

# Expression alterations define unique molecular characteristics of spinal ependymomas

Anbarasu Lourdasamy<sup>1</sup>, Ruman Rahman<sup>1</sup> and Richard G. Grundy<sup>1</sup>

<sup>1</sup>Children's Brain Tumour Research Centre, School of Medicine, Queen's Medical Centre University of Nottingham, Nottingham, UK

**Correspondence to:** Anbarasu Lourdasamy, **email:** anbarasu.lourdasamy@nottingham.ac.uk

**Keywords:** ependymoma; gene expression; meta-analysis; co-expression network

**Received:** February 20, 2015

**Accepted:** March 11, 2015

**Published:** March 30, 2015

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Ependymomas are glial tumors that originate in either intracranial or spinal regions. Although tumors from different regions are histologically similar, they are biologically distinct. We therefore sought to identify molecular characteristics of spinal ependymomas (SEPN) in order to better understand the disease biology of these tumors. Using gene expression profiles of 256 tumor samples, we identified increased expression of 1,866 genes in SEPN when compared to intracranial ependymomas. These genes are mainly related to anterior/posterior pattern specification, response to oxidative stress, glial cell differentiation, DNA repair, and PPAR signalling, and also significantly enriched with cellular senescence genes ( $P = 5.5 \times 10^{-03}$ ). In addition, a high number of significantly down-regulated genes in SEPN are localized to chromosome 22 (81 genes from chr22: 43,325,255 – 135,720,974;  $FDR = 1.77 \times 10^{-23}$  and 22 genes from chr22: 324,739 – 32,822,302;  $FDR = 2.07 \times 10^{-09}$ ) including *BRD1*, *EP300*, *HDAC10*, *HIRA*, *HIC2*, *MKL1*, and *NF2*. Evaluation of *NF2* co-expressed genes further confirms the enrichment of chromosome 22 regions. Finally, systematic integration of chromosome 22 genes with interactome and *NF2* co-expression data identifies key candidate genes. Our results reveal unique molecular characteristics of SEPN such as altered expression of cellular senescence and chromosome 22 genes.**

## INTRODUCTION

Ependymomas are primary tumors of the central nervous system (CNS) representing 3%–6% of all CNS tumor [1]. Histologically, they have been classified into three grades according to the World Health Organization (WHO): grade I (subependymomas and myxopapillary ependymomas), grade II (classic ependymomas), and grade III (anaplastic ependymomas) [2]. These tumors originate from either intracranial or spinal regions of the CNS. Spinal ependymoma (SEPN) constitutes approximately 34.5% of ependymomas and accounts for 60% spinal cord gliomas, making them the most common glial tumors of the adult spine [3-7]. While 90% of all childhood ependymomas are intracranial, SEPN are more commonly found in adults of 20 to 40 years of age [7] where standard treatment of SEPN is aiming for complete resection. Although, SEPN generally have better prognosis

than intracranial tumors, recurrence rate can be as high as 50%–70% without adjuvant therapy; however the beneficial role of adjuvant chemotherapy or radiotherapy in SEPN is controversial [4, 7]. Most currently known prognostic factors for SEPN are based on clinical and histological criteria, such as extent of tumor resection, and histological grade. The results of existing studies on these prognostic markers are contradictory. Therefore, there is a need to improve the understanding of the biology of SEPN in order to develop more accurate prognostic signatures and identify new therapeutic targets.

Few studies have examined the genetics of SEPN compared to intracranial ependymomas. Ebert *et al* found loss of heterozygosity (LOH) on chromosome 22q in grade II (6/14 cases) and grade III (1/3 case) SEPN, *NF2* mutations in grade II SEPN (6/14 cases), and found no mutations in intracranial and myxopapillary ependymomas [8]. In a study with 52 tumors from

**Table 1: Enrichment of KEGG pathways among significantly up-regulated genes in spinal ependymomas.**

Pathway	N	FDR
Protein processing in endoplasmic reticulum	28	$5.44 \times 10^{-06}$
Tight junction	21	$4.06 \times 10^{-04}$
Nucleotide excision repair	10	$1.98 \times 10^{-03}$
Cell adhesion molecules (CAMs)	19	$2.01 \times 10^{-03}$
Pyrimidine metabolism	15	$3.25 \times 10^{-03}$
Pyruvate metabolism	9	$4.17 \times 10^{-03}$
RNA transport	19	$6.02 \times 10^{-03}$
Purine metabolism	20	$6.11 \times 10^{-03}$
PPAR signaling pathway	11	$1.69 \times 10^{-02}$
Proteasome	8	$1.70 \times 10^{-02}$
Adherens junction	11	$1.79 \times 10^{-02}$
Oxidative phosphorylation	16	$2.04 \times 10^{-02}$
Focal adhesion	21	$2.54 \times 10^{-02}$

Enrichment P values for KEGG pathways were calculated using the Hyper-geometric test and corrected for multiple testing using Benjamini & Hoachberg method. The pathways with adjusted P values < 0.05 were selected. N: number of genes in the pathway; FDR: false discovery rate.

45 patients, Lamsuzs *et al* detected LOH on chr22q more frequently in intramedullary SEPN (14/20 cases) compared to intracranial ependymomas (6/25 cases) and found NF2 mutations in 5 out of 20 cases of SEPN [9]. Singh *et al* found the NF2 gene deletion in 5/15 SEPN and loss of the NF2 gene product, merlin in 5 out of 27 cases, all of which were from spinal [10].

Microarray-based expression studies have also been used to compare SEPN with intracranial tumors and correlate molecular signatures with clinical and histologic characteristics. A study by Korshunov *et al* examined 39 ependymal neoplasms including ten SEPN and detected 14 genes that were more highly expressed in SEPN compared to intracranial ependymomas including *HOXB5*, *PLA2G5* and *ITIH2* [11]. Lukashova-v Zangen *et al* reported five genes (*TFAM*, *EDN1*, *GAS2L1*, *HUMRTLH3* and *ADAM9*) that were preferentially expressed in SEPN by comparing grade II adult SEPN (n = 8) and adult intracranial ependymomas (n = 4) from a cohort of 47 ependymoma patients [12]. By comparing tumor samples from 16 SEPN and 16 intracranial ependymomas, Palm *et al* reported the over-expression of HOX genes in SEPN and up-regulation of several genes of Notch, Hedgehog, and BMP signaling pathways in intracranial ependymoma [13]. Finally, Taylor *et al* compared gene expression microarrays of tumor samples from SEPN (n = 3), supratentorial (n = 5), and posterior fossa (n = 21), and identified the expression signature for SEPN that consisted of 184 genes including HOX genes *HOXA7*, *A9*, *HOXB6*, *B7*, *HOXC8*, *C10*, and *IGF1* [14].

The aforementioned gene expression studies are limited to a small number of samples and are usually analysed in isolation. Meta-analysis approaches make it possible to combine multiple independent gene expression datasets and increase the statistical power for gene discovery. Such meta-analysis approaches have been successfully used to identify transcriptional signatures in cancer [15] and aging [16]. Individual gene expression studies are limited by systematic biases at both biological and technical levels, which hinder the broader application of their findings. Meta-analysis can control for such confounding factors by increasing the statistical power to detect consistent changes across multiple datasets. No systematic multi-cohort analysis has yet evaluated transcriptional alterations between SEPN and intracranial ependymomas. The present study uses microarray datasets from three independent cohorts to compare the biology of SEPN and intracranial ependymomas to identify unique molecular characteristics of SEPN such as altered expression of cellular senescence and chromosome 22 genes.

## RESULTS

We applied two different meta-analysis approaches to the normalized expression data from three independent studies to find differentially expressed genes. A total of 3,182 genes were identified as significantly differentially expressed (FDR < 0.05) between SEPN and intracranial ependymomas by both methods (Supplementary Table 1).

**Table 2: Top 25 genes that are highly correlated with NF2 gene expression.**

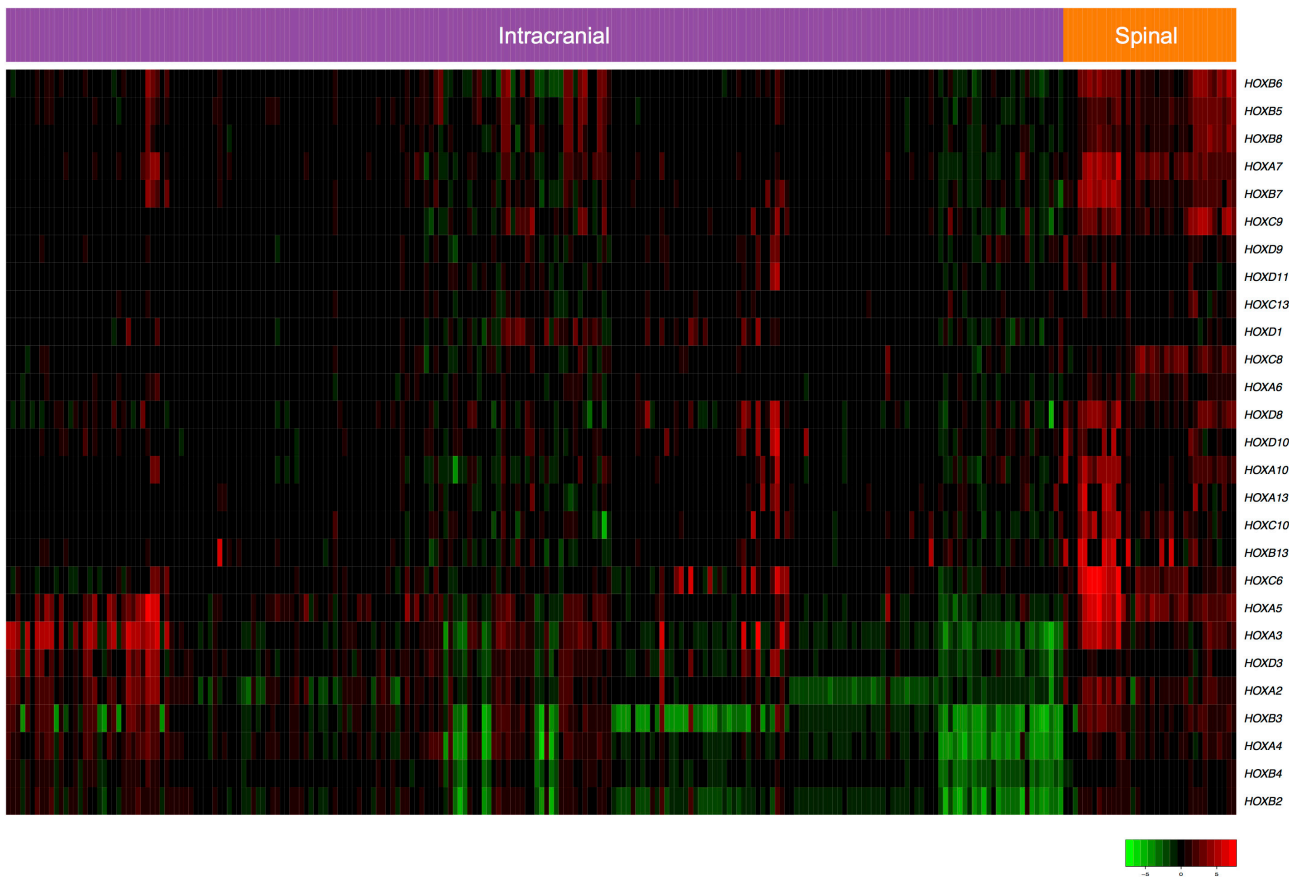
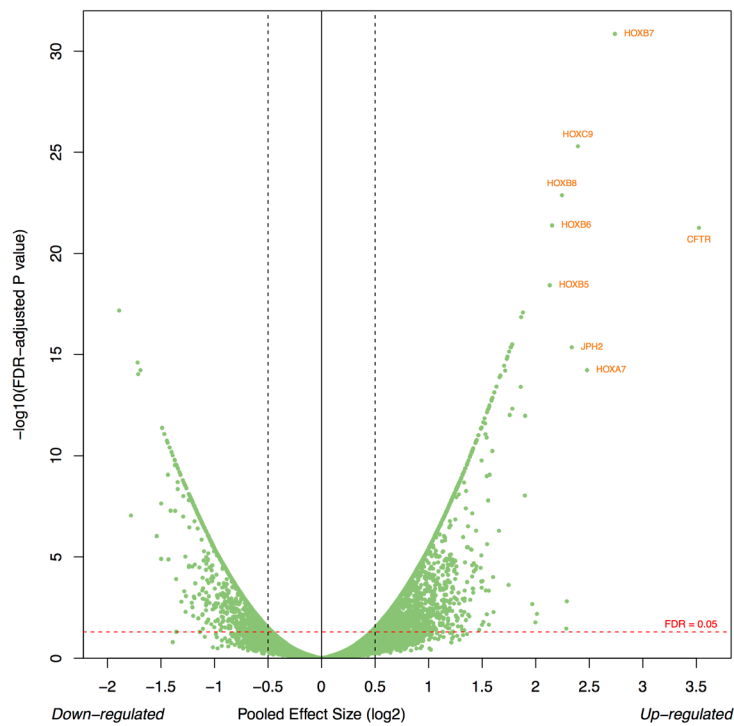
Gene	Location	ES	SE	FDR
<i>MIEF1</i>	22q13	0.91	0.09	$8.34 \times 10^{-19}$
<i>NDUFA6</i>	22q13.2	0.73	0.12	$4.54 \times 10^{-08}$
<i>ASCC2</i>	22q12.1	0.73	0.20	$1.78 \times 10^{-03}$
<i>ADSL</i>	22q13.2	0.69	0.17	$2.66 \times 10^{-04}$
<i>PITPNB</i>	22q12.1	0.68	0.08	$5.06 \times 10^{-14}$
<i>RPS19BP1</i>	22q13.1	0.68	0.09	$7.94 \times 10^{-12}$
<i>UFD1L</i>	22q11.21	0.64	0.10	$9.85 \times 10^{-09}$
<i>MTMR3</i>	22q12.2	0.63	0.15	$1.69 \times 10^{-04}$
<i>TCF20</i>	22q13.3	0.61	0.06	$1.58 \times 10^{-18}$
<i>DRG1</i>	22q12.2	0.61	0.13	$5.14 \times 10^{-05}$
<i>DGCR2</i>	22q11.21	0.61	0.08	$1.20 \times 10^{-11}$
<i>EIF3D</i>	22q13.1	0.61	0.16	$1.20 \times 10^{-03}$
<i>SAMM50</i>	22q13.31	0.61	0.09	$1.24 \times 10^{-09}$
<i>SRRD</i>	22q12.1	0.61	0.19	$5.65 \times 10^{-03}$
<i>ITPK1</i>	14q31	0.60	0.06	$6.35 \times 10^{-18}$
<i>EIF4ENIF1</i>	22q11.2	0.60	0.18	$3.37 \times 10^{-03}$
<i>MTHFD1</i>	14q24	0.60	0.08	$2.20 \times 10^{-10}$
<i>TTC38</i>	22q13	0.60	0.06	$1.46 \times 10^{-17}$
<i>UBE2L3</i>	22q11.21	0.59	0.16	$1.13 \times 10^{-03}$
<i>THOC5</i>	22q12.2	0.58	0.17	$1.96 \times 10^{-03}$
<i>TLL12</i>	22q13.31	0.58	0.06	$6.14 \times 10^{-17}$
<i>EIF3L</i>	22q	0.58	0.08	$3.51 \times 10^{-10}$
<i>ATXN10</i>	22q13.31	0.58	0.24	$3.19 \times 10^{-02}$
<i>FCSD2</i>	11q13.4	0.57	0.09	$4.51 \times 10^{-09}$
<i>NOL12</i>	22q13.1	0.56	0.06	$9.44 \times 10^{-16}$

Pearson correlation coefficients (r) between each gene expression and NF2 expression were calculated in each data set and combined using the DerSimonian-Laird random-effect meta-analytical approach by calculating Z-mean of r (ES) and its standard error (SE). Location: chromosome location; ES: Effect size; SE: standard error; FDR: false discovery rate.

### Expression analysis highlights diverse processes and pathways among over-expressed genes in SEPN

Of the 3,182 differentially expressed genes between SEPN and intracranial ependymomas, 59% (1866) were consistently up-regulated in SEPN (Figure 1A). The most significantly up-regulated genes included *HOXB7* (pooled effect size on log2, ES = 2.74, FDR =  $1.41 \times 10^{-31}$ ), *CTFR* (ES = 3.52, FDR =  $5.33 \times 10^{-22}$ ), *HOXB5* (ES = 2.13, FDR =  $3.72 \times 10^{-19}$ ), and *CTTNBP2* (ES = 1.77, FDR =  $4.29 \times 10^{-16}$ ). Other interesting genes that have increased expression in SEPN included *EZH1*, *IDH3*, *NEFL*, and *NELL2*. Of the 27 HOX genes annotated in the microarrays, 22 were significantly up-regulated in the SEPN (Figure 1B). Gene ontology (GO) analysis of over-expressed genes showed significant overrepresentation of genes involved in anterior/posterior pattern specification (27 genes; FDR =  $9.37 \times 10^{-10}$ ), apoptotic process (72

genes; FDR =  $6.70 \times 10^{-08}$ ), cell cycle (52 genes; FDR =  $1.68 \times 10^{-05}$ ), cilium assembly (13 genes; FDR =  $5.64 \times 10^{-05}$ ), and cell proliferation (35 genes; FDR =  $2.62 \times 10^{-03}$ ), as well as genes involved protein folding, DNA replication, mitochondrial electron transport, NADH to ubiquinone, response to oxidative stress, and cell redox homeostasis (Supplementary Table 2). In addition, genes involved in DNA repair processes such as nucleotide-excision repair and double-strand break; genes in signaling pathways such as positive regulation of I-kappaB kinase/NF-kappaB cascade, positive regulation of MAPK cascade; and glial cell differentiation were significantly enriched in SEPN (Supplementary Table 2). Our pathway analysis identified Protein processing in endoplasmic reticulum as the top canonical pathway followed by pathways involved in DNA repair, cell adhesion, metabolism, and PPAR signaling (Table 1).



**Figure 1: Gene expression profiles define unique molecular characteristics of SEPN.** **A.** Volcano plot showing the number of significantly differentially expressed genes in SEPN (FDR < 0.05). The x-axis represents the pooled effect sizes that are estimated from three independent microarray datasets by the meta-analysis of random-effects model and the y-axis represents  $-\log_{10}$  of false discovery rate (FDR). **B.** Heatmap depicting expression profiles of HOX gene that are significantly enriched in SEPN. The normalized value of each gene is indicated by colour intensity, with red/green representing high/low expression.

## Cellular senescence genes are over-expressed in SEPN

Cancer incidence and mortality increases with age and thus age is considered as a prime risk factor for several types of cancers, including gliomas [17, 18]. In addition, emerging evidences indicate that aging and cancer are closely related and mechanisms underlying the cellular senescence (CS) program may link these two processes [19]. To determine whether similar processes are shared between SEPN and CS, we compared the genes associated with CS with up-regulated genes in SEPN. We obtained a set of 261 candidate genes that are involved in CS from the published study [20]. The CS list is a manually curated set of candidate genes implicated by genetic or RNA interference (RNAi) interventions (gene knockout, partial or full loss-of-function mutations, RNAi-induced gene silencing, overexpression), which reportedly cause cells to induce, inhibit or reverse CS, and genes shown to be markers of CS. Among the up-regulated genes in SEPN there was a significant enrichment for the CS-associated genes (hypergeometric test, 34 genes;  $P = 5.5 \times 10^{-03}$ ), which includes oncogenes (*BCL2*, *CDK6*, *MDM2*, and *NR4A3*), cytokines and growth factors (*AGT*, *CYR61*, *FGF1*, and *IGF1*), transcription factors (*MECP2*, *PCGF2*, *PER2*, and *TP63*), member of RAS superfamily of small GTP-binding proteins (*RAC1*), and genes involved in the oxidative stress pathway (*SOD1*) (Figure 2 and Supplementary Table 3).

## Chromosome 22 genes are under-expressed in SEPN

One of the most frequent genomic alterations detected in sporadic ependymomas was the loss of chromosome 22, with the frequency ranging from 26% to 71% [8, 9]. In addition, loss of heterozygosity (LOH) on chromosome 22q has been found more frequently in intramedullary SEPN and more often in adults than the pediatric ones [21, 22]. To investigate whether chromosome loss alter expression of endogenous genes, we identified differentially expressed genes that are significantly down regulated in SEPN (1316 genes at  $FDR < 0.05$ , Supplementary Table 1) and tested for the enrichment of chromosomal regions. The positional gene enrichment analysis revealed that many down-regulated genes in SEPN localized to specific chromosomal regions rather than genes randomly distributed in the genome. Most prominently, a high number of genes were localized to a 92.4 Mbp region (81 genes from chr22: 43325255 – 135720974;  $FDR = 1.77 \times 10^{-23}$ ) and a 32.5 Mbp region (22 genes from chr22: 324739 – 32822302;  $FDR = 2.07 \times 10^{-09}$ ) of the chromosome 22 (Supplementary Table

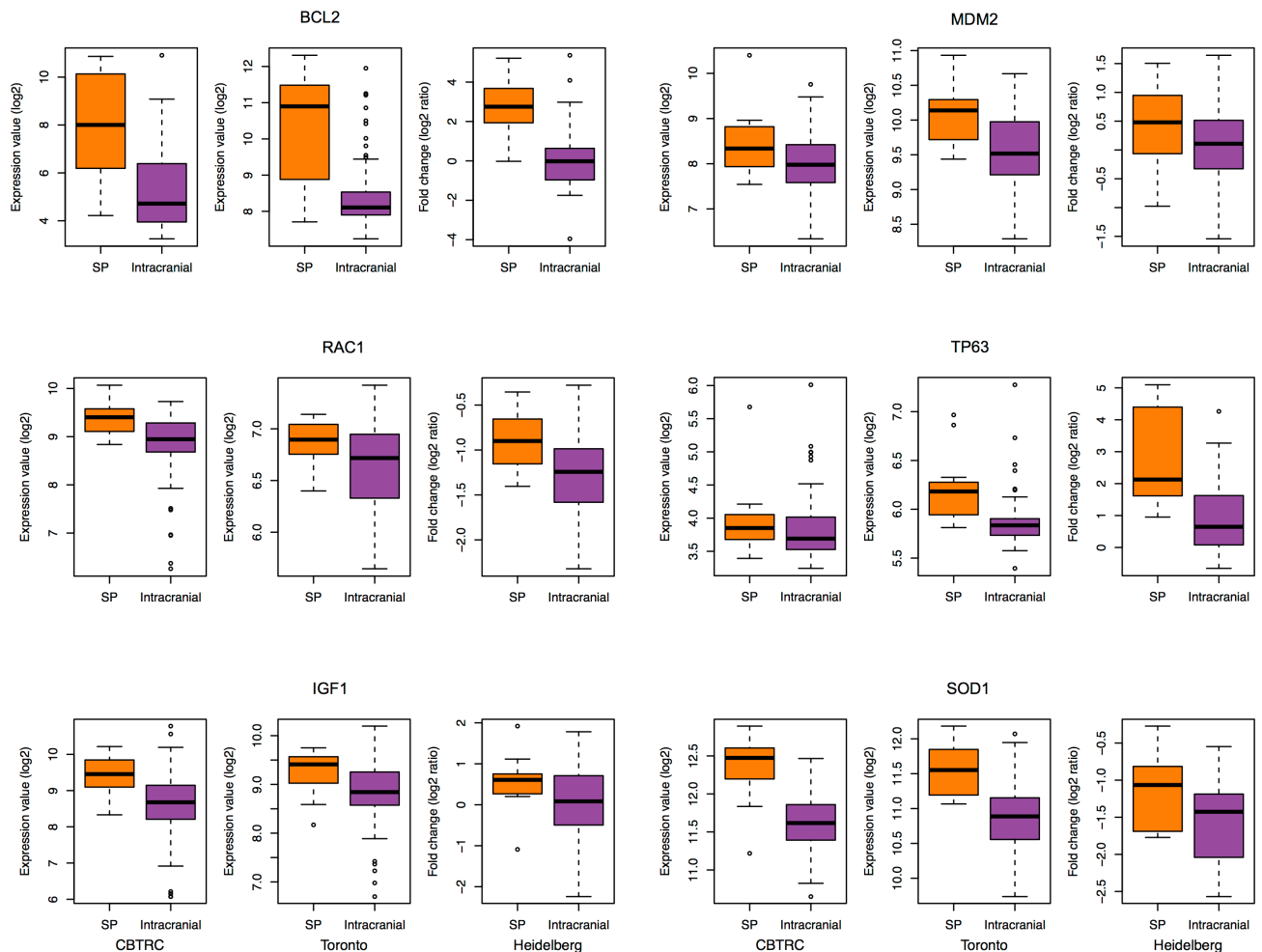
4). To confirm these observations, we also performed enrichment analysis of cytogenetic bands with the GSEA method. We observed that 51 genes located in chr22q13 (hypergeometric test,  $FDR = 1.41 \times 10^{-27}$ ), 27 genes in chr22q12 ( $FDR = 5.88 \times 10^{-13}$ ), and 40 genes in chr22q11 ( $FDR = 4.54 \times 10^{-11}$ ) were highly enriched among genes that were significantly down-regulated in SEPN (Supplementary Table 4). Furthermore, genes located in cytogenetic bands chr19p13 (59 genes;  $FDR = 5.88 \times 10^{-13}$ ), chr19q13 (59 genes;  $FDR = 1.34 \times 10^{-09}$ ), and chr20q13 (25 genes;  $FDR = 2.01 \times 10^{-06}$ ) were also enriched among down-regulated genes (Supplementary Table 4). Strikingly, 84% of genes located on chromosome 22 (271 out of 321 chr22 genes that were detected in the meta-analysis) were under expressed and 125 of them were significantly down-regulated in spinal when compared to intracranial ependymomas (Figure 3A). These included several cancer-associated genes such as *MKLI*, *EP300*, *NF2*, and *HIC2* and chromatin binding genes *BRD1*, *HIRA*, and *HDAC10* (Figure 3B).

## NF2-associated genes in ependymoma are located on chromosome 22

Our analysis of differential expression in relation to chromosome 22 identified a tumor suppressor gene, neurofibromin 2 (*NF2*), which is located at chromosome 22q12.2 and showed decreased expression in SEPN ( $ES = -1.05$ ,  $FDR = 8.75 \times 10^{-04}$ ). Higher incidence of spinal ependymomas in patients with neurofibromatosis type 2 and frequent of loss of the *NF2* gene in SEPN have been reported in numerous studies [23, 24]. This led us to further investigate the effect of *NF2* on the expression of other SEPN genes. We used three independent microarray datasets to identify potential co-expressed genes of *NF2* in an un-biased manner. The meta-analysis of correlations between *NF2* gene expression and other genes resulted in 260 genes ( $Z$  mean of correlations,  $r > 0.4$  and  $FDR < 0.05$ ) (Supplementary Table 5). The most highly correlated gene with *NF2*-expression, the gene encoding the mitochondrial membrane protein *MIEF1*, was significantly down-regulated in SEPN ( $ES = -1.17$ ,  $FDR = 4.68 \times 10^{-08}$ ) (Table 2). Strikingly, enrichment analysis for the cytogenetic band revealed the marked enrichments in the *NF2* co-expressed genes residing at 22q13, 22q12, 22q11, 22q, 14q23, 14q24, and 1q25 (Supplementary Table 4). From a total of 260 *NF2* co-expressed genes, 148 (57%) were significantly down-regulated in SEPN (Supplementary Table 4). Out of fifteen genes that are physically located close to *NF2* (flanking region of 500 Kb), fourteen were down-regulated in SEPN and nine of them were significantly correlated with *NF2* gene expression (Supplementary Table 5).

**Table 3: Prioritization of Chromosome 22 genes by network and NF2 gene co-expression analyses.**

Symbol	Gene name	Location	DAPPLE P value	Correlation with NF2 gene expression		
				ES	SE	FDR
<i>EP300</i>	E1A binding protein p300	22q13.2	0.002	0.45	0.19	0.039
<i>HIRA</i>	Histone cell cycle regulator	22q11.21	0.015	0.46	0.14	0.004
<i>MNI</i>	Meningioma (disrupted in balanced translocation) 1	22q12.1	0.017	0.10	0.08	0.223
<i>SGSM3</i>	Small G protein signaling modulator 3	22q13.1-q13.2	0.021	0.54	0.06	$1.04 \times 10^{-14}$
<i>SUSD2</i>	Sushi domain containing 2	22q11-q12	0.023	0.21	0.06	0.004
<i>SREBF2</i>	Sterol regulatory element binding transcription factor 2	22q13	0.025	0.52	0.06	$1.20 \times 10^{-13}$
<i>RASD2</i>	RASD family, member 2	22q13.1	0.033	0.26	0.07	$5.56 \times 10^{-04}$
<i>LZTR1</i>	Leucine-zipper-like transcription regulator 1	22q11.21	0.044	0.42	0.07	$1.59 \times 10^{-07}$



**Figure 2: Increased expression of cellular senescence genes in SEPN.** Box plots showing log<sub>2</sub> expression levels (y-axis) of six cellular senescence genes (*BCL2*, *MDM2*, *RAC1*, *TP63*, *IGF1*, and *SOD1*) from spinal (SP) and intracranial ependymomas. Expression values from three different studies: CBTRC, Toronto, and Heidelberg are displayed for each gene.

## Prioritisation of chromosome 22 genes by integration of the interactome with NF2-associated genes

We next reasoned that altered expression of chromosome 22 genes could help to identify potential candidate genes of ependymoma in addition to a *NF2* gene. To systematically evaluate and prioritize genes located on chromosome 22 that are associated with SEP, we used the integration of protein-protein interaction (PPI) data with gene expression data. We first constructed the PPI network using significantly down regulated genes in SEP that are located on chromosome 22 as seed genes excluding *NF2*. We found there were 11 of 125 chromosome 22 genes participating in the direct network (Supplementary Figure 1). The changes in connectivity in the inferred network were calculated by comparing random networks of equal size and seed genes were ranked by P-value of increased connectivity. This analysis identified eight genes that were ranked by the network algorithm with a P value < 0.05 (Table 3). These eight candidate genes represent the most highly connected, and therefore, potentially most functionally important ones. In order to investigate the relevance of these genes with the known candidate gene, we combine the results of *NF2* gene co-expression analysis with the P-value ranking of the network analysis. All candidate genes except the *MNI* were significantly correlated with the expression profile of *NF2* (Table 3).

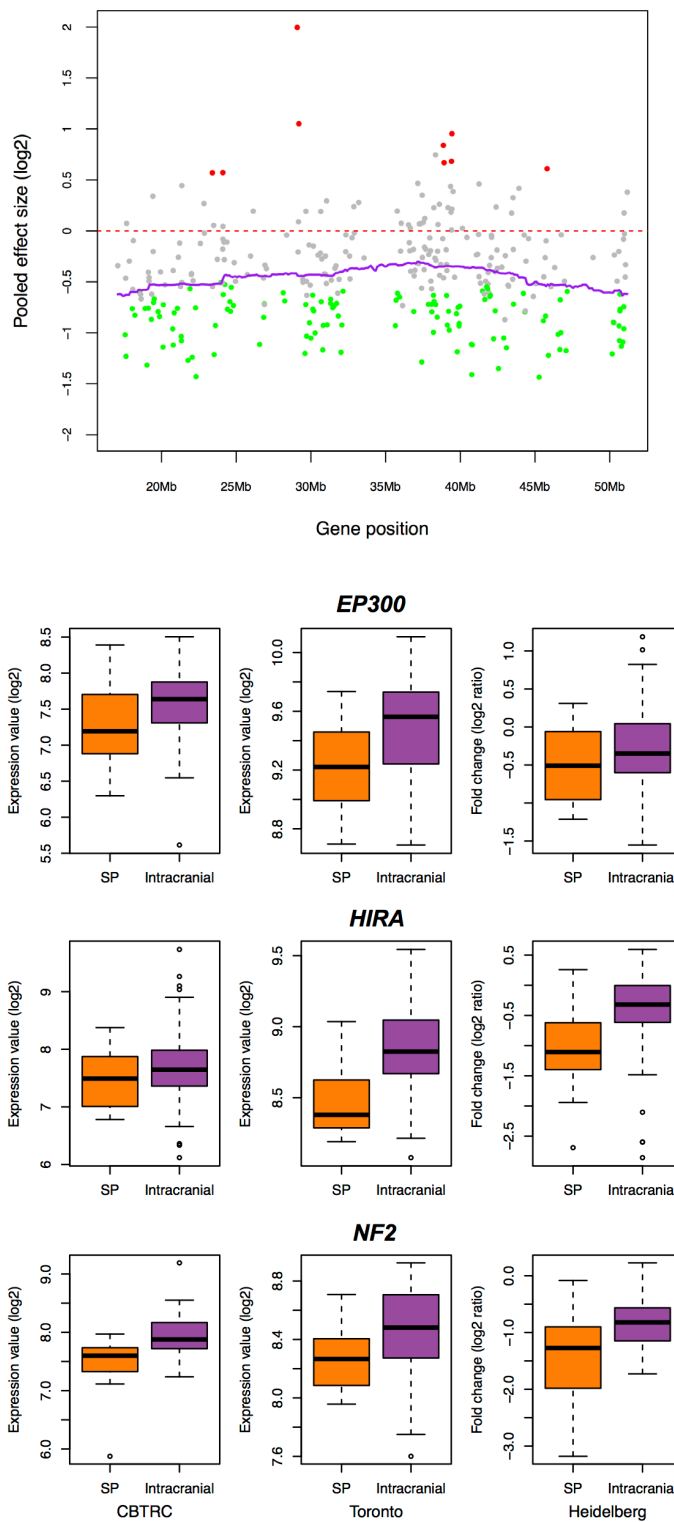
## DISCUSSION

Understanding distinct molecular characteristics exhibited by ependymomas according to their tumor location in the brain is becoming more important. Systematic analysis of molecular data from ependymomas have long been sought, however, there currently exists few studies that compared ependymomas arising in the spine to those intracranial. Here we demonstrate through meta-analysis approaches by combining multiple independent data sets that gene expression profiles of tumors from the spine display distinct patterns when compared with tumors from intracranial regions. To our knowledge, this is the first study that provides a comprehensive genome-wide gene expression profile and integrative analysis of SEP. Most, but not all, biological processes involved in the hallmarks of cancer are enriched with over-expressed genes in SEP. Genes related to cellular senescence are also highly enriched in SEP. In addition, we find that the majority of the genes from chromosomes 22 relatively decreased expression levels in tumors from the spine. Overall, we provide a systematic analysis of comprehensive gene expression data for assessing specific

biological processes of SEP.

Our results showed that SEP are characterized by diverse a set of known and novel biological processes and pathways. Previous work that used smaller study samples reported up-regulation of HOX genes in SEP when compared with ependymomas from intracranial regions [13, 14]. In the current study, we identified the up-regulation of multiple homeobox family members that include ANTP class homeoboxes (HOX) not previously implicated in SEP (*HOXA: A2, A3, A5, and A10; HOXB: B2, B3, B5, and B8; HOXC: C8 and C13; HOXD: D8, D9, and D10*, see Figure 1). The HOX genes, encode a family of evolutionarily conserved transcription factors that have fundamental roles in specifying anterior-posterior body patterning and development of the spine [25]. As they are involved in cellular fate determination and stem cell renewal, several studies investigated their role in other tumor types [26]. Particularly, HOX genes group 10 - 13 play important roles in the normal development of the lumbosacral region [25]. Indeed, the most significant GO biological process detected in this study was the anterior/posterior pattern specification. The combined analysis of large samples in our study from heterogeneous ependymoma patients recruited from different clinical settings provides confirmatory evidence of the association of HOX genes in SEP. Another important observation is the enrichment of genes involved in glial cell differentiation (*DNER, ERBB2, IGF1, METRN, and NFIB*) in SEP. This observation is consistent with the emerging evidence that radial-glia cells are likely cells of origin for ependymomas [14]. The up-regulation of these glial cell differentiation genes and homeobox family of transcription factors may transform radial-glia cells into cancer stem cells in the development of SEP. Our analysis also revealed novel biological processes such as antigen processing and presentation of antigen peptides via MHC class I, positive regulation of the I $\kappa$ B kinase/NF- $\kappa$ B cascade, positive regulation of the MAPK cascade, and proteolysis that have genes with relatively increased expression in SEP (Supplementary Table 1). These results provide data-driven hypotheses for future work, although we expect that further experimental evaluation will be necessary to understand the potential role of these biological processes in SEP.

Among the up-regulated genes in SEP were those that regulate DNA repair, DNA damage response, signal transduction by p53, response to oxidative stress, cell cycle, cell division, cell redox homeostasis, and mitochondrial electron transport (NADH to ubiquinone) (Supplementary Table 1). These genetic systems collectively play important roles in aging [27]. Several lines of evidence indicate that CS is a common process that links cancer and aging [19]. Senescent cells accumulate in ageing tissues, which may be due to an



**Figure 3: Down-regulated genes of SEPN contributed to the enrichment of chromosome 22.** **A.** Expression levels ( $\log_2$ ) of chromosome 22 genes that are detected in our study. Genes are sorted according to their position on the chromosome (x-axis) and coloured according to the statistical significance of their expression level (y-axis). Significantly down-regulated genes in SEPN ( $FDR < 0.05$ ) are marked with green, up-regulated ones with red, and non-significant ones with grey. Normalized  $\log_2$  expression values (dots) and kernel-smoothed expression values (purple line) are shown. **B.** Box plots showing  $\log_2$  expression levels (y-axis) of key genes that are located on chromosome (*EP300*, *HIRA*, and *NF2*) from spinal (SP) and intracranial ependymomas. Expression values from three different studies: CBTRC, Toronto, and Heidelberg are displayed for each gene.



increased senescence rate and/or decrease in the rate of clearance of senescent cells. The onset of CS in tumor cells can typically be activated by aberrant activation of oncogenes or loss-of-function of tumor suppressor genes, and also by several stressors, including DNA damage, oxidative stress, and signaling through either MAPK or IGF [19]. To determine a degree of CS convergence with SEPN, we compared the CS-associated genes with genes that are up-regulated in SEPN. Our analysis uncovered evidences of significant overlap between CS and SEPN at a molecular level, identifying core biological processes and genes they share. For example, oncogenes: *CDK6* and *MDM2*, tumor suppressor gene: *CHEK2*, oxidative stress genes: *SOD1* (encoding a member of the p53 family of transcription factors) and *TP63*, and IGF signaling genes: *IGF1*, *IGFBP5* and *IGFBP7*. Since the incidence of ependymomas from the spine increases with age, it is likely that the senescence pathway is involved in its development. Together, these findings add to the growing body of evidence that CS links cancer and ageing and that biological process in SEPN have a considerable degree of convergence with CS.

Loss of chromosome 22 and LOH have been frequently found in sporadic SEPN, and alteration of chromosome 22q is observed in 40% of SEPN [8, 13]. Analysis of gene expression profiles in the current study revealed the presence of many genes on chromosome 22 that were down-regulated in SEPN, indicating the functional consequences of chromosome 22 loss. The observed 92.4 and 32.5 Mb large domains of repressed transcription of chromosome 22 that include ‘hotspots’ regions 22q11, 22q12, and 22q13 is a result of a gene dosage effect due to unbalanced chromosomal alterations in SEPN. Potential tumor suppressor genes located within 22q12 include *NF2*, which exhibits markedly reduced mRNA expression in SEPN. The *NF2* gene encodes the FERM domain protein *Merlin*, which is co-ordinately regulated by intercellular adhesion and attachment to the extracellular matrix [28]. Increased incidence of CNS tumors including schwannoma, meningioma, and ependymoma has been reported in Neurofibromatosis type 2 patients who carry a single mutated *NF2* allele [23, 24]. The reduced expression of the tumor suppressor gene *NF2* in SEPN raises the question of the significance of expression alterations in other genes. From our meta-analysis of correlations with *NF2* gene expression, it is likely that similar expression alterations may occur in genes located in close physical proximity to *NF2*. Interestingly, genes located on 22q13, 22q12, 22q11 are significantly enriched among *NF2* co-expressed genes. Taken together these results further emphasize the loss of chromosome 22 in the transcriptional regulation of SEPN, and confirm the already reported importance of *NF2* transcriptional down-regulation.

The observed gene expression alterations over large regions of chromosome 22 suggest that these regions may

harbour potential candidate genes commonly affected in ependymomas. To prioritize 125 significantly down-regulated chromosome 22 genes in SEPN, we used the protein-protein interactions (PPI) network analysis with the assumption that proteins associated with disease tend to directly interact with each other. We thus identified 19 directly interacting proteins and investigated whether these PPI could help identify candidate genes within any of the 125 chromosome 22 genes identified in the current study. Our analysis identified eight genes, *EP300*, *HIRA*, *MNI*, *SGSM3*, *SUSD2*, *SREBF2*, *RASD2*, and *LZTR1*, and all of them except *MNI* are significantly correlated with *NF2* gene expression. These convergent lines of evidence strongly suggest that the prioritized genes may have pivotal role in ependymomas susceptibility. The tumor suppressor gene, *EP300* ranked highest among all eight genes, encodes histone acetyltransferase, *p300* that have widespread genomic effects on chromatin structure and gene expression as well as non-genomic effects on protein function [29]. It participates in the regulation of a wide range of biological processes such as proliferation, cell cycle regulation, apoptosis, differentiation, and DNA damage response [29, 30]. The *p300* protein functions primarily as a transcription cofactor for a number of nuclear proteins including known oncoproteins and tumor-suppressor proteins [29, 30]. An increasing body of evidence indicates that *p300* may be important in cancer [31, 32]. Interestingly, *p300* directly interacts with a transcription factor *NF-κB* as well as with the p65 protein that encoded by a gene *RELA*. The detection of frequent *C11orf95-RELA* gene fusion in supratentorial ependymomas further supports the potential role of *EP300* in ependymomas. Further functional investigation of *EP300* and other genes will provide pivotal information on the pathophysiology of ependymomas and potential therapeutic targets.

In summary, this study provides a comprehensive and global view of genes altered in SEPN. Importantly, enrichment of anterior/posterior pattern specification, response to oxidative stress, and cellular senescence genes distinguish the SEPN from intracranial ependymomas. We prioritized all chromosome 22 genes altered in ependymoma by comprehensive integration of distinct lines of evidence from different sources. The identification of these candidate genes provides an evidence-based rationale for functional studies that will help to further interrogate the initiation and/or progression of ependymomas.

## MATERIALS AND METHODS

### Microarray data

Three independent microarray datasets comprising a total of 262 expression profiles from tumors of ependymoma patients were used in this study. The supplementary table 6 reports the sample size, mean age, age range, gender, and the GSE identifier of the dataset in the Gene Expression Omnibus (GEO) database. Dataset 1 (CBTRC) contains tumor samples from intracranial (n = 71) and spinal (n = 12) regions of ependymoma patients with mean age of 10.6 and 25.6 years [33]. The CBTRC data were generated using the Affymetrix Human Genome U133 Plus 2.0 arrays. Dataset 2 (Toronto) consists of 85 intracranial tumors from patients with a mean age of 8.1 and 19 spinal tumors from patients with a mean of age of 35.6 years [34]. The Toronto data were generated using the Affymetrix human Exon 1.0 ST arrays. Dataset 3 (Heidelberg) consists of tumor samples from intracranial (n = 65) and spinal cord (n = 10) of patients with mean age of 18.7 and 35.3 years [34]. The Heidelberg data were generated using the Agilent microarrays.

### Microarray data analysis

#### Pre-processing and normalization

Expression intensity values were calculated at probeset level for the 83 Affymetrix U133 Plus 2.0 CEL files using the robust multi-array average (RMA) method [35]. Probesets that are 'absent' (present / absent call using MAS5) in all samples were filtered out from the analysis. Expression values were mapped from probeset to unique gene and the probeset with the highest mean expression value was selected when multiple probesets were mapped to the same gene. The final filtering step left a total of 18,166 genes. For the Affymetrix Exon 1.0 ST arrays, we used the Affymetrix Power Tools (APT) to generate gene-level (core meta-probeset) expression values from raw CEL files. Arrays were normalized using RMA, which included RMA background correction, quantile normalization, log transformation, and probeset summarization. Detection above background (DABG) was performed at both the probe and the probeset level using GC-matched background probes, and low variance probesets were excluded (17,001 genes). For the Heidelberg data set, the pre-processed data was directly obtained from the GEO database. Probes with more than 30% missing values were filtered out and the probeset with the highest average expression value was retained when removing duplicate probesets for a gene (18,913 genes).

### Differential expression analysis

Two different approaches were applied to the normalized data to identify differentially expressed genes between intracranial and spinal ependymomas. The first approach uses a meta-analysis that combines effect sizes from each dataset into a pooled effect size to estimate the amount of change in expression across all datasets. In each data set, the effect size was calculated for each gene using Hedges' adjusted *g*. A random effects model was used to combine effect sizes to obtain the pooled effect size and its standard error [36]. The *z*-statistic was computed as a ratio of the pooled effect size to its standard error for each gene, and the result was compared to a standard normal distribution to obtain a nominal *p*-value. *P*-values were corrected for multiple hypotheses testing using Benjamini-Hochberg correction. The second approach uses a non-parametric meta-analysis that combines *p*-values from individual studies. A moderated *t*-statistic based on empirical Bayesian method was calculated for each gene in each study [37]. Fisher's sum of logs was used to combine the *p*-values obtained from each study and were corrected for multiple hypotheses using Benjamini-Hochberg correction. The corrected *P* value, False Discovery Rate (FDR) less than 5% was used to select the differentially expressed genes between intracranial and spinal ependymomas.

### NF2 gene correlation analysis

The correlation between NF2 gene expression and other genes was calculated using the Pearson correlation method for each data set separately. DerSimonian-Laird random-effect meta-analytical approach implemented in the metacor R package was used to combine correlation coefficients obtained from each individual datasets [38]. The correlation coefficients were transformed to Fisher's *z*-scores. *P* values obtained from the meta-analysis were corrected with Benjamini-Hochberg method. We used the *Z*-mean of correlation coefficients > 0.4 and FDR < 0.05 to select significantly correlated genes with NF2 gene expression.

### Gene set enrichment and pathway analysis

The over-representation analyses for Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Panther pathways, were carried out with GeneCodis [39]. The REVIGO software was used to summarize and visualize significant GO terms [40]. The overlap between differentially expressed genes and chromosomal positions (C1: positional gene sets collection) was investigated using the molecular signature database (MSigDB) version 4.0 [41]. The CS gene set was obtained from the published study [20]. The significance

of statistically enriched functional categories, pathways, and gene sets was estimated with hypergeometric test and the p-values were corrected for the multiple comparisons by estimating the FDR.

### **Prioritization of down regulated genes located in chromosome 22**

Network analysis of protein-protein interactions (PPI) was used to prioritize significantly down regulated genes in SEPNe that are located in chromosome 22. The PPI networks (including direct and indirect) among these genes were extracted from InWeb, a well-characterized PPI database developed by Lage *et al* [42]. To evaluate whether down regulated genes located in chromosome 22 are significantly connected via PPIs, permutation test was used to assess the significance of networks built from PPI data. Disease Association Protein-Protein Link Evaluator (DAPPLE) was used to assess the significance of PPI network with 10,000 permutations [43].

### **ACKNOWLEDGMENTS**

This work was supported by the Air & Ground Charity, Gentleman's Night Out Fund and the Children's Brain Tumour Research Centre (CBTRC).

### **CONFLICTS OF INTEREST**

There is no conflict of interest that I should disclose.

### **REFERENCES**

1. Hasselblatt M. Ependymal tumors. Recent results in cancer research Fortschritte der Krebsforschung Progres dans les recherches sur le cancer. 2009; 171:51-66.
2. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW and Kleihues P. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. 2007; 114:97-109.
3. Abul-Kasim K, Thurnher MM, McKeever P and Sundgren PC. Intracranial spinal tumors: current classification and MRI features. Neuroradiology. 2008; 50:301-314.
4. Cooper PR. Outcome after operative treatment of intramedullary spinal cord tumors in adults: intermediate and long-term results in 51 patients. Neurosurgery. 1989; 25:855-859.
5. Helseth A and Mork SJ. Primary intraspinal neoplasms in Norway, 1955 to 1986. A population-based survey of 467 patients. Journal of neurosurgery. 1989; 71:842-845.
6. Parsa AT, Lee J, Parney IF, Weinstein P, McCormick PC and Ames C. Spinal cord and intradural-extraparenchymal spinal tumors: current best care practices and strategies. Journal of neuro-oncology. 2004; 69:291-318.
7. Ruda R, Gilbert M and Soffietti R. Ependymomas of the adult: molecular biology and treatment. Current opinion in neurology. 2008; 21:754-761.
8. Ebert C, von Haken M, Meyer-Puttlitz B, Wiestler OD, Reifenberger G, Pietsch T and von Deimling A. Molecular genetic analysis of ependymal tumors. NF2 mutations and chromosome 22q loss occur preferentially in intramedullary spinal ependymomas. The American journal of pathology. 1999; 155:627-632.
9. Lamszus K, Lachenmayer L, Heinemann U, Kluwe L, Finckh U, Hoppner W, Stavrou D, Fillbrandt R and Westphal M. Molecular genetic alterations on chromosomes 11 and 22 in ependymomas. International journal of cancer Journal international du cancer. 2001; 91:803-808.
10. Singh PK, Gutmann DH, Fuller CE, Newsham IF and Perry A. Differential involvement of protein 4.1 family members DAL-1 and NF2 in intracranial and intraspinal ependymomas. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2002; 15:526-531.
11. Korshunov A, Neben K, Wrobel G, Tews B, Benner A, Hahn M, Golanov A and Lichter P. Gene expression patterns in ependymomas correlate with tumor location, grade, and patient age. The American journal of pathology. 2003; 163:1721-1727.
12. Lukashova-v Zangen I, Kneitz S, Monoranu CM, Rutkowski S, Hinkes B, Vince GH, Huang B and Roggendorf W. Ependymoma gene expression profiles associated with histological subtype, proliferation, and patient survival. Acta Neuropathol. 2007; 113:325-337.
13. Palm T, Figarella-Branger D, Chapon F, Lacroix C, Gray F, Scaravilli F, Ellison DW, Salmon I, Vikkula M and Godfraind C. Expression profiling of ependymomas unravels localization and tumor grade-specific tumorigenesis. Cancer. 2009; 115:3955-3968.
14. Taylor MD, Poppleton H, Fuller C, Su X, Liu Y, Jensen P, Magdaleno S, Dalton J, Calabrese C, Board J, Macdonald T, Rutka J, Guha A, Gajjar A, Curran T and Gilbertson RJ. Radial glia cells are candidate stem cells of ependymoma. Cancer cell. 2005; 8:323-335.
15. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A and Chinnaiyan AM. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proceedings of the National Academy of Sciences of the United States of America. 2004; 101:9309-9314.
16. de Magalhaes JP, Curado J and Church GM. Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics. 2009; 25:875-881.
17. Schwartzbaum JA, Fisher JL, Aldape KD and Wrensch M. Epidemiology and molecular pathology of glioma. Nature

- clinical practice Neurology. 2006; 2:494-503; quiz 491 p following 516.
18. Kilari D and Mohile SG. Management of cancer in the older adult. *Clinics in geriatric medicine*. 2012; 28:33-49.
  19. Campisi J. Aging, cellular senescence, and cancer. *Annual review of physiology*. 2013; 75:685-705.
  20. Tacutu R, Budovsky A, Yanai H and Fraifeld VE. Molecular links between cellular senescence, longevity and age-related diseases - a systems biology perspective. *Aging*. 2011; 3:1178-1191.
  21. Hulsebos TJM, Oskam NT, Bijleveld EH, Westerveld A, Hermsen MA, van den Ouweland AMW, Hamel BC and Tijssen CC. Evidence for an ependymoma tumour suppressor gene in chromosome region 22pter-22q11.2. *Brit J Cancer*. 1999; 81:1150-1154.
  22. Kraus JA, de Millas W, Sorensen N, Herbold C, Schichor C, Tonn JC, Wiestler OD, von Deimling A and Pietsch T. Indications for a tumor suppressor gene at 22q11 involved in the pathogenesis of ependymal tumors and distinct from hSNF5/IN11. *Acta Neuropathol*. 2001; 102:69-74.
  23. Begnami MD, Palau M, Rushing EI, Santi M and Quezado M. Evaluation of NF2 gene deletion in sporadic schwannomas, meningiomas, and ependymomas by chromogenic in situ hybridization. *Hum Pathol*. 2007; 38:1345-1350.
  24. Birch BD, Johnson JP, Parsa A, Desai RD, Yoon JT, Lycette CA, Li YM and Bruce JN. Frequent type 2 neurofibromatosis gene transcript mutations in sporadic intramedullary spinal cord ependymomas. *Neurosurgery*. 1996; 39:135-140.
  25. Wellik DM. Hox patterning of the vertebrate axial skeleton. *Developmental dynamics : an official publication of the American Association of Anatomists*. 2007; 236:2454-2463.
  26. Cillo C, Faiella A, Cantile M and Boncinelli E. Homeobox genes and cancer. *Experimental cell research*. 1999; 248:1-9.
  27. Lopez-Otin C, Blasco MA, Partridge L, Serrano M and Kroemer G. The hallmarks of aging. *Cell*. 2013; 153:1194-1217.
  28. Bretscher A, Chambers D, Nguyen R and Reczek D. ERM-Merlin and EBP50 protein families in plasma membrane organization and function. *Annual review of cell and developmental biology*. 2000; 16:113-143.
  29. Giordano A and Avantaggiati ML. p300 and CBP: partners for life and death. *Journal of cellular physiology*. 1999; 181:218-230.
  30. Chan HM and La Thangue NB. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *Journal of cell science*. 2001; 114:2363-2373.
  31. Iyer NG, Ozdag H and Caldas C. p300/CBP and cancer. *Oncogene*. 2004; 23:4225-4231.
  32. Rajendran R, Garva R, Ashour H, Leung T, Stratford I, Krstic-Demonacos M and Demonacos C. Acetylation mediated by the p300/CBP-associated factor determines cellular energy metabolic pathways in cancer. *International journal of oncology*. 2013; 42:1961-1972.
  33. Johnson RA, Wright KD, Poppleton H, Mohankumar KM, Finkelstein D, Pounds SB, Rand V, Leary SE, White E, Eden C, Hogg T, Northcott P, Mack S, Neale G, Wang YD, Coyle B, et al. Cross-species genomics matches driver mutations and cell compartments to model ependymoma. *Nature*. 2010; 466:632-636.
  34. Witt H, Mack SC, Ryzhova M, Bender S, Sill M, Isserlin R, Benner A, Hielscher T, Milde T, Remke M, Jones DT, Northcott PA, Garzia L, Bertrand KC, Wittmann A, Yao Y, et al. Delineation of two clinically and molecularly distinct subgroups of posterior fossa ependymoma. *Cancer cell*. 2011; 20:143-157.
  35. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4:249-264.
  36. Ramasamy A, Mondry A, Holmes CC and Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS medicine*. 2008; 5:e184.
  37. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*. 2004; 3:Article3.
  38. Laliberté E. (2011). *metacore: Meta-analysis of correlation coefficients*. R package version).
  39. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM and Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome biology*. 2007; 8:R3.
  40. Supek F, Bosnjak M, Skunca N and Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*. 2011; 6:e21800.
  41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES and Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:15545-15550.
  42. Lage K, Karlberg EO, Stirling ZM, Olason PI, Pedersen AG, Rigina O, Hinsby AM, Tumer Z, Pociot F, Tommerup N, Moreau Y and Brunak S. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature biotechnology*. 2007; 25:309-316.
  43. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, International Inflammatory Bowel Disease Genetics C, Cotsapas C and Daly MJ. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS genetics*. 2011; 7:e1001273.