

# A Comparative Study of the Molecular Characteristics of Familial Gliomas and Other Cancers

JIE LU<sup>1\*</sup>, MARK G. BURNETT<sup>1</sup> and MAX SHPAK<sup>1,2,3</sup>

<sup>1</sup>NeuroTexas Institute Research Foundation, St. David's Medical Center, Austin, Texas, TX, U.S.A.;

<sup>2</sup>Center for Systems and Synthetic Biology, University of Texas, Austin, Texas, TX, U.S.A.;

<sup>3</sup>Fresh Pond Research Institute, Cambridge, MA, U.S.A.

**Abstract.** *Background: Familial cancers are those that occur among first-degree relatives without showing Mendelian patterns of inheritance. Materials and Methods: In this analysis, we compare the genomic characteristics of familial and sporadic cancers, with a focus on low-grade gliomas (LGGs) using sequence and expression data from the Cancer Genome Atlas. Results: Familial cancers show similar genomic and molecular biomarker profiles to sporadic cancers, consistent with the similarity in their clinical features. There are no statistically significant differences among somatic mutation, copy number variant, or gene expression patterns between familial and sporadic cancers; methylation profiles are the only class of molecular data to show significant differences. Conclusion: Familial cancers are likely driven by multiple, individually weak contributions to familiarity (i.e. large numbers of alleles and/or shared environmental risks). Consequently, these risk factors tend to be obscured by stronger confounding variables such as clinical or molecular variation among cancer subtypes.*

While the majority of cancers are sporadic, a significant fraction of the incidence of certain cancer types is accounted for by familial and hereditary cases. Hereditary cancers are those with Mendelian patterns of inheritance within families, the result of risk alleles with very high penetrance. Well-known examples include cancer risk alleles in the *BRCA1/2*

tumor suppressor genes where inherited mutations significantly increase the lifetime risk of breast, ovarian, and prostate cancers (1) by several orders of magnitude, and where cancer incidence is often characterized by significantly younger ages of onset and worse clinical outcomes (2). Additionally, hereditary cancers have often been found to have unique gene expression profiles and other molecular signatures (3-4).

In contrast, familial cancers are defined as occurrences of the same cancer type among several first-degree biological relatives within a family in the absence of clear Mendelian inheritance, suggesting the absence of risk alleles with individually strong effects. Familial cancers may be due to common behavioral and environmental risks (e.g. smoking) or, as a more interesting scenario, due to shared genetic backgrounds involving mutations across multiple loci. In the latter case, alleles at individual sites will have low penetrance and individually small phenotypic effects, so that the lifetime cancer risk of carriers is only slightly enhanced relative to individuals with "wildtype" alleles. For example, genome-wide association studies (GWAS) of glioma risk alleles reveal a pattern consistent with the genetic architecture of polygenic traits, with the majority of risk alleles being characterized by risk odds ratios of 1.2-1.5 (implying a very small increase in actual incidence because gliomas are rare in the general population). Liu *et al.* (5) identified 12 such risk alleles for familial gliomas using pedigree analysis, and Shete *et al.* (6) and Kinnersley *et al.* (7) each identified 5 glioma susceptibility loci using GWAS analyses (N.B. the 5 loci identified in (6) are distinct from the loci identified in (7)). While GWAS do not focus on familial cases per se, the low-penetrance risk alleles identified in GWAS potentially include those loci that may contribute to or associate with familial incidence.

There are many well-documented examples of clinical and molecular differences between the more common hereditary cancers and their sporadic counterparts, such as the much earlier ages of onset of breast and ovarian cancer and

\*Current address: Genetic Sciences Division, Thermo Fisher Scientific, Austin TX 78744, U.S.A.

Correspondence to: Max Shpak, St. David's Medical Center, 1015 E. 32nd Street, Suite 404, Austin, TX 78705, U.S.A. Tel: +1 5125448077, Fax: +1 5125445547, Email: shpak.max@gmail.com

Key Words: Familial cancer, low-grade glioma, gene expression, gene methylation, cancer risk alleles.

characteristic gene expression profiles in individuals with *BRCA1/2+* genotypes noted above. However, there have been relatively few studies attempting to determine whether familial cancers are etiologically distinct, or whether they have unique genomic profiles or other specific molecular signatures. Such analyses are particularly lacking for several types of brain cancers, including low-grade gliomas (LGG).

Gliomas are the most common class of primary brain cancers, with approximately 20,000 cases diagnosed annually in the United States (8). The most prevalent form is Glioblastoma multiforme (GBM), a WHO grade IV glioma, which can form *de novo* or in some cases progresses from low-grade astrocytomas (9). Survival times for low-grade gliomas (LGG) are typically 10 years or more following diagnosis, with most deaths due to recurrence as higher-grade tumors, while most GBM patients die within less than 1.5 years (10). The majority of glioma cases are sporadic and independent of an individual's genetic background; a comparatively smaller fraction of gliomas (<5%) are familial (9-10). In contrast, hereditary gliomas are extremely uncommon, and are primarily restricted to individuals with very rare genetic disorders such as Li-Fraumeni syndrome (inherited loss of function mutations in tumor suppressor *TP53*) and neurofibromatosis (due to inherited mutations in *NF1*, a tumor suppressor gene modulating activity in the RAS signaling pathway). These account for a statistically negligible fraction of glioma cases (11-12), in contrast to the ~5% which are familial in the broad sense (13-16).

The familial glioma consortium (*e.g.* (5-7,13-16)) has performed extensive GWAS analyses to identify the inherited risk alleles for gliomas, and as part of their study, Sadetzki *et al.* (17), documented the absence of statistically significant differences in the ages of onset between familial and sporadic occurrences. Results suggesting similarity in clinical characteristics between familial and sporadic tumors raise the question of whether the molecular profiles of cells in familial glioma tumors have any specific features, or whether, like the clinical features, sporadic and familial gliomas have similar molecular characteristics.

In this study, we will compare the molecular profiles of familial gliomas with the profiles of sporadic instances to determine whether there are any specific molecular signatures in the former, and to determine whether there are significant associations between familial incidence and specific histological/molecular subtypes of glioma. In order to validate and identify mutations contributing to familial incidence, we will also compare the germline mutations of patients with familial gliomas to sporadically occurring glioma cases. Finally, we will compare gene expression and methylation profiles between familial and sporadic cancers in six additional common cancer types, in order to determine whether there are any shared molecular characteristics among familial cancers in general.

Table I. TCGA datasets summarized by cancer type. The *p*-values are based on a *t*-test comparison of age differences.

	Total	Familial	Frequency	Age (Familial)	Age (Sporadic)	<i>p</i> -Value
BLCA	412	13	3.16	69.7	68.0	0.501
COAD	459	58	12.64	64.9	67.9	0.106
ESCA	185	12	6.49	62.4	60.9	0.656
LGG	347	14	4.04	41.4	42.1	0.770
PAAD	187	12	6.42	65.8	64.8	0.760
STAD	443	18	4.06	67.2	64.7	0.306
THCA	507	46	9.07	50.6	46.9	0.105

## Materials and Methods

**TCGA datasets.** The clinical and molecular data of 7 different cancer types were downloaded from the Cancer Genome Atlas (TCGA: <https://tcga-data.nci.nih.gov/tcga/>): low-grade gliomas (LGG), bladder urothelial carcinoma (BLCA), esophageal carcinoma (ESCA), colon adenocarcinoma (COAD), pancreatic adenocarcinoma (PAAD), stomach adenocarcinoma (STAD) and thyroid carcinoma (THCA), based on availability of data on familial incidence. We identified familial cancers as those with a total of two or more cases within a family, restricting the definition to first-degree biological relatives (*i.e.* parents, children, or siblings) diagnosed with the same cancer type. Samples with no information on family history were excluded from the analyses. Clinical and demographic data describing family history of specific cancer type, and age at initial diagnosis are summarized in Table I. For LGG, familial cases were further categorized based on histologic diagnosis, *i.e.* astrocytomas, oligodendrogliomas and mixed (oligoastrocytomas).

**Genomic data processing.** RNA-seq and DNA methylation data were retrieved for all 7 cancer types. The level 3 RNA-seq gene expression data was transformed to a base-2 logarithmic scale. Level 3 DNA methylation data from two platforms (Illumina methylation arrays 27 and 450) were combined by intersecting the probe sets, excluding the 10.1% of samples with more than 5% missing values. Missing values in the remaining probes were imputed using the median value across samples.

For more detailed analysis of LGG patient data, additional genomic profiling including somatic mutation, germline mutation (Affymetrix SNP array), and Copy Number Variation (CNV) were retrieved from TCGA. Single nucleotide polymorphism (SNPs) data for germline point mutations were identified from the Level 2 TCGA SNP data. These SNPs were further processed to exclude low quality genotype calls (>10% missing values), rare alleles (<5% minor allele frequencies), and those loci with allele frequencies not in Hardy-Weinberg equilibrium ( $p < 10^{-4}$  in a Chi-square association test) using the pipeline described in (18). For the Level 3 CNV data, a weighted average CNV score was computed if a gene spanned multiple CNV segments where score weights for each gene were assigned in proportion to the fraction of the gene spanned by the respective segments.

**Statistical analysis.** To determine whether any germline or somatic mutations are enriched among familial gliomas, Fisher's exact test was performed on the relative frequencies of each SNP allele in

Table II. Differential gene expression analysis and gene ontology enrichment. The up/down- regulated genes refer to familial in comparison to sporadic cases. The gene ontology enrichment analysis is based on DEGs defined by unadjusted  $p$ -values.

Cancer Type	Number of DEGs ( $p < 0.05$ )	Up/Down regulated	$q < 0.05$	Bootstrap $p$ -value	Enriched Term	Fold Enrichment	$q$ -value
STAD	91	19/72	11	0.44	Zymogen	9.23	0.001
					Collagen alpha 1(V) chain	115.56	0.011
					Serine protease	9.33	0.028
					Protease	4.37	0.041
LGG	22	15/7	0	0.65	Extracellular region part	7.99	0.009
					Extracellular region	4.45	0.016
PAAD	84	25/59	3	0.15	Ionic channel	7.56	0.001
					Gated channel activity	6.28	0.005
					Ion transport	4.58	0.010
					Voltage-gated channel	9.62	0.021
					Glycoprotein	1.78	0.024
THCA	18	17/1	0	0.03	Metabolism of xenobiotics by cytochrome <i>P450</i>	36.32	0.035
BLCA	60	19/41	5	0.98	N/A		
COAD	8	8/0	0	0.47	N/A		
ESCA	285	195/90	35	0.02	N/A		

the familial vs. sporadic cases. Significantly differentially expressed or differentially methylated genes (DEGs/DMGs) were identified as follows: a Student's  $t$ -test was performed for each probe to compare mean expression or methylation levels between the familial vs. sporadic sample sets. The significance of association between CNVs and familial gliomas was tested using logistic regression, where a univariate model was fitted for each gene, with the (continuous) copy number as a predictor of  $Y=0,1$  (familial vs. sporadic). Because each application of the  $t$ -test to these data involves thousands of comparisons and therefore a high probability of false positive results, we applied the Benjamin-Hochberg False Discovery Rate (FDR) to compute the adjusted  $q$ -values (19). We used  $q < 0.05$  and absolute 2-fold change ( $\log_2 \text{FCI} > 1$ ) as cutoffs for DEGs and  $q < 0.01$  for DMGs. When testing for significant association in a specific gene, the unadjusted  $p < 0.05$  was used as the criterion for statistical significance. To evaluate whether the identification of DEGs/DMGs in familial cases was a consequence of sampling bias, a bootstrap procedure was performed as follows:  $n$  samples (where  $n$  is the total number of familial cases) were randomly selected with replacement from all samples over 100 iterations. In each iteration, the number of DEGs/DMGs was tallied, so that the  $p$ -value could be computed as the frequency at which an equal or greater number than the observed DEGs/DMGs in familial cases was observed in the bootstrapped data. The subsets of DEGs/DMGs with significant familial association were analyzed for enrichment with respect to structural or functional features using the DAVID gene ontology (GO) tool <https://david.ncifcrf.gov/>. Significantly enriched GO terms were called with  $q < 0.05$ . For LGG, unsupervised hierarchical clustering of gene expression values was performed to compare the expression profiles of familial vs. sporadic cases (in order to determine whether the familial incidences of glioma form a subset defined by a single node in the dendrogram characterized by some subset of DEGs). In the cluster analysis, each sample is represented as a vector of expression values and

classified by pairwise Pearson Correlation Coefficient distance. Unless otherwise indicated, all other data processing and all analyses were implemented in Python 2.7.5 and R 3.0.3.

## Results

*Comparative study across cancer types.* The percentages of familial cancer cases range from 3.16% in BLCA to 12.64% in COAD (Table I). According to a  $t$ -test comparison, there is no significant tendency towards a younger age of onset between familial and sporadic cases in any of the cancer types.

*Gene expression.* The results of the differential gene expression and gene ontology enrichment analyses are summarized in Table II. In four of the cancer types (STAD, PAAD, BLCA and ESCA), we found very few genes that were significantly differentially expressed with respect to the FDR-adjusted  $q < 0.05$  cutoff, and none at all in LGG, COAD, and THCA. Therefore, for the gene expression data, we loosened the stringency and set unadjusted  $p < 0.05$  as the DEG cutoff for enrichment analysis. Among these sets of genes, significant enrichment in functional or structural categories is seen in cellular structural macromolecules, *e.g.* collagen alpha 1(V) chain in STAD, ionic channel in PAAD and extracellular region proteins in LGG. There are also genes related to digestion and metabolism among the DEGs, *e.g.* zymogen in STAD and metabolism of xenobiotics by cytochrome *P450* in THCA. There is no evidence of enrichment with respect to well-known cancer-related pathways such as signal transduction, transcriptional regulation, nor is there differential

Table III. Shared DEGs between 2 familial cancer types.

Gene	Up-regulated	Down-regulated	Description
<i>AGR3</i>	ESCA, PADD		Anterior gradient 3
<i>AGR9</i>	PADD	BLCA	Critical in the development of the neuromuscular junction
<i>CAPN9</i>	ESCA, PADD		Calpains
<i>CLDN4</i>	ESCA, STAD		Claudins
<i>COL11A2</i>	STAD	ESCA	Alpha chains of type XI collagen
<i>GRIA4</i>	THCA	PAAD	Glutamate receptors
<i>NLRP3</i>	COAD, THCA		Pyrin-like protein
<i>SLC23A2</i>	BLCA, ESCA		Sodium-dependent vitamin C transporters

Table IV. Differentially methylated genes in familial cancers.

Cancer Type	Number of DEMs ( $q < 0.01$ )	Hyper-methylated	Hypo-methylated	Bootstrap $p$ -value
BLCA	236	18	218	0.35
COAD	69*	10	59	0.33
ESCA	70	1	69	0.36
LGG	286	7	279	0.22
PAAD	21	14	7	0.17
THCA	38	14	24	0.60
STAD	564	4	560	0.18

\*A cutoff of  $q < 0.05$  was applied to COAD data.

expression in known tumor suppressor or oncogenes in any one of the seven cancer types. In summary, the gene expression profiles of familial cancers are nearly indistinguishable from corresponding sporadic cancers.

To identify potential shared characteristics among the seven cancer types, we compared the lists of DEGs and found 8 genes that are shared by 2 cancer types (Table III, no genes are shared by more than 2 cancer types). Five of these 8 shared genes are up-regulated in both cancer types, the remaining 3 are differentially expressed in opposite directions among types (*i.e.* up-regulated in one cancer type and down-regulated in the other cancer type).

**DNA methylation.** For most cancer types, a large number of DMGs were called under the cut-off of  $q < 0.01$ , with the exception of COAD, for which a less stringent cutoff of  $q < 0.05$  was applied due to the much higher number of DMGs (Table IV). There is a strong tendency for most of the DMGs to be hypomethylated rather than hypermethylated in all cancer types except in PAAD. Figure 1 shows the distribution of the difference in mean methylation scores between FLGG and SLGG across genes, with a mean and standard deviation at  $-0.0299$  and  $0.0298$  for this highly skewed distribution, which qualitatively suggests a strong trend of demethylation in the FLGG genomes. However, the statistical significance of

differential methylation between familial and sporadic cancers is not supported by bootstrap analysis of the data ( $p > 0.10$  for all data classes), perhaps as a consequence of imbalanced sample size.

### Additional Analyses of Familial LGG

**Germline mutations.** We first consider the 7 known familial glioma-associated SNPs identified in (4) and (5). These SNPs were retrieved from TCGA Affymetrix SNP array. Comparisons of familial incidence between mutant and reference genotypes identified odds ratios (OR)  $> 1$ , indicating higher occurrence of the variants in familial LGGs (FLGGs) versus sporadic LGGs (SLGGs). As can be seen from the summary of this data in Table V, despite the relatively large ORs, only rs565934 (*SOX5*) had a near-significant association (OR=3.11;  $p=0.07$ ) with FLGG.

Furthermore, among the 749,999 germline SNPs that passed our quality filters, 33954 were found to be significantly associated with FLGGs ( $p < 0.05$ ), but none of these associations are statistically significant following FDR correction. Most of these SNPs had ORs near zero for the variant alleles, suggesting under-representation of these variant alleles in FLGGs. However, there are 5 SNPs with OR $>5$ , four of which are associated with *PPP1R3A* (protein phosphatase 1).

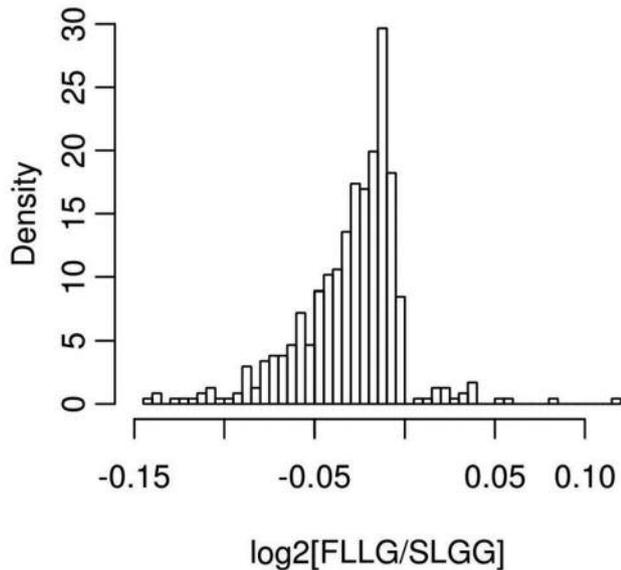


Figure 1. Heatmap of the gene expression levels in FLGG and SLGG samples. Hierarchical clustering (HC) on the expression levels of DEGs between FLGG and SLGG was used to classify the samples, with a row dendrogram (clustering of samples) based on Pearson correlation coefficient, the column dendrogram on a Spearman correlation coefficient.

**Somatic mutations.** We did not identify any somatic mutations with significant enrichment with respect to FLGG or SLGG after FDR correction ( $q < 0.05$ ). The somatic mutation most strongly associated with FLGG is a site in the coding region of *SPATA31E1*, with an OR=32.9 and  $p \sim 10^{-4}$ . This somatic mutation is found in 4 out of 14 FLGG cases, compared to 4 out of 333 in SLGGs. *SPATA31E1* is a homolog of a spermatogenesis-associated gene in the SPATA31 subfamily. Other somatic mutations strongly associated (OR>20) with FLGG include the ATPase *ATAD2b*, teneurin transmembrane protein *TENM1*, chromodomain protein *CDYL2*, CCR4-NOT transcription complex subunit 4 *CNOT4*, oculocutaneous albinism II *OCA2*, proliferation-associated 2G4 *PA2G4*, transcriptional regulating factor 1 *TRERF1* and zinc finger protein 485 *ZNF485*.

**Cluster analysis of gene expression profiles.** If we consider solely the set of DEGs between FLGG and SLGG, there is a weak tendency for FLGGs to fall into the astrocytoma subset (OR=2.91,  $p=0.055$ ). However, hierarchical clustering of samples by DEG expression levels does not identify any subtree (defined by a single node) containing FLGGs, although there does appear to be a relatively high incidence of familial cases (and astrocytomas) in the upper subtree in Figure 2. In contrast, the clustering analysis based on the subset of genes in the upper 5th percentile of variance in

Table V. The odds ratios of known familial risk SNPs in TCGA LGG dataset, with  $p$ -values based on a Fisher exact test.

SNPs	Associated gene	OR	$p$ -Value
rs1530364	<i>WNT9B</i>	5.63	0.49
rs2418087	<i>STYK1</i>	6.35	0.26
rs12050604	<i>UBR1</i>	1.49	1.00
rs7961199	<i>GRIN2B</i>	1.24	0.74
rs6764952	N/A	1.03	1.00
rs4689705	<i>SORCS2</i>	1.95	0.51
rs565934	<i>SOX5</i>	3.11	0.07

expression level indicates that neither FLGGs nor histologic classes formed any distinct subclusters (not shown).

**Copy number variations.** While 734 genes were found to have associations with  $p < 0.05$  with FLGGs (of these genes, 27 had associations with  $p < 0.001$ ), none of these association is statistically significant following FDR correction. Interestingly, the 27 genes with the strongest association with LGG had negative coefficients, suggesting a possible weak association between FLGG and a decrease in copy number in these genes. Among those with the strongest negative associations are the myosin regulatory subunit genes *MYL12A,B* and the homeobox protein *TGIF1* (a complete table of CNVs with FLGG associations with  $p < 0.001$  is available from the corresponding author upon request).

## Discussion

In contrast to the marked differences between certain hereditary cancers and their sporadic counterparts, this analysis of familial gliomas found that their typical age of onset and molecular profiles are generally similar to those in sporadic gliomas. For example, even though there was limited data on patient survival times, the fact that most patients in both cohorts survived beyond the evaluation period indicates that neither familial nor sporadic LGG tumors are more aggressive, nor are familial LGG's inherently more likely to progress to invariably and rapidly fatal GBMs. We also found similar ages of onset when sporadic vs. familial cases are compared in six other common cancer types.

The lack of clinical and etiological differences between familial and sporadic cancers (as described in (17) and confirmed by comparisons of patient ages in this study) is consistent with the similarly weak differences between familial and sporadic cases with respect to their genomic and molecular features. Very few genes are significantly differentially expressed between familial and sporadic cases

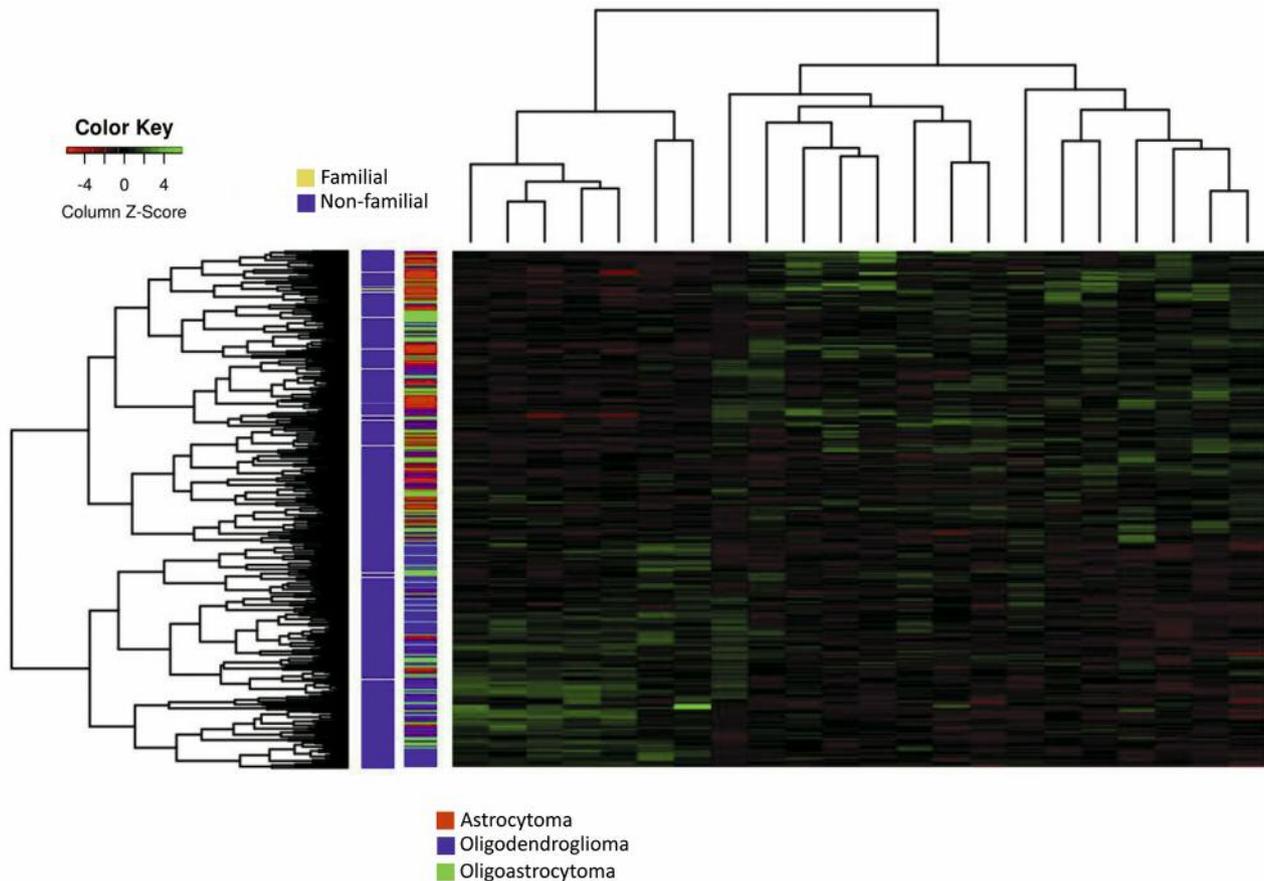


Figure 2. The differences in mean methylation scores between familial and sporadic LGG samples for all methylation probes, plotted as a frequency distribution.

if FDR-adjusted probabilities are used as a criterion (indeed, no genes in LGG are significantly differentially expressed according to the adjusted  $q$ -values). In contrast, significant differential methylation of genes is robust under FDR adjustment, and unlike the recent study (20) comparing long-term survivors to non-long term survivors among high-grade glioma cases, this cannot be attributed to differences in patient age (since there is no significance age difference between FLGG and SLGG patients). However, even DMG's statistical significance does not hold under bootstrap permutation tests in the LGG samples or in the data sets for most of the other cancer types considered in this study. The general absence of molecular biomarkers specific to familial cancers, together with comparable expected ages of onset, suggest that further clinical differences are probably minimal. Consequently, we should not expect different patient prognoses or responses to therapy between familial and sporadic gliomas.

The observed similarity in molecular profiles between sporadic and familial cancers is probably a consequence of

several contributing factors. First, in contrast to hereditary cancers, such as *BRCA*+ breast and ovarian cancers, which are typically driven by single mutations with high penetrance and strong effects, familial cancers are usually the consequence of multiple, individually weak risk alleles (shared genetic background) and/or in some cases a similarly weak and variable set of shared environmental risks. As a result, familial cancers are expected to be genetically heterogeneous across different families.

Furthermore, there is the potentially confounding issue of molecular heterogeneity within each cancer type. For example, in LGG there are three principal subtypes defined by the cells from which the tumor is derived: astrocytomas, oligodendrogliomas, and mixed (oligoastrocytomas), none of which have a strong association with familial occurrence. Likewise, the TCGA data doesn't distinguish between grades among the LGGs. This heterogeneity within LGG (and many other cancer types) is likely to be associated with greater molecular differences among the subclasses defined by histology and grade than any

differences that may be due to familial *vs.* sporadic origin. Additional sources of heterogeneity and error are due to the definition of familial cases in TCGA data. Specifically, TCGA records whether an LGG patient has immediate family members with some form of primary brain tumor, not necessarily LGG, with the same caveat likely to hold among the other cancer types.

Nevertheless, the fact that at least one of the SNPs identified in association with familial gliomas in previous GWAS and pedigree studies (rs565934, a site in the *SOX5* gene) is disproportionately represented in the familial LGG data supports the validity of the classification of samples as familial *vs.* sporadic based on TCGA's clinical data. The association of this SNP with familial cases in spite of the lack of additional molecular markers strongly associated with familiarity further supports the hypothesis that there may be multiple genetic backgrounds and gene expression profiles associated with familial gliomas, as opposed to a single set of genotypes defining familiarity.

While most of our results suggest an overall absence of molecular signatures specific to familial cancers, the occurrence (for some cases) of common genetic backgrounds due to shared multiple low-risk alleles implies that additional biomarkers of familial cancer may be found with more refined data sets and enhanced statistical power. This would require much larger samples of familial cancers within each type, as well as more specific division of cancers into histological and molecular subtypes to minimize the noise due to confounding variables. For example, large sample sizes allowed Bondy *et al.* (15) to identify germline mutations that were risk alleles for familial gliomas (only one of which, *SOX5*, appeared as nearly significant in our analysis), while restricting analysis to a single subtype led van Nistelrooij *et al.* (21) to identify younger average ages of onset and a small number of significant gene expression biomarkers that distinguish familial and sporadic esophageal cancer linked to Barrett's esophagus. However, if such differences exist, they are almost certainly weak and of limited clinical or biological importance. The sample sizes in the current study, which have >10-20 familial cases per cancer type, should provide sufficient statistical power to detect genomic signatures of familiarity if they were stereotypical across individuals and if the statistical association between these genetic biomarkers and familiarity were strong.

In future studies, we propose to perform extensions of our analyses on familial GBM once additional data become available. Unfortunately, TCGA does not record familial incidence as a clinical variable in their current GBM data sets, and as was the case with the LGG data, there are insufficiently many samples within any subtype of GBM to provide sufficient statistical power in analyses restricted to a single molecular or histological subtype.

## Acknowledgements

The Authors would like to thank Matt Cowperthwaite for his comments on the manuscript, and the St. David's Foundation Impact fund for its financial support.

## References

- Petrucelli N, Daly MB and Feldman G: Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. *Genet Med* 12(5): 245-259, 2010.
- Litton JK, Ready K, Chen H, Gutierrez-Barrera A, Etzel CJ, Meric-Bernstam F, Gonzalez-Angulo AM, Le-Petross H, Lu K, Hortobagyi GN and Arun BK: Earlier age of onset of BRCA mutation-related cancers in subsequent generations. *Cancer* 118(2): 321-325, 2012.
- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, and Friend SH: Gene expression profiling predicts clinical outcomes of breast cancer. *Nature* 514(6871): 530-536, 2002.
- Piccolo SR, Andrulis IL, Cohen AL, Conner T, Moos PJ, Spira AE, Buys SS, Johnson WE and Bild AH: Gene-expression patterns in peripheral blood classify familial breast cancer susceptibility. *BMC Medical Genomics* 8: 72, 2015.
- Liu Y, Melin BS, Rajaraman P, Wang Z, Linet M, Shete S, Amos CI, Lau CC, Scheurer ME, Tsavachidis S, Armstrong GN, Houlston RS, Hosking FJ, Claus EB, Barnholtz-Sloan J, Lai R, Il'yasova D, Schildkraut J, Sadetzki S, Johansen C, Bernstein JL, Olson SH, Jenkins RB, LaChance D, Vick NA, Wrensch M, Davis F, McCarthy BJ, Andersson U, Thompson PA, Chanock S; Gliogene Consortium and Bondy ML: Insight in glioma susceptibility through an analysis of 6p22.3 12p13.33-12.1, 17q22-23.2 and 18q23 SNP genotypes in familiar and non-familiar glioma. *Hum Genet* 131(9): 1507-1517, 2012.
- Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, Simon M, Marie Y, Boisselier B, Delattre JY, Hoang-Xuan K, El Hallani S, Idbaih A, Zelenika D, Andersson U, Henriksson R, Bergenheim AT, Feychting M, Lönn S, Ahlbom A, Schramm J, Linnebank M, Hemminki K, Kumar R, Hepworth SJ, Price A, Armstrong G, Liu Y, Gu X, Yu R, Lau C, Schoemaker M, Muir K, Swerdlow A, Lathrop M, Bondy M and Houlston RS: Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* 41(8): 899-904, 2009.
- Kinnersley B, Labussiere M, Holroyd M, Di Stefano A-L, Broderick P, Vijayakrishnan J, Mokhtari K, Delattre J-Y, Gousias K, Schramm J, Schoemaker MJ, Fleming SJ, Herms S, Heilmann S, Schreiber S, Wichmann H-E, Nothen MM, Swerdlow A, Lathrop M, Simon M, Bondy M, Sanson M and Houlston RS: Genome-wide association study identifies multiple susceptibility loci for glioma. *Nat Commun* 6: 8559, 2015.
- Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, Wolinsky Y, Kurchko C and Barnholtz-Sloan JS: CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro Oncol* 17(Suppl 4): iv1-62, 2015.

- 9 Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW: The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol* 131(6): 97-109, 2016.
- 10 Ohgaki H and Kleihues P: Epidemiology and etiology of gliomas. *Acta Neuropathol* 109(1): 93-108, 2005.
- 11 Johansson G, Andersson U and Melin B: Recent developments in brain tumor predisposing syndromes. *Acta Oncologica* 55: 401-411, 2016.
- 12 Kyritsis AP, Bondy ML, Rao JS and Sioka C: Inherited predisposition to glioma. *Neuro Oncol* 12(1): 104-113, 2010.
- 13 Melin B, Dahlin AM, Andersson U, Wang Z, Henriksson R, Hallmans G, Bondy ML, Johansen C, Feychting M, Ahlbom A, Kitahara CM, Wang SS, Ruder AM, Carreón T, Butler MA, Inskip PD, Purdue M, Hsing AW, Mechanic L, Gillanders E, Yeager M, Linet M, Chanock SJ, Hartge P and Rajaraman P: Known glioma risk loci are associated with glioma with a family history of brain tumours – a case-control gene association study. *Int J Cancer* 132(10): 2465-2468, 2013.
- 14 Blumenthal DT and Cannon-Albright LA: Familiality in brain tumors. *Neurology* 71(13): 1015-1020, 2008.
- 15 Bondy ML, Scheurer ME, Malmer B, Barnholtz-Sloan JS, Davis FG, Il'yasova D, Kruchko C, McCarthy BJ, Rajaraman P, Schwartzbaum JA, Sadtzki S, Schlehofer B, Tihan T, Wiemels JL, Wrensch M and Buffler PA (Brain Tumor Epidemiology Consortium): Brain tumor epidemiology: consensus from the Brain Tumor Epidemiology Consortium. *Cancer* 113(7 Suppl): 1953-1968, 2008.
- 16 Malmer B, Adatto P, Armstrong G, Barnholtz-Sloan J, Bernstein JL, Claus E, Davis F, Houlston R, Il'yasova D, Jenkins R, Johansen C, Lai R, Lau C, McCarthy B, Nielsen H, Olson SH, Sadtzki S, Shete S, Wiklund F, Wrensch M, Yang P and Bondy M: GLIOGENE an international consortium to understand familial glioma. *Cancer Epid. Biomarkers Prev* 16: 1730-1734, 2007.
- 17 Sadtzki S, Bruchim R, Oberman B, Armstrong GN, Lau CC, Claus EB, Barnholtz-Sloan JS, Il'yasova D, Schildkraut J, Johansen C, Housteon RS, Shete S, Amos CI, Bernstein JL, Olson SH, Jenkins RB, Lachance D, Vick NA, Merrell R, Wrensch M, Davis FG, McCarthy BJ, Lai R, Melin BS and Bondy ML: Description of selected characteristics of familial glioma gene patients – results from the Gliogene Consortium. *Eur J Cancer* 49(6): 1335-1358, 2013.
- 18 Shpak M, Hall AW, Goldberg MM, Derryberry DZ, Ni Y, Iyer VR and Cowperthwaite MC: An eQTL analysis of the human glioblastoma multiforme genome. *Genomics* 103(4): 252-263, 2014.
- 19 Storey JD and Tibshirani R: Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100(16): 9440-9445, 2003.
- 20 Lu J, Cowperthwaite MC, Burnett MG and Shpak M: Molecular predictors of long-term survival in Glioblastoma multiforme patients. *PLoS ONE* 11(4): e0154313, 2016.
- 21 van Nistelrooij AMJ, Dinjens WNM, Wagner A, Spaander MCW, van Lanschot JJB and Wijnhoven BPL: Hereditary factors in esophageal adenocarcinoma. *Gastrointest. Tumors* 1(2): 93-98, 2014.

*Received August 25, 2016*  
*Revised September 27, 2016*  
*Accepted October 10, 2016*