–883. [PubMed: 22257669]

42. Lovén J, et al. Revisiting global gene expression analysis. Cell. 2012; 151:476–482. [PubMed: 23101621]

43. Verhaak RG, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. Cancer Cell. 2010; 17:98–110. [PubMed: 20129251]

44. De Sousa E Melo F, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat Med. 2013; 19:614–618. [PubMed: 23584090]

45. Sadanandam A, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med. 2013; 19:619–625. [PubMed: 23584089]

46. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26:1572–1573. [PubMed: 20427518]

47. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010; 11:367. [PubMed: 20598126]

48. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA. 2001; 98:5116–5121. [PubMed: 11309499]

49. Isella C, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2015; 47:312–319. [PubMed: 25706627]

50. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]

51. Cotney J, et al. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. Genome Res. 2012; 22:1069–1080. [PubMed: 22421546]

52. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

53. Zhang Y, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]

54. Chipumuro E, et al. CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. Cell. 2014; 159:1126–1139. [PubMed: 25416950]

55. Ross-Innes CS, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature. 2012; 481:389–393. [PubMed: 22217937]

56. Bailey TL, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37:W202–W208. [PubMed: 19458158]
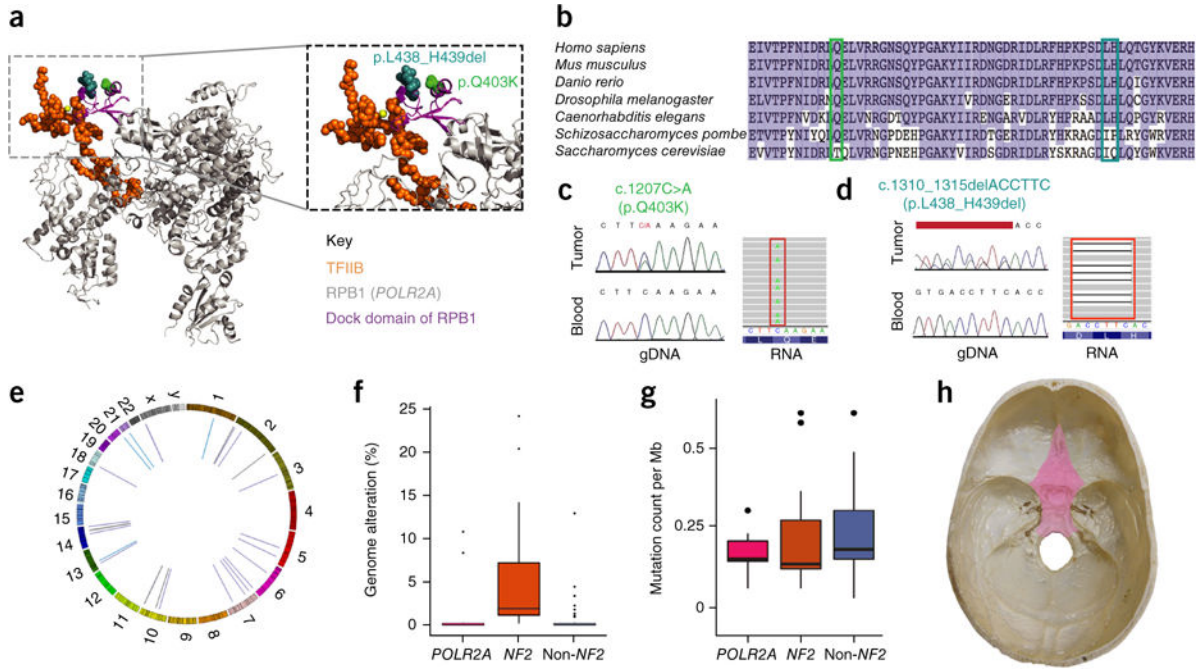
**Figure 1. *POLR2A* mutations define a distinct subset of benign meningiomas**
**(a)** The recurrent somatic *POLR2A* mutations localize to the highly conserved dock domain, where interactions with TFIIB mediate the formation of the pre-initiation complex. Alterations found in human meningioma were mapped to the *Saccharomyces cerevisiae* RPB1 structure (PDB ID 1R5U). Green, RPB1 p.Gln403Lys equivalent residue; teal, RPB1 p.Leu438_His439del equivalent residues; yellow, zinc ion. **(b)** Alignment of the protein sequences of the large Pol II subunit for seven species shows the dock domain, including the two residues altered by recurrent meningioma mutations, to be highly conserved. **(c.d)** Representative Sanger chromatograms of recurrent *POLR2A* mutations in tumor and matching blood show that the mutations are somatic. Mutant alleles are confirmed by RNA-seq reads. gDNA, genomic DNA. **(e)** A representative Circos plot of a whole-genome-sequenced *POLR2A* meningioma does not show any large-scale chromosomal rearrangements. **(f)** The percentage of the genome altered by copy-number variations between meningioma subgroups. The low proportion of the genome altered by copy-number variation was similar in *POLR2A* mutant tumors and other non-*NF2* mutant meningiomas. **(g)** The number of somatic mutations (normalized per megabase pair of sequencing data) from whole-exome sequencing data that are protein-altering and predicted to be damaging. In **f** and **g**, lines indicate median values, box edges indicate 25th (bottom) and 75th (top) percentiles, and whiskers and points show the outliers. **(h)** *POLR2A* meningiomas are enriched along the skull base, near the tuberculum sellae region (highlighted in pink).
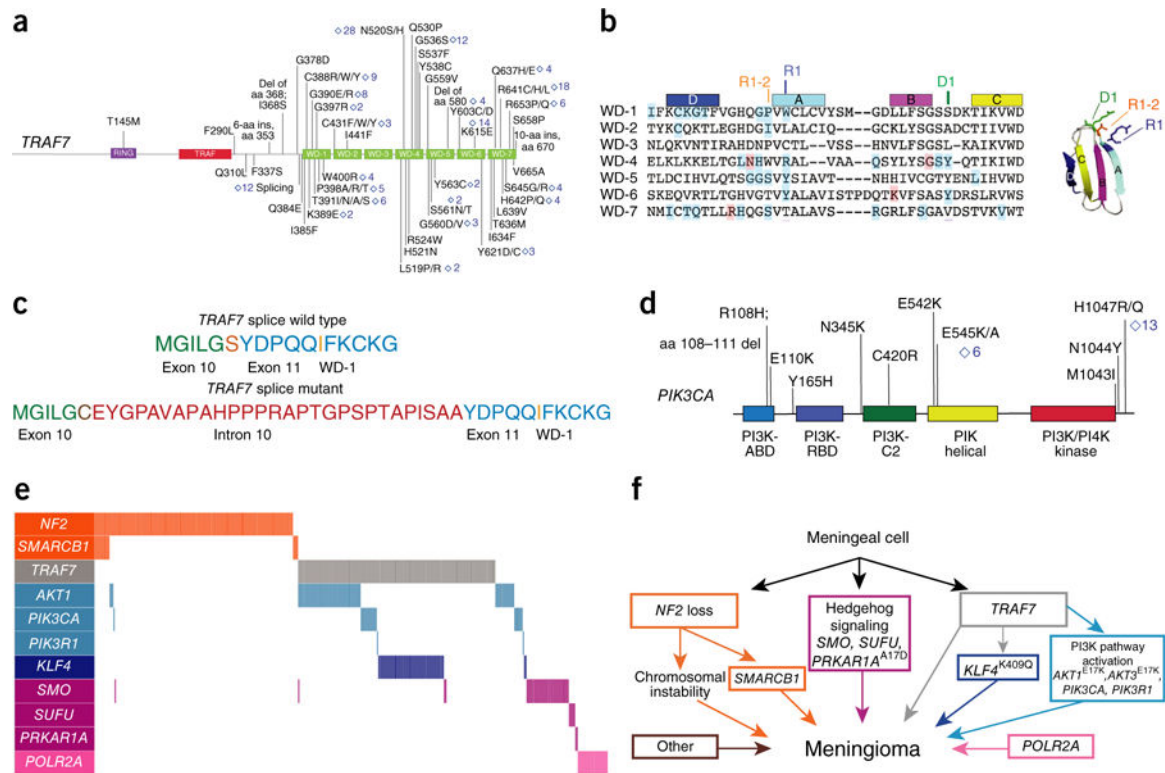
**Figure 2. Meningioma driver genes in five major pathways account for the formation of >80% of benign meningiomas**

**(a)** The distribution of the 182 *TRAF7* mutations showed highly recurrent missense mutations that clustered in the WD40 repeat domains or mutations affecting the splice junction immediately upstream of the WD40 repeat domain ($n = 12$), but not frameshift or nonsense mutations. Del, deletion; ins, insertion; aa, ammo acid. **(b)** Alignment of TRAF7 WD40 repeats. Left: the previously reported R1–2, R1, and D1 WD40 mutation hot spots[24]. Boxes A, B. C, and D identify recurrent structural motifs shared by WD40 repeat domains. Blue-shaded amino acids mark mutations detected in meningioma, whereas red-shaded amino acids represent highly recurrent meningioma mutations (seen in more than ten tumors). Right: typical WD40 repeat (here, FBW7 WD40 repeat 3, PDB ID 2QVQ) with color-coded structural motifs and hot-spot mutations. **(c)** Recurrent *TRAF7* splice mutations promote intron inclusion, resulting in a 28-amino-acid in-frame insertion plus a Ser-to-Cys missense alteration. **(d)** Distribution of *PIK3CA* mutants showing hot spots at p.Glu545 and p.His1047, which have been previously reported to result in constitutive PI3K signaling[8, 25]. **(e)** Distribution of the driver genes detected in the benign (WHO grade I) meningioma cohort. Each bar represents a meningioma sample. Only genes screened via targeted next-generation sequencing and/or Sanger sequencing are plotted. Our analysis identified somatic mutations in *POLR2A, PIK3R1, SUFU*, and *PRKAR1A* that were mutually exclusive with the previously identified meningioma genes. **(f)** Benign meningiomas form mutually exclusive subgroups based on mutational background.
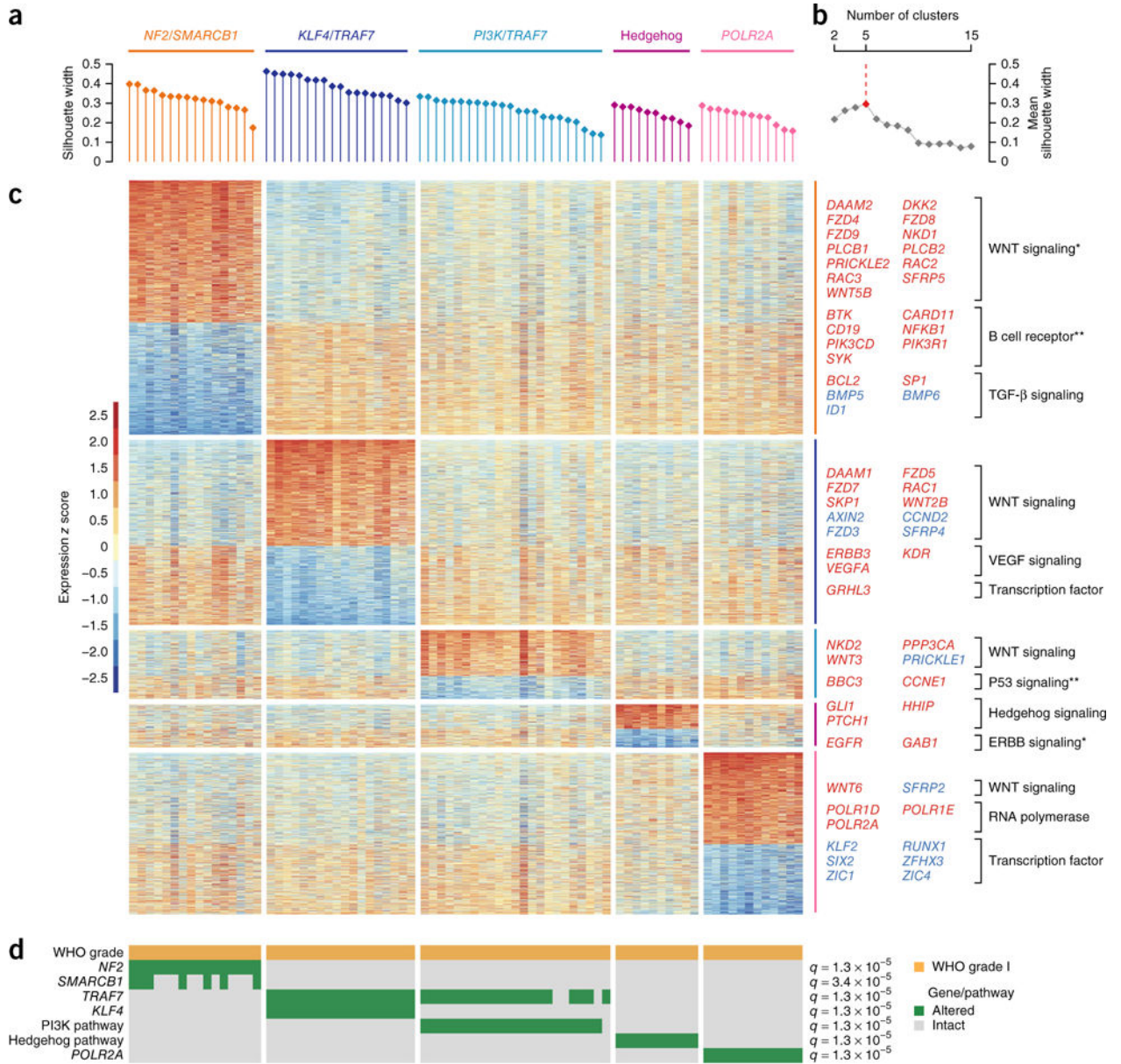
**Figure 3. Meningioma subgroups cluster according to their genome-wide transcription profile and show significant differences in the expression of critical developmental regulators**
(a) Unsupervised hierarchical clustering of gene expression in meningioma tumor samples. The width of each clustering silhouette is a measure of the tumor's match within the indicated subgroup. Each point represents a single tumor. (b) The number of clusters used for analysis was chosen to maximize the mean width of the clustering silhouette. Clustering with five subgroups gave optimal results. (c) Subgroup-specific genes were selected for heat-map generation in a supervised approach via a classification algorithm. For each subgroup, representative genes and the pathways they belong to are presented on the right. Upregulated genes are shown in red, and downregulated ones in blue. Pathway enrichment was performed using a hypergeometric test: $*P < 0.05$, $**q < 0.1$. (d) Samples clustered according to underlying driver mutation, $q$-values were calculated via two-sided Fisher's exact test.
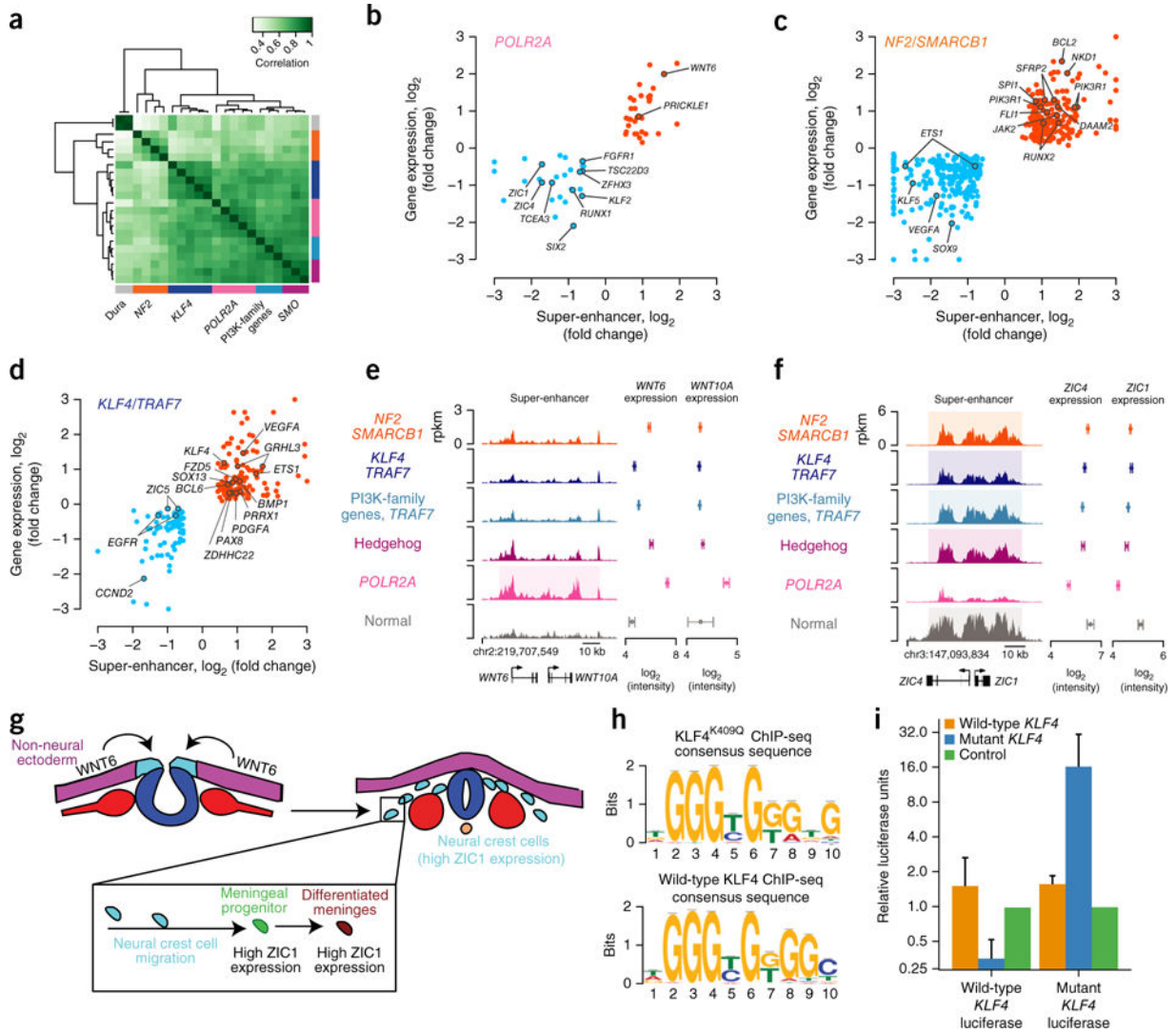
**Figure 4. Super-enhancer-driven genes classify meningioma subgroups**
**(a)** Unsupervised hierarchical clustering based on super-enhancer binding classifies meningioma subgroups. Super-enhancer binding scores were derived from affinity based on H3K27ac ChIP-seq read counts. **(b–d)** Concordant differential gene expression and differentially bound super-enhancers between meningioma subgroups. Red points represent a differentially high-binding super-enhancer (FDR < 0.05) with concordant upregulation of its nearby target gene (FDR < 0.05). Blue points represent a differentially low-binding super-enhancer that correlates with downregulation of its target gene. **(e,f)** Differential super-enhancer binding and gene expression was detected at the *WNT6/WNT10A* **(e)** and *ZIC1/ZIC4* **(f)** loci in *POLR2A* mutant tumors compared with other meningioma subgroups and control dura ('normal'). Background shading represents super-enhancers present in two or more samples per subtype. Expression is $\log_2$ transformed, rpkm, reads per kilobase of transcript per million mapped reads. **(g)** Schematic outlining the roles of WNT6 and ZIC1 in controlling neural crest development and meningeal cell differentiation. During embryonic development, WNT6 is expressed and secreted by the non-neural ectoderm (purple) to

trigger induction of neural crest cells (light blue), whereas ZIC1 is expressed by neural crest cells and is a marker for meningeal cells. **(h)** Compared to the wild-type KLF4 consensus binding sequence, the two 3′-most nucleotides of the KLF4 p.Lys409Gln mutant consensus DNA binding sequence were altered to T and G, respectively. **(i)** Using luciferase assays, we confirmed that KLF4 p.Lys409Gln binds to this novel mutant motif *in vitro*, activating higher downstream gene expression than either the control or wild-type KLF4. Error bars represent s.d., calculated over three replicates.