Systems Biomedicine

ISSN: 2162-8130 (Print) 2162-8149 (Online) Journal homepage: http://www.tandfonline.com/loi/kssb20

Systematic use of computational methods allows stratification of treatment responders in glioblastoma multiforme

Riku Louhimo, Viljami Aittomäki, Ali Faisal, Marko Laakso, Ping Chen, Kristian Ovaska, Erkka Valo, Leo Lahti, Vladimir Rogojin, Samuel Kaski & Sampsa Hautaniemi

To cite this article: Riku Louhimo, Viljami Aittomäki, Ali Faisal, Marko Laakso, Ping Chen, Kristian Ovaska, Erkka Valo, Leo Lahti, Vladimir Rogojin, Samuel Kaski & Sampsa Hautaniemi (2013) Systematic use of computational methods allows stratification of treatment responders in glioblastoma multiforme, Systems Biomedicine, 1:2, 130-136, DOI: 10.4161/sysb.28904

To link to this article: https://doi.org/10.4161/sysb.28904

Published online: 14 Apr 2014.

RESEARCH PAPER

Systems Biomedicine 1:2, 130–136; April/May/June 2013; © 2013 Landes Bioscience

Systematic use of computational methods allows stratification of treatment responders in glioblastoma multiforme

Riku Louhimo,[1] Viljami Aittomäki,[1,†] Ali Faisal,[2,†] Marko Laakso,[1,†] Ping Chen,[1] Kristian Ovaska,[1] Erkka Valo,[1] Leo Lahti,[3,‡] Vladimir Rogojin,[1] Samuel Kaski,[2,4] and Sampsa Hautaniemi[1,*]

[1]Systems Biology Laboratory; Genome-scale Biology Research Program; University of Helsinki; Helsinki, Finland; [2]Helsinki Institute of Information Technology (HIIT); Department of Information and Computer Science; Aalto University; Espoo, Finland; [3]Department of Veterinary Bioscience; University of Helsinki; Helsinki, Finland; [4]Department of Computer Science; University of Helsinki; Helsinki, Finland

[†]These authors contributed equally to this work.

[‡]Current affiliation: Department of Veterinary Biosciences; University of Helsinki; Helsinki, Finland

Keywords: data integration, data mining, dependency analysis

Background: Cancers are complex diseases whose comprehensive characterization requires genome-scale molecular data at multiple levels from genetics to transcriptomics and clinical data. Using our recently published Anduril bioinformatics framework and novel computational approaches, such as dependency analysis, we identify key variables at miRNA, copy number variation, expression, methylation, and pathway levels in glioblastoma multiforme (GBM) progression and drug resistance. Furthermore, we identify characteristics of clinically relevant subgroups, such as patients treated with temozolomide and patients with an EGFRvIII mutation, which is a constitutively active variant of EGFR.

Results: We identify several novel genomic regions and transcript profiles that may contribute to GBM progression and drug resistance. All results and Anduril scripts are available at http://csbi.ltdk.helsinki.fi/camda/.

Conclusions: Our results highlight the need for approaches that define context at several levels in order to identify genomic regions or transcript profiles playing key roles in cancer progression and drug resistance.

Background

Glioblastoma multiforme (GBM) is the most frequent and aggressive brain tumor type with an incidence of 2–3 cases in 100,000 people per year. Over the past 25 years the advances in GBM treatment have been very modest and the median survival of a GBM patient has remained at 15 months.[1] To improve diagnosis and treatment of GBM, The Cancer Genome Atlas (TCGA) consortium provides, for hundreds of GBM primary tumors, high-throughput molecular data at genetic, transcriptomic, and epigenetic levels together with clinical data.[2] These data sets thus provide a basis for gaining a holistic view on GBM progression and drug resistance. Integration of such massive amounts of data requires a computational infrastructure, however, that allows systematic data processing and interpretation. We recently introduced a computational platform, Anduril, which facilitates the analysis and integration of large-scale data, systematic software development and the rapid use of different bio-databases.[3] Anduril provides a framework for joining reusable algorithms (components) into executable workflows. A component can be implemented with any programming language, including R, MATLAB, Java, and C++, which allows us to take advantage of the whole range of efforts of the bioinformatics community. The system also allows us to parallelize computationally demanding tasks.

A major challenge in studies characterizing complex diseases is that despite seemingly similar phenotypes, such as finding a malignant tumor in brain, there can be considerable heterogeneity in histopathological and molecular levels between samples in a cohort. Therefore, it is crucial to identify medically relevant subgroups of patients and key variables, such as genomic regions, genes, or pathways that possibly stratify a cohort. Such groups or key variables can be identified by integrating molecular data with clinical and treatment data. For example, for GBM, Verhaak and colleagues suggested four subtypes of GBM based on gene expression profiles.[4] When identifying such subgroups based on molecular level information, it is also important to consider how patients have been treated. For example, the GBM patients in

*Correspondence to: Sampsa Hautaniemi; Email: sampsa.hautaniemi@helsinki.fi
Submitted: 09/12/2011; Revised: 12/05/2011; Accepted: 07/12/2013; Published Online: 04/14/2014
http://dx.doi.org/10.4161/sysb.28904

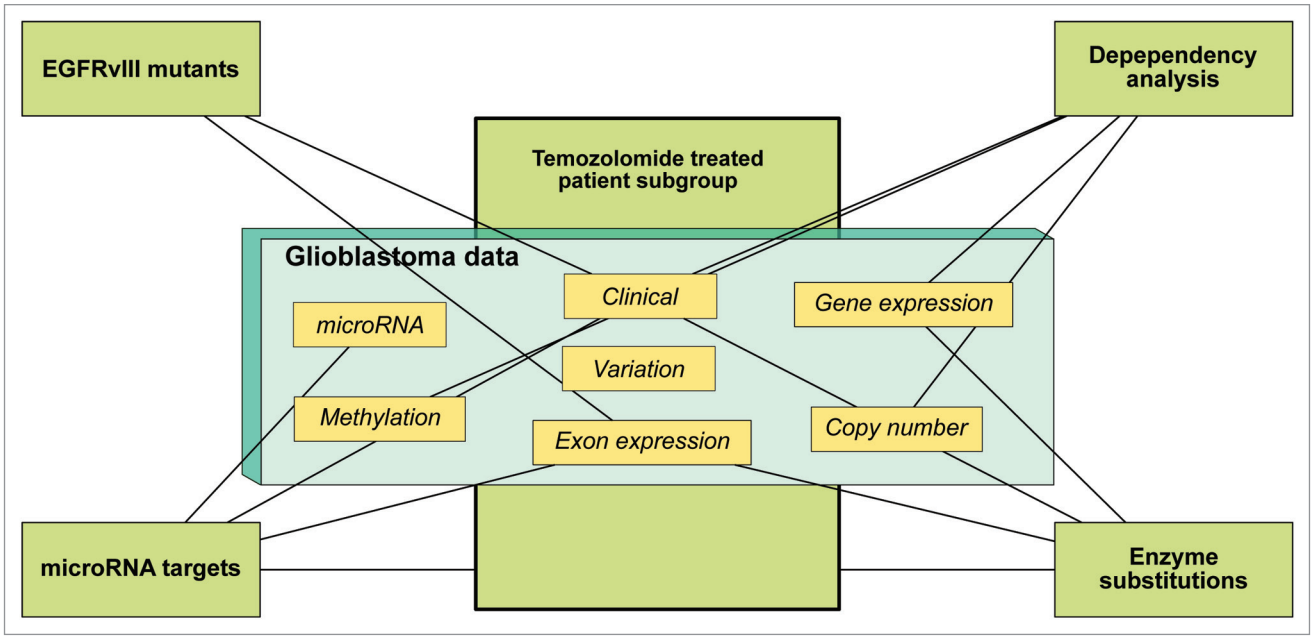

Figure 1. A schematical overview of the analysis pipeline. Connections between subanalyses (green boxes) and glioblastoma data (yellow) illustrate the primary data sources for each analysis. Connections between subanalyses illustrate connections in their results.

TCGA have been treated using a wide spectrum of drugs, and while some patients have received only single drug chemotherapy, others have been administered up to 15 different compounds in combination therapy.

In this study, our main objective is to present computational approaches that allow the identification of genomic regions and transcript profiles that have a significant association with survival or a positive drug response for subsets of the GBM samples. We identify several genes in samples treated with temozolomide, which is one of the most potent GBM drugs to date, and key chromosomal regions that show survival association. In addition, we introduce a method to identify samples with the EGFRvIII mutant form. We also characterize putative alternative pathways in GBM that have been activated by enzyme substitutions, and integrate microRNA and methylation with gene expression data to characterize possible causes for the enzyme deactivation. Our results identify novel regions associated with GBM progression and drug resistance, and highlight the need for comprehensive integrated data analysis of complex biological disorders.

Results

We developed the Anduril workflow that automatically downloads, annotates, and analyzes gene expression, copy number, single nucleotide polymorphism (SNP), microRNA (miRNA), methylation, and clinical data from TCGA. This structured workflow enables us to develop integrative approaches to analyze these data (**Fig. 1**), and greatly facilitates data and analysis management. Details and parameters of data processing for each individual data type are available at http://csbi.ltdk.helsinki.fi/camda/.

Automated TCGA data import

The first step in the analysis of TCGA data are to retrieve it. As the TCGA repository is frequently updated, fetching the data needs to be automatic and periodical. To this end we implemented an Anduril component (GetFromTcga) that automatically imports data from the TCGA data portal. These data can then be integrated into Anduril workflows. GetFromTcga automatically generates file and sample reference tables as well as reports relationships between the data files and the samples that those files contain. The latest versions of data are imported by default, although a user can request any earlier versions of the data. In this way, GetFromTcga enables fine grained selection of data to be imported, as well as an automated data download and update functionality. Since GetFromTcga is incorporated into our workflow, our analysis always contains the most up-to-date version of the TCGA data. In the present work, all data were accessed on May 6, 2011.

Biomarker candidate search for temozolomide treated GBM patients

GBM patients with a methylated promoter of the MGMT DNA-repair gene treated with temozolomide adjuvant therapy and radiotherapy have a longer median survival of 15 to 21 months.[5] Though the treatment is not curative and mutations in mismatch repair genes have been shown to override MGMT repression by rendering tumors resistant to alkylating agents,[6] temozolomide in GBM is a prime example of the power of "personalized medicine," i.e., choosing the therapeutic strategy using a molecular biomarker status of the patient. Here our objective was to establish a workflow that allows for a rapid search for candidate biomarkers for a given treatment strategy. We used 76 GBM patients treated with adjuvant temozolomide and checked whether gene or alternative spliced transcript variant

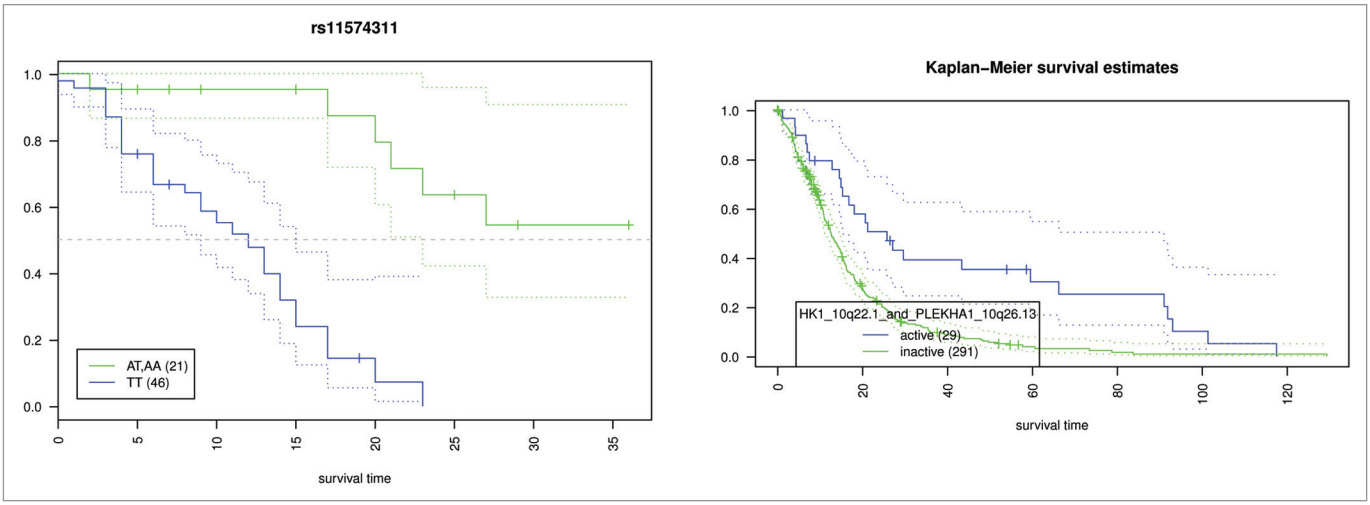

Figure 2. Kaplan-Meier survival plots a) Kaplan-Meier analysis of rs11574311 in *WRN* for 67 temozolomide treated patients that were also genotyped. Heterozygous or rare homozygote patients (AA/AT) have significantly better survival than patients with the wildtype homozygote (TT). *WRN* has been suggested to be involved in DNA damage repair, which makes it an interesting candidate for further studies. b) Kaplan-Meier analysis of a combination of two regions centered at the genes *HK1* and *PLEKHA1*. Patients with a high dependency between copy number alteration and gene expression in all these regions ("active") have better survival association than patients having low dependency ("inactive"). *x* axis: months; *y* axis: percentage of GBM patients alive. Dotted lines: 95% confidence intervals.

expression, copy number alterations, SNPs, methylation patterns, or miRNA have survival effects using Kaplan-Meier analysis (logrank test, $P < 0.05$).

This analysis suggested several interesting genes for further investigation. For instance, chromosome X open reading frame 1 (*CXorf1*) is significantly less expressed in GBM and associated with survival in both gene and exon platform data ($P < 0.0004$). As another example, Werner syndrome, RecQ helicase-like (*WRN*) has two SNPs that are significantly associated with survival ($P < 0.0001$) as shown in **Figure 2**. Both of these effects are absent if all patients are included in the analysis. While the number of temozolomide-treated tumors is small, our efforts provide a comprehensive workflow to identify variables that play a key role in temozolomide sensitivity, and provide candidate biomarkers for temozolomide treatment.

Dependency analysis of differentially expressed genes within copy number alteration and methylated regions

Genomic instability is a hallmark of cancer and high-throughput measurements of copy number aberration data have become commonplace in cancers. Given that copy number measurements are noisy, one of the most successful approaches in increasing the reliability of putative driver genes involved in tumor progression and drug resistance is integration of copy number data to transcriptomics data. Our objective is to first identify chromosomal regions that have high dependencies between gene expression and copy number changes, and then form patient groups from each of the identified region and a survival analysis to check whether the identified genomic aberrations have survival associations in GBM.

We used our recently developed similarity constrained canonical correlation analysis approach (simCCA) as we have observed that with suitable constraints/prior it performs better than other learning methods.[7] The method is based on a Bayesian formulation of classical canonical correlation analysis.[8] It detects linear dependencies between two data sources by searching for their maximally correlated low-dimensional representation. Briefly, the model defines a chromosomal region via a window that is centered at a gene and spans across ten neighboring genes within the chromosomal arm. The window is slid across all chromosomal arms and a dependency score and each sample's contribution toward the score for each region is calculated. A high score reveals a correlating expression and corresponding chromosomal change; high-scoring regions with $q < 0.05$ were selected for further analysis. The significance of the regions was estimated by a permutation test, using the observed dependency score as a test statistic. The columns (samples) in the gene expression data matrix were randomly rearranged, which distorts the relationship of genes and the copy-number changes. One thousand random permutations were formed and their dependency scores computed. Significance of each chromosomal region was then determined as the proportion of random scores that are greater than the observed dependency score. For each identified region, patient-wise contribution scores were ordered and three groups were formed based on the 10th percentile, the 90th percentile and the rest.

For each identified region we checked the corresponding patient groups for significant survival associations ($P < 0.05$) using both Kaplan-Meier and Cox models. The analysis was also repeated for methylation and gene-expression data-pair, taking into account their inverse relationship (this is achieved by placing a negative identity prior on the transformation matrix; for details refer to ref. 7). The found survival associated regions were filtered by controlling the false positives using two different approaches; 1) a multiple hypothesis correction was applied where we restricted the overall false discovery rate to 5%, 2) an integrative analysis was performed where we searched for common significant ($P <$

Table 1. Survival associated chromosomal regions identified using Kaplan-Meier and Cox analysis

				Enricment Test *q* values								
				Clinical factors							Temozolomide	
	Data set used	Region	Genes in the region	White	Female	Male	Age < 30	Age < 40	Age > 50	Age > 60	Upreg	Downreg
Kaplan Meier	cgh + exp	10p13	**MCM10** SEC61A2 OPTN CDC123 OLAH RPP38 PRPF18 PTER CAMK1D HSPA14	1	0.186	1	0.254	0.006	1	1	1	1
		10q22.1	UNC5B CHST3 SUPV3L1 HKDC1 **HK1** DDX21 SGPL1 COL13A1 SLC29A3 KIAA1279	0.254	0.378	1	1	1	1	9.26078E-05	1	1
		10q22.1	HNRNPH3 UNC5B SUPV3L1 **HKDC1** HK1 DDX21 SGPL1 COL13A1 SLC29A3 KIAA1279	1	1	1	1	0.056	1	1	1	1
		10q26.13	SEC23IP **PLEKHA1** BCCIP WDR11 PTPRE TACC2 BUB3 FAM175B ACADSB INPP5F	1	1	1	1	1	0.056	0.812	1	1
	methyl + exp	21q22.2	DOPEY2 ETS2 CBR1 MORC3 SLC37A1 PKNOX1 CRYAA TTC3 MX2 SH3BGR	1	1	1	1	0.873	1	1	0.804	1
Cox	cgh + methyl + exp	9q24.3	**SMARCA2** DNAJA1 TYRP1 SH3GL2 TEK IFNA8 SNAPC3 NUDT2 KCNV2 PDCD1LG2	1	1	0.378	1	1	1	1	1	0.885

The region is centered at the gene shown in bold.

0.05) regions found in both copy-number/gene-expression and methylation/gene-expression analysis. The former approach is a strict test for regions at the cost of missing some of the true positives from two data-sources; copy-number and gene-expression, while the latter provides shared active regions common in the three data-sources; copy-number, gene-expression and methylation.

The analysis of copy-number/gene-expression data sets identified three significant chromosomal regions (10p13, 10q22.1, 10q26.13) using Kaplan-Meier ($q < 0.05$). Many genes in these regions had expression profiles that correlated with copy number aberrations, such as *HK1, HKDC1, MCM10, DDX21*, and *SLC29A3*. None of the genes identified in these regions could be associated to temozolomide treatment (**Table 1**). In addition to copy number alterations, gene expression levels are affected by methylation patterns. The results from methylation/gene-expression reveal one statistically significant Kaplan-Meier survival associated region ($q < 0.05$) at 21q22.2. This region is centered at *ETS2* which has both tumor suppressive and promoting properties depending in different tumor types.[9,10]

There are several clinical factors (age, race, and gender) for each patient that may bias the groupings based on copy number and gene expression data. Thus, we used Fisher exact test to check for the bias as follows. The first grouping is induced by the quantile clustering on the patient-wise contribution scores while the second grouping is formed from any of the binary clinical variables considered separately. The clinical factors were transformed to binary variables: race: white/non-white, gender: male/female, and age was discretized into four classes: < 30, < 40, > 50, and > 60 years.

The resulting q values indicate no bias from the tested clinical factors (**Table 1**).

Cox survival analysis can be regarded as an extension of Kaplan-Meier analysis that controls for the effect of covariate variables. The integrative analysis of copy-number/gene-expression and methylation/gene-expression revealed one shared significant region ($P < 0.05$) 9p24.3 based on Cox analysis. Kaplan-Meier analysis did not find regions shared between these two data-pairs.

Identification of microRNAs and their target genes with survival association

MicroRNAs (miRNAs) are short non-coding RNAs that typically negatively regulate gene expression and alterations in miRNA expressions are frequently associated with human cancers.[11] In GBM miRNAs play a key role in many hallmarks of glioblastoma, including cell proliferation, invasion, glioma stem cell behavior, and angiogenesis.[12]

Here, we integrated gene expression data from exon arrays of 437 GBM patients to miRNA expression data (level 3) from 309 patients to investigate the relationship between cancer related regulatory miRNAs and their potential targets in GBM. Exon arrays were preprocessed by Multiple Exon Array Preprocessing (MEAP) algorithm,[13] which reduces bias due to probe sequence composition with a Bayesian estimation model for probe background correction and quantifies reliable expression data at exon, alternative splicing and gene levels. In this study, we summarized gene expression data from probes mapped to the majority of gene's alternative splicing variants (60%). Initial targets of 63 differentially expressed miRNAs in GBM were collected from miRBase,[14] microRNAorg,[15] miRNAMap2,[16] and TargetScan,[17] which contain both validated and predicted targets. These targets were further filtered by Pearson correlation ($r < -0.4$) and Kaplan-Meier survival analysis ($P < 0.05$).

Our miRNA and gene expression integration analysis resulted in 10 candidate miRNAs along with their 56 negatively regulated targets with survival association ($P < 0.05$). Interestingly, the region 10p13, which was identified with copy number and expression data integration, harbors *SEC61A2*, that has high negative correlations with *hsa-mir-23a* ($r < -0.5$) and *hsa-mir-34a* ($r = -0.478$), and *CAMK1D*, that has high negative correlation with *hsa-mir-25* ($r = -0.424$).

Identification of EGFRvIII patients

Epidermal growth factor receptor (*EGFR*) is the most frequently amplified and overexpressed gene in GBM and its amplification is a prognostic marker. EGFRvIII is a variant of *EGFR* with genomic deletion of exons 2–7, corresponding to amino acids 6–273 in the extracellular domain.[18] This variant is not capable of binding ligands but is constitutively active and contributes to tumor progression. The EGFRvIII mutation is found in 20–30% of glioblastoma patients and it is often combined with *EGFR* amplification.[18]

We developed a method to systematically find EGFRvIII mutations in TCGA data based on Affymetrix exon expression arrays. Exon arrays have the advantage of having high probe resolution in exons. They also enable simultaneous profiling of expression and targeted deletions. Probe sets targeting *EGFR* are divided into two groups based on whether they match exons deleted in EGFRvIII mutations. Probe sets with low median signal are considered outliers and are removed from statistical analysis. After filtering, there are six probe sets in the deletion group and 32 probe sets in the non-deletion group. In EGFRvIII-mutant samples, deletion-group probe sets should have lower signal than other probe sets. For each tumor sample, a one tailed t test is applied between these two groups. Samples with a low P value are considered likely candidates for an EGFRvIII mutation.

Using a P value threshold of 0.1, 103 out of 397 samples (26%) were classified as potentially having EGFRvIII mutations. In accordance with varied literature on the survival effect of EGFRvIII,[18] we did not see a survival difference between candidate EGFRvIII patients and non-EGFRvIII patients ($P < 0.41$).

To find potential drug correlations with EGFRvIII phenotype, we selected the subset of patients (n = 196) who have received temozolomide treatment. This drug was selected due to the large number of patients receiving the drug; similar analysis can be conducted for any drug having a suitable number of receiving patients. Temozolomide-treated patients have better survival compared with other patients both in the whole patient set ($P < 1.4 \times 10^{-5}$) as well as in EGFRvIII patients ($P < 0.051$). We compared the survival of EGFRvIII vs. non-EGFRvIII patients in the temozolomide-treated patient group. No statistical survival association was found ($P < 0.17$), suggesting that the EGFRvIII mutation does not affect survival in temozolomide-treated patients.

From context to reactions: Enzyme substitutions in GBM

Enzyme activities are often aberrated in cancer cells and they play a key role in drug resistance. We focused in alternative pathways by linking the identified genomic regions and the transcript profiles with survival effect in GBM to enzymatic reactions. We searched for pairs of enzymes with the same enzyme commission (EC) number, i.e., catalyzing the same reaction, where the same sample has the first enzyme upregulated and the second one downregulated. Our hypothesis is that these enzymes may represent activations of alternative pathways in tumors that mediate tumor progression and drug resistance. Biological properties, such as regulation, binding, efficacy and specificity, typically differ between the original and the substituting enzyme although they share the same EC number. Such differences in vital metabolic processes may be used for therapeutic purposes.[19]

Enzyme substitutions found among the differentially expressed genes (DEG)[3] were followed up at the sample level. The gene expression profile of each sample was compared against the median signals of the control population and a fold change limit of two was used to call downregulation and upregulation. The intersection of the common substitution pairs and the DEGs of each individual sample was used to seed de novo pathway construction. Pathways were constructed using the Moksiskaan database, which describes known interactions between the genes, proteins, and drugs.[20] The de novo pathways were constructed by selecting the genes, drugs, molecular functions, and biological processes connected to the enzyme substitution pairs found from the expression sample profile. Small molecule mediated

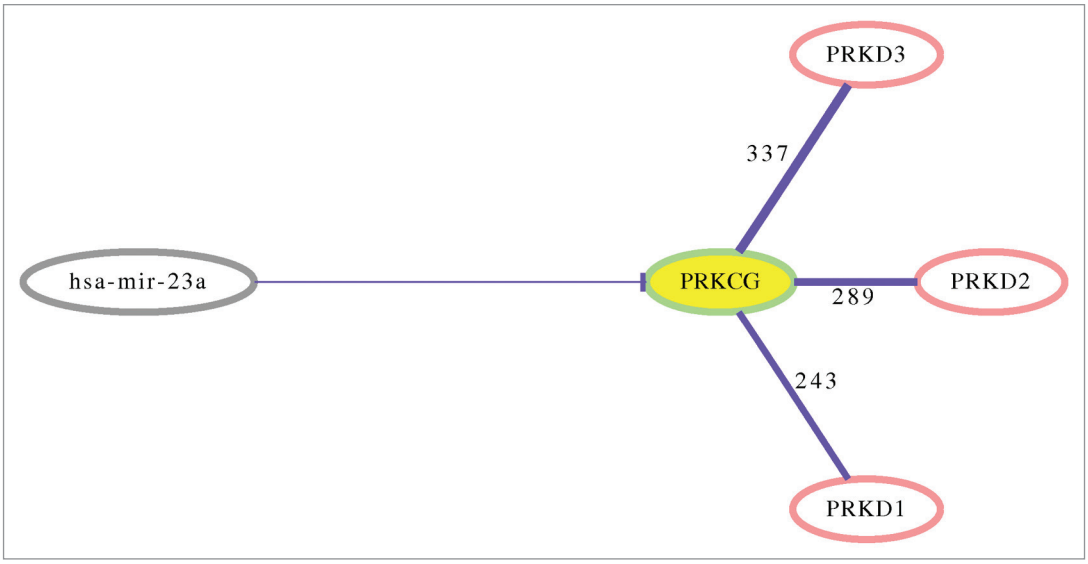

Figure 3. An example of enzyme-pair analysis. Upregulated and downregulated genes are in red and green, respectively. The sample numbers for an enzyme pair are displayed on edges. Nodes filled with orange or yellow are validated or predicted miRNA targets in databases. MiRNA gene *hsa-mir-23a* has a survival association and *PRKCG* belongs to the KEGG glioma pathway. Here we found three upregulated enzymes from the same protein kinase C group (EC 2.7.11.13).

dependencies were excluded as our primary interest in this study is in signaling cascades.

The results reveal a number of interesting substitutions where known GBM related genes have a central role, such as 295 samples with a phosphorylase, glycogen, muscle (*PYGM*) → phosphorylase, glycogen, liver (*PYGL*) transition affecting the glycogen metabolism. We speculate that these changes are related to the active glycolysis that provides energy to the tumor cells and possibly aids their survival in hypoxia. Total of 129 samples shared a uridine phosphorylase 2 (*UPP2*) → uridine phosphorylase 1 (*UPP1*) transitions possibly related to the maintenance of the active DNA replication. These enzymes are important mediators of fluorouracil that is a widely used anticancer drug.[21]

In addition to a global search, we also integrated the identified miRNA-gene pairs (Section 2.4) to the substituted enzymes that were also targets of the survival associated miRNAs. We found several downregulated enzymes in common enzyme substitutions that are likely to be regulated by miRNAs. For example, *hsa-mir-23a* has a survival effect and its target *PRKCG* participates a reaction that is associated with GBM progression as shown in **Figure 3**.

Conclusions

Multi-dimensional data provided by TCGA require a computational platform, such as Anduril, that allows for inclusion of a large spectrum of computational methods and facilitates collaboration of a team of bioinformaticians. We have presented here several tools to identify key variables defining context for GBM progression and drug resistance, such as probabilistic dependency analysis that provides a flexible and robust approach toward multi-view data integration in functional genomics.

Our results highlight the need for a large number of samples; when the data set is divided into clinically or therapeutically interesting subgroups, the number of samples decreases rapidly, which poses challenges for genome-scale analysis. For example, we introduced a novel approach to identify EGFRvIII variants using exon array data. However, no EGFRvIII patients were treated with EGFR kinase inhibitors, which prohibited characterization of possible biomarkers related to poor efficacy of the EGFR inhibitors in GBM. Accordingly, we showed a proof-of-principle analysis with temozolomide treated EGFRvIII patients, which resulted in no association between constitutive EGFR signaling to temozolomide efficacy.

Our results also highlight the need for advanced algorithms to define context at several levels in order to identify genomic regions or transcript profiles that play a key role in cancer progression and drug resistance. Here we have shown two novel approaches to identify context at miRNA, gene expression, methylation, and copy number alteration levels and association to survival of patients treated with temozolomide. We further linked the miRNA results to cell network level using a novel concept of finding enzyme substitution pairs. For instance, our results suggest that *hsa-mir-23a* has a survival association ($P < 0.029$) and its expression correlates ($r = -0.55$) with the *PRKCG* gene (**Fig. 2**). Enzyme substitutes are feasible drug targets as their activity is higher in cancer cells than healthy cells. Cancer cells may have become dependent on the substitute if they have lost the wild type enzyme or are otherwise unable to use it. The inhibition of these gene products may reduce the malignancy of the cells if they have become addicted to reactions mediated by these genes.

In summary, we have presented multiple computational approaches to identify clinically relevant subgroups using large-scale and heterogeneous data. All methods are implemented

on a common computational infrastructure Anduril, which allows reusing the methods in other large-scale studies as well as integrating the analysis results. Our results provide several genes and genomic regions that have survival effect in GBM or a clinically defined subset, such as temozolomide-treated patients.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

RL developed the analysis workflow, performed analysis and wrote manuscript. VA performed temozolomide analysis. AF and LL performed dependency analysis and wrote corresponding sections. ML analyzed the enzyme substitution patterns and wrote corresponding sections. PC analyzed miRNA/gene pairs and wrote corresponding sections. KO performed EGFRvIII analysis and wrote corresponding sections. VR developed the TCGA data import tool and wrote corresponding sections. EV contributed to the development of the analysis workflow. SK supervised dependency analysis, contributed to result interpretation and wrote dependency analysis related manuscript parts. SH initiated the project, contributed to results interpretation, wrote the manuscript, and supervised the project. All writers read and approved the final manuscript.

This work was supported by the Academy of Finland (projects: 125826); Sigrid Jusélius Foundation; Biocentrum Helsinki; Finnish Cancer Associations; Helsinki Biomedical Graduate School (RL, KO); and the Finnish Doctoral Programme in Computational Sciences (ML, PC, AF).

References

1. Huse JT, Holland EC. Targeting brain cancer: advances in the molecular pathology of malignant glioma and medulloblastoma. Nat Rev Cancer 2010; 10:319-31; PMID:20414201; http://dx.doi.org/10.1038/nrc2818
2. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008; 455:1061-8; PMID:18772890; http://dx.doi.org/10.1038/nature07385
3. Ovaska K, Laakso M, Haapa-Paananen S, Louhimo R, Chen P, Aittomäki V, Valo E, Núñez-Fontarnau J, Rantanen V, Karinen S, et al. Large-scale data integration framework provides a comprehensive view on glioblastoma multiforme. Genome Med 2010; 2:65; PMID:20822536; http://dx.doi.org/10.1186/gm186
4. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, et al.; Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010; 17:98-110; PMID:20129251; http://dx.doi.org/10.1016/j.ccr.2009.12.020
5. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005; 352:997-1003; PMID:15758010; http://dx.doi.org/10.1056/NEJMoa043331
6. Liu L, Markowitz S, Gerson SL. Mismatch repair mutations override alkyltransferase in conferring resistance to temozolomide but not to 1,3-bis(2-chloroethyl)nitrosourea. Cancer Res 1996; 56:5375-9; PMID:8968088
7. Lahti L, Myllykangas S, Knuutila S. S K: Dependency detection with similarity constraints. In Proc. MLSP 2009, IEHE International Workshop on Machine Learning for Signal Processing, IEEE 2009:89-94
8. Klami A, Kaski S. Local dependent components. In Proceedings of the 24th International Conference on Machine Learning (ICML 2007). Edited by Ghahramani Z, Omni Press 2007:425-433
9. Tynan JA, Wen F, Muller WJ, Oshima RG. Ets2-dependent microenvironmental support of mouse mammary tumors. Oncogene 2005; 24:6870-6; PMID:16007139; http://dx.doi.org/10.1038/sj.onc.1208856
10. Sussan TE, Yang A, Li F, Ostrowski MC, Reeves RH. Trisomy represses Apc$^{(Min)}$-mediated tumours in mouse models of Down's syndrome. Nature 2008; 451:73-5; PMID:18172498; http://dx.doi.org/10.1038/nature06446
11. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006; 6:857-66; PMID:17060945; http://dx.doi.org/10.1038/nrc1997
12. Lawler S, Chiocca EA. Emerging functions of microRNAs in glioblastoma. J Neurooncol 2009; 92:297-306; PMID:19357957; http://dx.doi.org/10.1007/s11060-009-9843-2
13. Chen P, Lepikhova T, Hu Y, Monni O, Hautaniemi S. Comprehensive exon array data processing method for quantitative analysis of alternative spliced variants. Nucleic Acids Res 2011; 39:e123; PMID:21745820; http://dx.doi.org/10.1093/nar/gkr513
14. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008; 36(suppl 1):D154-8; PMID:17991681; http://dx.doi.org/10.1093/nar/gkm952
15. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. Nucleic Acids Res 2008; 36(suppl 1):D149-53; PMID:18158296; http://dx.doi.org/10.1093/nar/gkm995
16. Hsu SD, Chu CH, Tsou AP, Chen SJ, Chen HC, Hsu PWC, Wong YH, Chen YH, Chen GH, Huang HD. miRNAMap 2.0: genomic maps of microRNAs in metazoan genomes. Nucleic Acids Res 2008; 36(suppl1):D165-9; PMID:18029362; http://dx.doi.org/10.1093/nar/gkm1012
17. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. Cell 2003; 115:787-98; PMID:14697198; http://dx.doi.org/10.1016/S0092-8674(03)01018-3
18. Gan HK, Kaye AH, Luwor RB. The EGFRvIII variant in glioblastoma multiforme. J Clin Neurosci 2009; 16:748-54; PMID:19324552; http://dx.doi.org/10.1016/j.jocn.2008.12.005
19. Vander Heiden MG. Targeting cancer metabolism: a therapeutic window opens. Nat Rev Drug Discov 2011; 10:671-84; PMID:21878982; http://dx.doi.org/10.1038/nrd3504
20. Laakso M, Hautaniemi S. Integrative platform to translate gene sets to networks. Bioinformatics 2010; 26:1802-3; PMID:20507894; http://dx.doi.org/10.1093/bioinformatics/btq277
21. Longley DB, Harkin DP, Johnston PG. 5-fluorouracil: mechanisms of action and clinical strategies. Nat Rev Cancer 2003; 3:330-8; PMID:12724731; http://dx.doi.org/10.1038/nrc1074