Sex-specific gene and pathway modeling of inherited glioma risk

Quinn T. Ostrom[1-3], Warren Coleman[4], William Huang[5], Joshua B. Rubin[6], Justin D. Lathia[7], Michael E. Berens[8], Gil Speyer[8], Peter Liao[1], Margaret R. Wrensch[9], Jeanette E Eckel-Passow[10], Georgina Armstrong[2], Terri Rice[9], John K. Wiencke[9], Lucie S. McCoy[9], Helen M. Hansen[9], Christopher I. Amos[11], Jonine L. Bernstein[12], Elizabeth B. Claus[13,14], Richard S. Houlston[15], Dora Il'yasova[16-18], Robert B. Jenkins[19], Christoffer Johansen[20], Daniel H. Lachance[21], Rose K. Lai[22], Ryan T. Merrell[23], Sara H. Olson[12], Siegal Sadetzki[24,25], Joellen M. Schildkraut[26], Sanjay Shete[27], Ulrika Andersson[28], Preetha Rajaraman[29], Stephen J. Chanock[29,30], Martha S. Linet[29], Zhaoming Wang[29,30], Meredith Yeager[29,30] (on behalf of the GliomaScan consortium^), Beatrice Melin[28], Melissa L. Bondy[2], Jill. S. Barnholtz-Sloan[1]

1. Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, Ohio, United States
2. Department of Medicine, Section of Epidemiology and Population Sciences, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas
3. Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio, United States
4. University School, Chagrin Falls, Ohio, United States
5. Case Western Reserve University, Cleveland, Ohio, United States
6. Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri; Department of Neuroscience, Washington University School of Medicine, St. Louis, Missouri, United States
7. Department of Stem Cell Biology and Regenerative Medicine, Cleveland Clinic Foundation, Cleveland, Ohio, United States
8. Cancer and Cell Biology Division, The Translational Genomics Research Institute, Phoenix, Arizona, United States

9. Department of Neurological Surgery, School of Medicine, University of California, San Francisco, San Francisco, California, United States
10. Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, United States
11. Institute for Clinical and Translational Research, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, United States
12. Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, United States
13. School of Public Health, Yale University, New Haven, Connecticut, United States
14. Department of Neurosurgery, Brigham and Women's Hospital, Boston, Massachusetts, United States
15. Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, United Kingdom
16. Department of Epidemiology and Biostatistics, School of Public Health, Georgia State University, Atlanta, Georgia, United States
17. Cancer Control and Prevention Program, Department of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina, United States
18. Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina, United States
19. Department of Laboratory Medicine and Pathology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota, United States
20. Oncology clinic, Finsen Center, Rigshospitalet and Survivorship Research Unit, The Danish Cancer Society Research Center, Copenhagen, Denmark
21. Department of Neurology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota
22. Departments of Neurology and Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States
23. Department of Neurology, NorthShore University HealthSystem, Evanston, Illinois, United States

24. Cancer and Radiation Epidemiology Unit, Gertner Institute, Chaim Sheba Medical Center, Tel Hashomer, Israel
25. Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel
26. Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, Virginia, United States
27. Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, United States
28. Department of Radiation Sciences, Faculty of Medicine, Umeå University, Umeå, Sweden
29. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States
30. Core Genotyping Facility, National Cancer Institute, SAIC-Frederick, Inc, Gaithersburg, Maryland, United States

Corresponding author

Jill S. Barnholtz-Sloan

Case Comprehensive Cancer Center,

Case Western Reserve University School of Medicine,

11100 Euclid Ave,

Cleveland, Ohio, 44106

Telephone: 216-368-1506

Fax: 216-368-2606

Email: jsb42@case.edu

^ Membership of the Gliomascan Consortium is presented in the acknowledgements

Short title: Sex-specific germline gene and pathway analysis in glioma

Funding: QTO is supported by a Research Training Grant from the Cancer Prevention and Research Institute of Texas (CPRIT; RP160097T). JWC was supported by the Grant S Roth Memorial Fund. WH was supported by the Young Scientist Summer Research Program. The GICC was supported by grants from the National Institutes of Health, Bethesda, Maryland (R01CA139020, R01CA52689, P50097257, and P30CA125123). Additional support was provided by the McNair Medical Institute and the Population Sciences Biorepository at Baylor College of Medicine.

In Sweden work was additionally supported by Acta Oncologica through the Royal Swedish Academy of Science (BM salary) and The Swedish Research council and Swedish Cancer foundation.

The UCSF Adult Glioma Study was supported by the National Institutes of Health (grant numbers R01CA52689, P50CA097257, R01CA126831, and R01CA139020), the Loglio Collective, the National Brain Tumor Foundation, the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research, the Robert Magnin Newman Endowed Chair in Neuro-oncology, and by donations from families and friends of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen. This project also was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 RR024131. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries,

under agreement # U58DP003862-01 awarded to the California Department of Public Health. The ideas and opinions expressed herein are those of the author(s) and endorsement by the State of California Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred.
UK10K data generation and access was organized by the UK10K consortium and funded by the Wellcome Trust.

Conflict of Interest: There are no conflicts of interest to report.

Total word count: 5872/6000

ABSTRACT

Background: To date, genome-wide association studies (GWAS) have identified 25 risk variants for glioma, explaining 30% of heritable risk. Most histologies occur with significantly higher incidence in males, and this difference is not explained by currently-known risk factors. A previous GWAS identified sex-specific glioma risk variants, and this analysis aims to further elucidate risk variation by sex using gene- and pathway-based approaches.

Methods: Results from the Glioma International Case-Control Study were used as a testing set, and results from three GWAS were combined via meta-analysis and used as a validation set. Using summary statistics for nominally significant autosomal SNPs ($p<0.01$ in a previous meta-analysis) and nominally significant X-chromosome SNPs ($p<0.01$), three algorithms (Pascal, BimBam, and GATES) were used to generate gene-scores, and Pascal was used to generate pathway-scores. Results were considered statistically significant in the discovery set when $p<3.3x10^{-6}$ and in the validation set when $p<0.001$ in 2/3 algorithms.

Results: 25 genes within 5 regions and 19 genes within 6 regions reached statistical significance in at least 2/3 algorithms in males and females, respectively. *EGFR* was significantly associated with all glioma and glioblastoma in males only, and a female-specific association in *TERT*, all of which remained nominally significant after conditioning on known risk loci. There were nominal associations with the Biocarta telomeres pathway in both males and females.

Conclusions: These results provide additional evidence that there may be differences by sex in genetic risk for glioma. Additional analyses may further elucidate the biological processes through which this risk is conferred.

Importance of the Study: Glioma, like most other cancer histologies, occurs at greater frequency in males than in females, and this difference is not explained by currently known risk factors. A previous sex-stratified analysis of these datasets identified three risk loci that varied in association between males and females. In this analysis, we attempted to leverage the summary statistics generated by this previous

analysis for additional discovery using gene- and pathway-based approaches. After conditioning on previously identified genetic risk loci, *EGFR* was significantly associated with all glioma and glioblastoma in males only, and a female-specific association in *TERT*. There were also nominal associations with the Telomeres, Telomerase, Cellular Aging, and Immortality pathway in both males and females. These results provide additional evidence that there may be biologically relevant significant differences by sex in genetic risk for glioma

INTRODUCTION

Glioma is the most common type of primary malignant brain tumor in the United States (US), with an average annual age-adjusted incidence rate of 6.0 per 100,000 population.[1,2] Glioma can be broadly classified into glioblastoma (GBM, 61.9% of gliomas in adults 18+ in the US) and lower-grade glioma (non-GBM glioma, 24.2% of adult gliomas). These tumors occur more commonly in people of European ancestry, in males and in older adults. Most glioma histologies occur with a 30-50% higher incidence in males, and this male preponderance of glial tumors increases with age (**Supplementary Figure 1**).[2]

Many environmental exposures have been investigated as sources of glioma risk, but the only validated risk factors for these tumors are ionizing radiation (which increases risk) and history of allergies or other atopic disease (which decreases risk).[3,4] A minority of glioma risk is thought to arise from heritable genetic risk factors, and the contribution of common low-penetrance single nucleotide polymorphisms (SNPs) to the heritability of glioma is estimated to be ~25%.[5] A recent glioma genome-wide association study (GWAS) meta-analysis validated 12 previously reported risk loci, and identified 13 new risk loci, and these 25 loci in total are estimated to account for ~30% of heritable glioma risk.[6] This suggests that there are both undiscovered environmental risk (which accounts for ~75% of disease incidence variance) and genetic risk factors (accounting for ~70% of heritable risk).[5,6]

Each individual GWAS results in regression estimates for hundreds of thousands of SNPs, only several hundred of which may meet the criteria for statistical significance to be prioritized for further investigation. While this process is appropriate for identifying individual loci that contribute to the development of disease, there is likely additional information about disease risk within these results that do not meet the stringent statistical significance thresholds used in GWAS (usually $p<5x10^{-8}$). Gliomas are known to be biologically complex, and as a result additional single-SNP analyses may not be appropriate to discover additional sources of genetic risk for these tumors. Multi-SNP methods--such as

gene or pathway-based approaches--can allow for additional discovery in a manner that complements single-SNP approaches, while substantially reducing the multiple testing burden associated with GWAS.[7]

While it is not likely that autosomal genomic sequence varies significantly by sex in the population, previous research has suggested that sex-related genetic variation may occur at the transcriptional and regulatory level.[8-10] One of the primary ways through which SNPs are thought to affect phenotype is through variation in gene regulation and expression.[11] Sex-specific variation in regulatory processes may also affect the relationship between SNPs and phenotype, as a result it may be possible that allele frequencies in risk SNPs vary between affected males and females. A recent sex-stratified GWAS identified three glioma risk loci that differ in effect by sex.[12] These three SNPs explain 1.4% of phenotypic variance in a pooled glioma sample (1.3% in males and 2.2% in females), and 0.6% of variance in GBM (0.9% in males and 0.7% in females). Other analyses have also identified sex-specific sources of risk for glioma, including an association study focused on the cAMP pathway identified SNPs in Adenylate Cylase 8 as a sex-specific modifier of risk for low-grade astrocytoma in Neurofibromatosis Type 1.[13] Genetic risk for complex traits is increasingly understood to be polygenic, and variations in risk may be the result of variation at hundreds of locations across the genome. Each individual SNP may only provide a very small contribution to genetic risk for a trait, and the mechanistic relationship between these individual, low-effect SNPs and phenotype is hard to estimate. Additional analyses using gene- and pathway-based approaches may further elucidate sex differences in genetic risk for glioma.

There is no consensus on the best method for generating gene- and pathway-based test statistics from GWAS summary statistics, and many different approaches have been developed. These pathway approaches have been utilized in other cancers to some success, but have not been widely used in glioma. The primary aim of this analysis was to contrast multiple gene-based approaches for leveraging currently existing sex-specific glioma summary statistics, as well as to assess whether these approaches may identify additional sources of genetic risk for glioma that may vary by sex.

METHODS

Summary statistics generated as part of a prior sex-specific GWAS[12] were used to estimate sex-specific gene and pathway scores. Data from four studies were divided into a testing set, and validation set. Results from the Glioma International Case-Control Study (GICC)[14,15] were used as a testing set (**Figure 1a**), and results from three prior glioma GWAS (San Francisco Adult Glioma Study GWAS,[16] MD Anderson Glioma GWAS,[17] and National Cancer Institute's Gliomascan[18]) were combined via inverse-variance weighted fixed effects meta-analysis in META[19] and used as a validation set for any statistically significant genes and pathways (**Figure 1a**). See **Supplementary Table 1** for an overview of characteristics for individuals included in these datasets, and **Figure 1** for an overview of the study schematic. Details of case ascertainment, genotyping, quality control, imputation and primary analysis of these datasets are available in Melin, et al. (GICC), Wrensch, et al. (San Francisco Adult Glioma Study GWAS), Shete, et al. (MD Anderson Glioma GWAS), and Rajaraman, et al. (Gliomascan).[14-18]

Sex-specific summary statistics study for autosomal markers were previously generated[20] using sex-stratified logistic regression models in SNPTEST[21] to estimate sex-specific betas (β_M and β_F), standard errors (SE_M and SE_F), and p-values (p_M and p_F) (**Figure 1a**). Summary statistics for the three studies used as a validation set were combined using META.[19] Only SNPs with minor allele frequency (MAF) ≥ 0.01, imputation INFO score ≥ 0.7 and $p<0.01$ in a previous eight-study pooled-sex meta-analysis[15], which included the four datasets used in this analysis, were used to generate gene- and pathway-scores. X chromosome data were available from GICC set only, and analyzed using logistic regression model in SNPTEST module 'newml' assuming complete inactivation of one allele in females, and males are treated as homozygous females. X chromosome SNPs with MAF ≥ 0.01, imputation INFO score ≥ 0.7, and single SNP association $p<0.01$ were used for generation of gene scores. Linkage disequilibrium (LD) information was based on structure within the European cases from the 1,000 genomes project phase 3

dataset. [22] All analyses were performed separately for males and females to identify genes and pathways with germline variation between cases and controls. Genes were prioritized that were identified by at least two of the three selected algorithms (**Figure 1b**). Analyses were conducted for glioma overall and for glioblastoma only by sex within each dataset.

Three algorithms (Pascal,[23] BimBam,[24] and GATES[25]) were used to generate gene-scores. Gene-based effects were assessed using SNPs within 50kb of each gene (using 5' and 3' UTR) as defined using the UCSC hg19 assembly. Pascal[23] calculates gene-scores using the VEGAS[26] scoring algorithm, generates a gene-based test statistic using sum-of-chi-squares (SOCS) correcting for LD structure (based on a reference set). Genes that are in LD are considered to be a 'fusion gene' and have only one gene-score calculated. BimBam[24] (as implemented in FAST using summary statistics[27]) is a Bayesian regression approach. This method calculates an average Bayes Factor for all K possible models within a gene, where K is the number of SNPs. The model then uses a Laplace method to estimate posterior distributions of the model's parameters, and distribution models are obtained using the Fletcher-Reeves conjugate gradient algorithm. GATES[25] (as implemented in FAST[27]) uses a modified Sims test that combines SNP-based p-values, using the p-value correlation matrix to estimate the number of independent SNPs within the gene. The resulting gene-based p-values approximate a uniform distribution. For all methods implemented within FAST, SNPs were excluded if they were in complete LD (r^2=1) with another SNP in the gene, which limited the amount of SNPs evaluated within each gene.

Pathway-scores were generated using Pascal,[23] using gene and fusion-gene-scores generated by the Pascal algorithm (**Figure 1c**). The pathway-score was then calculated using both independent and fusion genes. A parameter free enrichment strategy was used to calculate pathway-scores using either a chi-squared method (gene-score p-values were ranked and transformed to a uniform distribution, these values were then transformed by a chi-square quantile function, and summed) or an empirical sampling method (gene-scores are transformed with chi-square quantile function and summed, then Monte Carlo estimate of the

p-values were obtained by sampling random sets of the same size). Results from each gene and pathway algorithm were compared within each sex as well as between sexes. Pathway information was obtained from KEGG,[28] Reactome,[29] and Biocarta[30] (as defined in MSigDB[31,32]).

For genes within regions that contain SNPs previously identified as significant by GWAS, conditional analyses were run for all SNPs within those regions using SNPTEST and adjusted gene-scores were calculated. All figures were generated using R 3.3.2, ggplot2, graphite, network, Intergraph, ggnetwork, igraph, gridExtra, and LocusZoom.[33-39]

RESULTS

159,706 SNPs from the testing set and 163,115 SNPs from the validation set were included in gene-based analyses. Gene-scores were generated for ~16,000 genes and were considered significant at $p<3.3x10^{-6}$ (based on a Bonferroni correction for 15,000 tests). P-values in the validation set were considered significant at $p<0.001$ (based on a Bonferroni correction for 50 tests, for 25 total genes tested in each sex).

Among males, 25 genes within five regions had scores that reached the set significance threshold ($p<3.3x10^{-6}$) in at least two of three evaluated algorithms in all glioma or glioblastoma (See **Figure 2** and **Supplementary Table 2** for the strongest associations within each region of the six regions where genes met the set significance threshold). Among females, 19 genes within six regions had scores that reached the set significance threshold ($p<3.3x10^{-6}$) in at least two of three evaluated algorithms in all glioma or glioblastoma (See **Figure 2** and **Supplementary Table 3** for the strongest associations within each of the six regions where genes met the set significance threshold). Solute carrier family 6, member 18 (*SLC6A18*), Telomerase reverse transcriptase (*TERT*), and cyclin dependent kinase inhibitor 2B (*CDKN2B*), and stathmin 3 (*STMN3*) reached the set significance threshold in both males and females in glioblastoma, while *SLC6A18*, *TERT*, and *STMN3* reached the set significance threshold in both sexes in all glioma. All shared associations validated.

Epidermal growth factor receptor *(EGFR)*, dynein axonemal heavy chain 2 (*DNAH2*), and several genes surrounding regulator of telomere elongation helicase 1 (*RTEL1*) on chromosome 20 (with the strongest association in *RTEL1*-TNF receptor superfamily member 6b [*RTEL1-TNFRSF6B*]) reached the significance threshold in males only (**Figure 2**, **Supplementary Table 2**). In all glioma, *CDKN2A* reached the set significance threshold in males only. All genes validated in males. Blepharophimosis, epicanthus inversus and ptosis, candidate 1 (non-protein coding) (*BPESC1*) reached the significance threshold in all glioma in females only (**Figure 2**, **Supplementary Table 3**), but this association was not confirmed in in the validation set. The association in *EGFR* was nominally significant in males after conditioning on three SNPs previously identified by GWAS within this gene (rs75061358, rs723527, and rs11979158), including one (rs11979158) that has previously been identified as having a sex-specific effect (**Supplementary Tables 4-5)**. When conditional single-SNP associations were examined by sex and histology in *EGFR*, there was nominally significant peak apparent in both males and females, with no single SNP that approached genome-wide significance (**Supplementary Figure 3**). The association at *TERT* was nominally significant for females in glioblastoma only after conditioning on the previous identified SNP (**Figure 3, Supplementary Table 5**). When conditional single-SNP associations were examined by sex and histology in *TERT*, a single SNP of nominal significance (rs7705526) was identified upstream of the previously identified SNP (rs10069690, **Supplementary Figure 3**). This region was apparent in males and females. Associations in *STMN3* and *RTEL1-TNFRSF6B* remained nominally significant after conditioning in both males and females (**Figure 3, Supplementary Tables 4-5**). When conditional single-SNP associations were examined by sex and histology in *RTEL1-TNFRSF6B*, nominally significant SNPs were identified across the gene in both males and females, with no apparent additional signal (**Supplementary Figure 4**). There was no substantial difference in effect size by sex in the most significant SNP in either gene (**Supplementary Figure 5**).

There were 202,886 X chromosome SNPs with MAF≥0.01 and INFO score≥0.7 in the GICC dataset. Gene-scores were calculated for 56 X chromosome genes with at least five SNPs, and associations were considered significant at $p<8.3x10^{-4}$ (based on a Bonferroni correction for 60 tests). There were 12 genes within 4 chromosomal regions that reached the significance threshold in at least two of three algorithms (Results from the strongest association in each region are shown in **Table 1**). Shroom family member 2 (*SHROOM2)* (Xp22.2), and armadillo repeat containing, X-linked 2 (*ARMCX2*) (Xq22.1) were significantly associated with both all glioma, and glioblastoma, while dystrophin (*DMD*) (Xq21.2-p21.1) was significantly associated with all glioma only and zinc finger protein 185 with LIM domain (*ZNF185*) was significantly associated with glioblastoma only.

There were 1,077 pathways in the combined KEGG, Biocarta, and Reactome sets, and associations were considered statistically significant in the discovery set at $p<5x10^{-5}$ (based on a Bonferroni correction for 1,000 tests), and significant in the discovery set at $p<0.00883$ (based on a Bonferroni correction for 6 tests). No pathways reached the set significance threshold, but there were several nominally significant associations. The Telomeres, Telomerase, Cellular Aging, and Immortality pathway reached nominal significance in both males and females in all glioma, and glioblastoma (**Table 2**). When the gene-scores for the genes contained within this pathway were examined, the association with this pathway was driven primarily by strong associations in *TERT*, and *TP53* (**Figure 4**). There were nominally significant associations in *POLR2A* (in both males and females) and *PRKCA* (in males only), both genes that have not been significantly associated with glioma to date. Further interrogation of the single-SNP results for these genes found no associations significant at the $p<5x10^{-4}$ level in either sex or histology group.

Nominally significant associations were identified in five cancer-specific KEGG pathways: bladder cancer, glioma (**Supplementary Figure 6**), melanoma (**Supplementary Figure 7**), non-small cell lung cancer, and pancreatic cancer (**Table 2**). There is significant overlap between these gene-sets **(Supplementary Figure 8)**, and when the gene-scores used to build each pathway were examined all the

associations appear to be driven largely by strong associations in *EGFR* and *CDKN2A* which are members of all KEGG cancer pathways found to be nominally associated with glioma in this analysis. Pathway analyses were run using single-SNP results including conditional analyses for all SNPs within a 2mb window around the previously identified SNPs near *TERT*, *EGFR*, *CDKN2B*, *TP53*, and *RTEL1*. All pathway associations no longer reached the significance threshold when analysis included conditioned results (**Table 2**).

DISCUSSION

This represents the first genome-wide sex-specific gene- or pathway-based analysis for germline risk variants in glioma. Gene-based tests are an efficient way to increase power to detect associations of low effect size, where multiple variants within a region may contribute to increased risk. Multi-marker tests, such as gene- or pathway-based tests, allow investigators to leverage previously existing GWAS summary statistics for discovery as well as to increase power when strength of association for single-SNP associations may be low. Incidence of glioma is significantly higher in males as compared to females, and currently identified environmental risk factors do not explain this variation in incidence[1]. A previous sex-specific GWAS analysis identified three loci with sex-specific effects, including a previously identified SNP in 7p11.2 (rs11979158, proximate to *EGFR*).[12] As a primary goal, this analysis aimed to compare existing gene- and pathway-based methods in the context of heritable genetic risk for glioma, and additionally to explore additional potential sources of genetic risk that may contribute to sex differences in genetic risk for glioma. All autosomal genes identified by this analysis were proximate to previously identified GWAS hits. After conditioning on previously identified SNPs, associations at *TERT*, *EGFR* and *RTEL1* remained nominally significant. The results of this conditional analysis suggest that there are remaining sources of genetic risk for glioma within these regions, including one apparent region in *TERT* (**Supplementary Figure 3**), with a single-SNP association that approaches genome-wide significance in both sexes (**Supplementary Figure 5**). There were no differences by sex in effect size and direction,

which suggests that while there may be remaining genetic risk associations to be detected within these genes, they do not have sex specificity.

Four regions on the X chromosome (Xp22.2, Xp21.2-p21.1, Xq22.1, and Xq28) contained genes that reached the significance threshold in at least two of three algorithms (**Table 1**). These genes have not been previously associated with glioma. SNPs near *SHROOM2* (Xp22.2) were previously associated with prostate and colon cancer.[40-42] There are no known associations with inherited variants in the other three regions and increased risk for cancer, though all contain genes that have been shown to be dysregulated in some cancer cells (e.g. *DMD*, *ARMCX2*, *ZNF185*).[43-47] Without a validation set, it is not possible to know if these are true associations or the result of type 1 error. Further exploration of these genes is necessary to determine their true relationship with glioma risk.

The Biocarta *Telomeres, Telomerase, Cellular Aging, and Immortality* pathway reached nominal significance in both males and females in all glioma, and glioblastoma (**Table 2**). This pathway contains *EGFR*, *TERT*, and *TP53*, all of which contain SNPs identified by glioma GWAS. Inherited variants affecting telomere length have been associated with many complex diseases, including glioma.[48-50] Both age and sex are known to affect telomere length, and previous research has suggested that males have shorter telomeres and higher rates of telomere attrition with aging.[51,52] An analysis comparing a weighted genetic score based on eight SNPs associated with leukocyte telomere length found that telomere length was ~5% longer in glioma cases versus controls.[53] The significance of the telomere maintenance pathway may explain the remaining significant association in the regions surrounding *TERT*, *EGFR*, and *RTEL1*, as any variants affecting telomere length could contribute to glioma risk. In addition to the strong associations in genes associated with SNPs previously identified by GWAS, there were nominally significant associations in *POLR2A* (in both males and females) and *PRKCA* (in males only).

The numerous KEGG cancer pathways found to be significant in this analysis are likely due to the strength of association in genes (*CDKN2A* and *EGFR*) that are members of all identified cancer-specific KEGG pathways. While these associations are driven by strong associations in these specific genes, they may also be evidence of shared sources of genetic risk between these cancers and glioma. Both the glioma and melanoma pathways, driven by strong associations in *CDKN2A*, were significantly associated with all glioma in males (**Supplementary Figures 6-7)**. Previous analyses suggested an association between genetic risk for glioma and melanoma, both in terms of known cancer syndromes (most notably Melanoma-neural system tumor syndrome, caused by inherited variants in *CDKN2A*[4]), familial glioma and sporadic disease.[54-56] Persons with a previous diagnosis of melanoma are estimated to have incidence of glioma that is 1.42 times that of the general population, while relatives of glioma patients ha are diagnosed with melanoma approximately 2-4 times as frequently as the general population.[54-56] Melanoma GWAS to date have identified 21 genetic risk loci,[57,58] including SNPs near *CDNK2A* and *TERT*, genes that have also been associated with glioma.[6] The identified SNPs in these genes do not account for a large proportion of risk in either melanoma or glioma, but there is evidence that innate telomere length and variation in telomere maintenance pathways may contribute to risk in both diseases.[59] When pathway-analyses were re-run using single-SNP results conditioned on known GWAS hits, pathway associations no longer reached the significance threshold. Gene-specific p-values for *TERT*, *EGFR*, and *RTEL1* were lowest for conditional analyses performed in Pascal, the algorithm used to calculate pathway-scores as compared to the other two algorithms. The SOCS approach used by Pascal may be conservative than others if there are many genes with null association and few genes with significant associations. Other pathway scoring algorithms that are more sensitive to a smaller set of strong associations may be more sensitive in identifying pathway associations.

All genetic association tests require consideration of the implicit assumptions about the genetic architecture of the disease and population of interest. GWAS approaches have attempted to identify single variants that have a causal relationship with a phenotype, which requires that this variant occur repeatedly

within the study population. Gene- and pathway-based tests assume that the aggregate effect of variants within a gene or pathway affect disease risk, but do not require that all individuals possess the same variant. These approaches are most appropriate for complex diseases where risk for disease is polygenic. In contrast to the logistic regression methods utilized by GWAS, the three analytic approaches used here do not generate measures of the magnitude of association. These methods test for enrichment of associations at single SNPs within genes, without consideration of the magnitude or direction of association. Further analysis of the identified regions is necessary to

While multi-marker tests can increase power to detect associations as compared to single-SNP tests, different methods may be better suited to particular types of genetic architecture. Methods vary in their performance based on whether a gene has one strong signal versus multiple signals of lower significance. One of the genes identified by this analysis is known to have at least two independent GWAS signals (*EGFR*),[6] and as a result, its identification in some methods may be affected by this bias.

There is a well-known bias in GWAS towards large genes,[60] which are often enriched for tag SNPs, and this bias may influence the results of this analysis. All of the algorithms used for this analysis can be affected by gene size. Large genes with many SNPs of minimal significance and few SNPs of large effect may 'dilute' the gene-score in methods based on summed scores, such as Pascal. All three of the algorithms used for this analysis 'prune' SNPs based on linkage disequilibrium statistics in attempt to obtain a set of independent SNPs. For large genes that contain multiple haplotype blocks, results may still be biased towards large genes. This analysis used a relatively large window surrounding the defined genes (+/- 50kb) which may further bias analyses towards large genes. While estimates of average gene size range from 10-15kb, the average size for genes included in the annotation file for this analysis was 43.32 kb. The three major genes identified by this analysis range in size, but all are larger than the estimated average: *TERT* (29.84kb), *EGFR* (137.92kb), and *RTEL-TNFRSF6B* (39.8kb). Genes that did not remain significant after conditioning tended to be smaller: *BPESC1* (20.98kb), *SLC6A18* (20.84kb), *CDKN2B*

(6.4kb), *DNAH2* (26.14kb), and *STMN3* (13.06kb). As a result, it is possible that these methods may be biased towards identifying smaller amounts of remaining signal in larger genes as compared to smaller genes. This analysis utilized a 50kb window surrounding a gene, and it is likely that changing this window may change the identified associations. These methods will also fail to identify any associations in intergenic regions, particularly the region at 8q24.21 that has previously been identified as having a sex-specific association.[12]

Patterns of linkage disequilibrium within the study population may also significantly affect the performance of a method. Results for methods that use LD information, including all algorithms evaluated in this analysis, may also be significantly altered by the reference populations to estimate LD. All of the included methods attempt to adjust for potential score inflation due to LD, using the 1,000 EUR super population as a reference set. FAST does this by pruning SNPs that are in complete linkage (r^2=1), while Pascal does this by generating 'fusion' gene-scores for genes that are in linkage with each other. These 'fusion' genes are used along with single-gene-scores to generate pathway-scores to decrease inflation of p-values due to the physical proximity of genes.[23] Due to variations in adjustment for LD used in the two programs, the number of included SNPs by each gene varied slightly. Both methods require that the identifier for each variant in the summary statistics be present in the LD reference file, and as a result these methods are not able to incorporate variants that do not have a standard reference SNP cluster ID (RSID). FAST additionally limits the dataset by requiring that all markers be bi-allelic SNPs, and does not accept indels.

Different multi-marker approaches may also perform better than others based on the computational resources available for an analysis. Permutation-based tests are more computationally intensive as compared to parametric tests, especially when gene-scores are calculated genome-wide. Both Pascal and GATES do not rely on permutations for estimating p-values, which significantly decreases analysis time. BimBam uses permutations to calculate exact p-values, as a result these analyses require more time to

complete. The number of permutations used to calculate determines the boundaries for an exact p-value (ranging from 1 to 1/*n*, where *n* is the number of permutations), which may result in increasing permutations for increased p-value specificity. For more stringent p-value cut-offs, such as when testing multiple phenotypes in multiple groups, the number of permutations required may substantially increase analysis time. While multi-marker tests do substantially decrease the multiple testing burden as compared to genome-wide single SNP approaches, it is still important to consider multiple testing when conducting these tests. In cases such as this analysis, where multiple phenotypes are being tested within population strata, multiple testing correction strategies should be used. The p-value threshold used for this analysis are adjusted only for the number of genes/pathways within the testing phase. Use of more stringent testing cut-offs may result in prioritization of fewer genes. Use of a more stringent threshold in the testing phase of $p<7.8x10^{-7}$ (based on a Bonferroni correction for 64,000 tests, for 16,000 genes in two phenotypes and two sexes) would result in *DNAH2* not reaching the set significance threshold in males, and *BPESC1* and *STMN3* not reaching the significance threshold in females (**Supplementary Tables 2** and **3**). A Bonferroni correction is known to be conservative, and use of these very stringent cut offs may result in rejection of 'real' associations. Use of two-stage testing and a validation stage provide an additional safeguard against type 1 error.

In addition to the technical limitations of the three algorithms utilized for this analysis, there are several limitations. All glioma cases from the included four GWAS datasets were recruited at time of first diagnosis, and the assigned diagnoses represent the primary tumor type according to the prevailing histologic criteria at that time. There may also be variation in the histologies contained within each set by sex. The proportion of each dataset that is composed of glioblastoma as compared to lower grade gliomas varies by both study and sex (**Supplementary Table 1**). Less than 50% of female glioma cases in the testing set are glioblastoma, whereas over 50% of female cases are glioblastoma in the validation sets. Glioma is a heterogonous disease, and due to all of these factors, it is likely that heterogeneity exists between the utilized datasets.

CONCLUSIONS

Multi-marker tests, such as gene- or pathway-based tests, allow investigators to leverage previously existing summary statistics and increase power when strength of single-SNP associations may be low. This analysis aimed to explore additional potential sources of genetic risk that may contribute to sex differences in genetic risk for glioma. There was a nominally significant association between germline variants in *RTEL1* in both males and females after conditioning on previously identified SNPs. A significant association was detected between germline variants in the telomere maintenance pathway in both males and females, which builds on previous evidence of the relationship between inherited variants related to increased telomere length and increased risk for glioma. There was also a male specific association in *EGFR*, and a female-specific association in *TERT* that remained nominally significant after conditioning on previous GWAS hits. The results of this analysis confirm previously known information about inherited glioma risk, and provide potential mechanistic explanations for how these variants may affect the process of gliomagenesis.

FUNDING

QTO is supported by a Research Training Grant from the Cancer Prevention and Research Institute of Texas (CPRIT; RP160097T). JWC was supported by the Grant S Roth Memorial Fund. WH was supported by the Young Scientist Summer Research Program. The GICC was supported by grants from the National Institutes of Health, Bethesda, Maryland (R01CA139020, R01CA52689, P50097257, and P30CA125123). Additional support was provided by the McNair Medical Institute and the Population Sciences Biorepository at Baylor College of Medicine.

In Sweden work was additionally supported by Acta Oncologica through the Royal Swedish Academy of Science (BM salary) and The Swedish Research council and Swedish Cancer foundation. The UCSF Adult Glioma Study was supported by the National Institutes of Health (grant numbers R01CA52689,

P50CA097257, R01CA126831, and R01CA139020), the Loglio Collective, the National Brain Tumor Foundation, the Stanley D. Lewis and Virginia S. Lewis Endowed Chair in Brain Tumor Research, the Robert Magnin Newman Endowed Chair in Neuro-oncology, and by donations from families and friends of John Berardi, Helen Glaser, Elvera Olsen, Raymond E. Cooper, and William Martinusen. This project also was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through UCSF-CTSI Grant Number UL1 RR024131. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #U58DP003862-01 awarded to the California Department of Public Health. The ideas and opinions expressed herein are those of the author(s) and endorsement by the State of California Department of Public Health, the National Cancer Institute, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors is not intended nor should be inferred. UK10K data generation and access was organized by the UK10K consortium and funded by the Wellcome Trust.

ACKNOWLEDGMENTS

The results of this study were previously presented at the 2017 Annual Meeting of the Society for Neuro-Oncology. A pre-publication version of this manuscript was previously made available on bioRxiv (doi: https://doi.org/10.1101/235408)

Other significant contributors for the UCSF Adult Glioma Study include: M Berger, P Bracci, S Chang, J Clarke, A Molinaro, A Perry, M Pezmecki, M Prados, I Smirnov, T Tihan, K Walsh, J Wiemels, S Zheng. Glioma scan group comprised: Laura E. Beane Freeman, Stella Koutros, Demetrius Albanes, Kala Visvanathan, Victoria L. Stevens, Roger Henriksson, Dominique S. Michaud, Maria Feychting, Anders Ahlbom , Graham G. Giles Roger Milne , Roberta McKean-Cowdin, Loic Le Marchand, Meir Stampfer, Avima M. Ruder, Tania Carreon, Goran Hallmans, Anne Zeleniuch-Jacquotte, J. Michael Gaziano, Howard D. Sesso, Mark P. Purdue, Emily White , Ulrike Peters, Howard D. Sesso, Julie Buring.
We are grateful to all the patients and individuals for their participation and we would also like to thank the clinicians and other hospital staff, cancer registries and study staff in respective centers who contributed to the blood sample and data collection.

REFERENCES

1. Ostrom QT, Gittleman H, Liao P, et al. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro. Oncol.* 2017; 19(suppl_5):v1-v88.

2. Ostrom QT, Gittleman H, Xu J, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013. *Neuro. Oncol.* 2016; 18:v1–v75.

3. Amirian ES, Zhou R, Wrensch MR, et al. Approaching a Scientific Consensus on the Association between Allergies and Glioma Risk: A Report from the Glioma International Case-Control Study. *Cancer Epidemiol Biomarkers Prev.* 2016; 25(2):282-290.

4. Ostrom QT, Bauchet L, Davis F, et al. The epidemiology of glioma in adults: a "state of the science" review. *Neuro. Oncol.* 2014; 16(7):896-913.

5. Kinnersley B, Mitchell JS, Gousias K, et al. Quantifying the heritability of glioma using genome-wide complex trait analysis. *Scientific reports.* 2015; 5:17267.

6. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* 2017; 49(5):789-794.

7. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet.* 2007; 81(6):1278-1283.

8. Reinius B, Saetre P, Leonard JA, et al. An evolutionarily conserved sexual signature in the primate brain. *PLoS genetics.* 2008; 4(6):e1000100.

9. Rinn JL, Snyder M. Sexual dimorphism in mammalian gene expression. *Trends in genetics : TIG.* 2005; 21(5):298-305.

10. Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. *Nature reviews. Genetics.* 2007; 8(9):689-698.

11. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet.* 2015; 16(4):197-212.

12. Ostrom QT, Kinnersley B, Wrensch MR, et al. Sex-specific glioma genome-wide association study identifies new risk locus at 3p21.31 in females, and finds sex-differences in risk at 8q24.21. *Scientific reports.* 2018; 8(1):7352.

13. Warrington NM, Sun T, Luo J, et al. The cyclic AMP pathway is a sex-specific modifier of glioma risk in type I neurofibromatosis patients. *Cancer Res.* 2015; 75(1):16-21.

14. Amirian ES, Armstrong GN, Zhou R, et al. The Glioma International Case-Control Study: A Report From the Genetic Epidemiology of Glioma International Consortium. *Am. J. Epidemiol.* 2016; 183(2):85-91.

15. Melin BS, Barnholtz-Sloan JS, Wrensch MR, et al. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nat. Genet.* 2017.

16. Wrensch M, Jenkins RB, Chang JS, et al. Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat. Genet.* 2009; 41(8):905-908.

17. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat. Genet.* 2009; 41(8):899-904.

18. Rajaraman P, Melin BS, Wang Z, et al. Genome-wide association study of glioma and meta-analysis. *Human Genetics* 2012; 131(12):1877-1888.

19. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature genetics.* 2010; 42(5):436-440.

20. Ostrom QT, Kinnersley B, Wrensch MR, et al. Sex-specific genome-wide association study in glioma identifies new risk locus at 3p21.31 in females, and finds sex-differences in risk at 8q24.21. *bioRxiv* 2017; https://www.biorxiv.org/content/biorxiv/early/2017/12/18/229112.full.pdf.

21. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics.* 2007; 39(7):906-913.

22. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* 2015; 526(7571):68-74.

23. Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS computational biology.* 2016; 12(1):e1004714.

24. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS genetics.* 2007; 3(7):e114.

25. Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet.* 2011; 88(3):283-293.

26. Liu JZ, McRae AF, Nyholt DR, et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet.* 2010; 87(1):139-145.

27. Chanda P, Huang H, Arking DE, Bader JS. Fast association tests for genes with FAST. *PLoS One.* 2013; 8(7):e68585.

28. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28(1):27-30.

29. D'Eustachio P. Reactome knowledgebase of human biological pathways and processes. *Methods in molecular biology.* 2011; 694:49-61.

30. Nishimura D. BioCarta. *Biotech Software & Internet Report.* 2001; 2(3):117-120.

31. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011; 27(12):1739-1740.

32. Subramanian A, Tamayo P, Mootha VK, et al. 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(43):15545-15550.

33. R Core Team. R: A language and environment for statistical computing. 2017; http://www.R-project.org/.

34. Wickham H. ggplot2: elegant graphics for data analysis. 2009; http://had.co.nz/ggplot2/book.

35. Csardi G NT. The igraph software package for complex network research. *InterJournal.* 2006; Complex Systems:1695.

36. Briatte F. ggnetwork: Geometries to Plot Networks with 'ggplot2'. R package version 0.5.1. 2016; https://CRAN.R-project.org/package=ggnetwork.

37. Sales G, Calura E, Romualdi C. graphite: GRAPH Interaction from pathway Topological Environment. R package version 1.16.0. 2015; http://www.bioconductor.org/packages/release/bioc/html/graphite.html.

38. Auguie B. gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. 2017; https://CRAN.R-project.org/package=gridExtra.

39. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010; 26(18):2336-2337.

40. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nature genetics.* 2013; 45(4):385-391, 391e381-382.

41. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nature genetics.* 2012; 44(7):770-776.

42. Closa A, Cordero D, Sanz-Pamplona R, et al. Identification of candidate susceptibility genes for colorectal cancer through eQTL analysis. *Carcinogenesis.* 2014; 35(9):2039-2046.

43. Darras BT, Miller DT, Urion DK. Dystrophinopathies. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews(R)*. Seattle (WA)1993.

44. Pantaleo MA, Astolfi A, Urbini M, et al. Dystrophin deregulation is associated with tumor progression in KIT/PDGFRA mutant gastrointestinal stromal tumors. *Clinical sarcoma research.* 2014; 4:9.

45. Wang Y, Marino-Enriquez A, Bennett RR, et al. Dystrophin is a tumor suppressor in human cancers with myogenic programs. *Nature genetics.* 2014; 46(6):601-606.

46. Hatzfeld M. The armadillo family of structural proteins. *International review of cytology.* 1999; 186:179-224.

47. Kurochkin IV, Yonemitsu N, Funahashi SI, Nomura H. ALEX1, a novel human armadillo repeat protein that is expressed differentially in normal tissues and carcinomas. *Biochemical and biophysical research communications.* 2001; 280(1):340-347.

48. Codd V, Nelson CP, Albrecht E, et al. Identification of seven loci affecting mean telomere length and their association with disease. *Nature genetics.* 2013; 45(4):422-427, 427e421-422.

49. Walsh KM, Codd V, Smirnov IV, et al. Variants near TERT and TERC influencing telomere length are associated with high-grade glioma risk. *Nature genetics.* 2014; 46(7):731-735.

50. Walsh KM, Wiencke JK, Lachance DH, et al. Telomere maintenance and the etiology of adult glioma. *Neuro. Oncol.* 2015; 17(11):1445-1452.

51. Mayer S, Bruderlein S, Perner S, et al. Sex-specific telomere length profiles and age-dependent erosion dynamics of individual chromosome arms in humans. *Cytogenet. Genome Res.* 2006; 112(3-4):194-201.

52. Barrett JH, Iles MM, Dunning AM, Pooley KA. Telomere length and common disease: study design and analytical challenges. *Hum Genet.* 2015; 134(7):679-689.

53. Walsh KM, Codd V, Rice T, et al. Longer genotypically-estimated leukocyte telomere length is associated with increased adult glioma risk. *Oncotarget.* 2015; 6(40):42468-42477.

54. Scarbrough PM, Akushevich I, Wrensch M, Il'yasova D. Exploring the association between melanoma and glioma risks. *Annals of epidemiology.* 2014; 24(6):469-474.

55. Scheurer ME, Etzel CJ, Liu M, et al. Familial aggregation of glioma: a pooled analysis. *Am. J. Epidemiol.* 2010; 172(10):1099-1107.

56. Paunu N, Pukkala E, Laippala P, et al. Cancer incidence in families with multiple glioma patients. *International journal of cancer.* 2002; 97(6):819-822.

57. Ransohoff KJ, Wu W, Cho HG, et al. Two-stage genome-wide association study identifies a novel susceptibility locus associated with melanoma. *Oncotarget.* 2017; 8(11):17586-17592.

58. Kocarnik JM, Park SL, Han J, et al. Replication of associations between GWAS SNPs and melanoma risk in the Population Architecture Using Genomics and Epidemiology (PAGE) Study. *The Journal of investigative dermatology.* 2014; 134(7):2049-2052.

59. Endicott AA, Taylor JW, Walsh KM. Telomere length connects melanoma and glioma predispositions. *Aging.* 2016; 8(3):423-424.

60. Mirina A, Atzmon G, Ye K, Bergman A. Gene size matters. *PLoS One.* 2012; 7(11):e49093.

FIGURE KEY

Figure 1. Study schematic for a) generation of discovery and validation summary statistic sets, b) generation, prioritization, and validation of gene-scores , c) generation, prioritization, and validation of pathway-scores.

Figure 2. Gene-scores for prioritized genes by algorithm, histology, and sex for a) BPESC1 (3q23), B) TERT (5p15.33), C) EGFR (7p11.2), D) CDKN2B (9p21.3), E) DNAH2 (17p13.1), F) RTEL1-TNFRSF6B (20q13.33).

Figure 3. Conditional gene-scores for prioritized genes by algorithm, histology, and sex for A) TERT (5p15.33), B) EGFR (7p11.2), C) CDKN2B (9p21.3), D) DNAH2 (17p13.1), E) RTEL1-TNFRSF6B (20q13.33).

Figure 4. Gene-scores for genes in the Biocarta telomere pathway for all glioma in a) males, and b) females, and for glioblastoma in c) males and d) females.