Uniwersytet Warszawski
Wydział Biologii

Marta Koblowska, PhD
Head of the Department of Systems Biology
Faculty of Biology, University of Warsaw

Warsaw 10.02.2021

Review Report

on the doctoral thesis of ADRIA-JAUME ROURA CANALDA
entitled: **„A multi-omics evaluation of somatic mutations, transcriptomic dysregulation, chromatin accessibility and remodeling in High-Grade Gliomas"**

PhD thesis completed at the Laboratory of Molecular Neurobiology of the Nencki Institute of Experimental Biology Polish Academy of Sciences under supervision of Prof. dr hab. Bożena Kamińska-Kaczmarek and Dr hab. Bartosz Wojtaś

Proper gene regulation ensuring the expression of the right genes at the right time and place is a key for the right functioning of cells and organisms. The complexity and multi-level nature of this process is one of the main research topics in biology. Thanks to the involvement of thousands of scientists and enormous resources, we now know a lot about the rules governing gene expression in eukaryotic organisms, especially in humans. The emergence of high-throughput technologies, allowing for the simultaneous observation of many thousands of changes at both the genome and transcriptome levels, in combination with methods aimed at studying chromatin and the nuclear architecture, has allowed for the discovery of many new principles of gene regulation. Despite the enormous amount of results accumulated in recent years, we are aware that much remains to be discovered. The great challenge of the present time is not only asking the right questions and obtaining high quality results, but also to properly integrate the already collected data and to analyze them. This can lead scientists to many important breakthroughs. Striving for a better understanding of the already described regulatory phenomena, the relationships between them and discovering new ways of controlling genes is a great challenge, even more important because it is the only hope to find treatments for one of the most deadly human diseases - cancer. The basis of this disease are disturbances in the pattern of gene expression resulting from defects in the genome, epigenome and chromatin structure, leading to carcinogenic changes in the cellular phenotype. The detailed study of cancer cells and changes in regulatory elements that have caused gene perturbations, identification of key reasons leading to cancer formation, and discovery of cancer "vulnerabilities" will help us to win the fight for the lives of many patients worldwide.

Project background

Gliomas are primary brain tumors with extensive intra-tumoral heterogeneity, divided into four grades with increased malignancy according to the World Health Organization (WHO) classification. The most malignant form is glioblastoma multiforme GBM with very poor prognosis, irrespective of optimal, state-of-the-art treatment. The majority of patients survive only several months and usually die about 15 months after diagnosis.

ADRIA-JAUME ROURA CANALDA's doctoral dissertation is aimed at better understanding the underlying molecular basis for recurrence and pathobiology of High Grade Gliomas (HGGs).

For this study PhD Student used samples from 16 patients. These were fragments of primary and recurrent tumours, a total of 14 pairs of GBMs and 2 anaplastic astrocytomas WHO grade III. For two patients, additional samples were taken after 3rd recurring tumour removal. In total, 35 fresh frozen glioma samples were used in presented work. In addition some experiments were performed on human established GBM LN18 and LN229 cells.
To characterize the molecular basis of HGGs and identify alterations relevant to tumor recurrence, studies were performed using selected high-throughput technologies such as: targeted sequencing to determine somatic mutations and copy number alterations (CNAs) in primary and recurrent tumors, RNA-seq for transcriptome analysis, and the ATAC-seq method determining chromatin accessibility throughout the genome in this study used to identify HGGs-specific transcription factor binding sites and chromatin reorganization in GMB LN18 cells following downregulation of two major chromatin remodelers SMARCA2 and SMARCA4. In addition, gathered data from various analyses were integrated with external data previously obtained in the laboratory or collected as part of The Cancer Genome Atlas (TCGA) consortium. Selected results from high throughput experiments were confirmed by classical methods such as immuochemistry staining, EMSA supershift and Western blots.

Formal description of the dissertation

The dissertation presented for the review is 160 pages long and has a classical structure. It begins with acknowledgements, followed by a table of contents, figures and tables, a list of abbreviations and a two-page abstract in Polish and English. The abstract clearly presents the current state of knowledge, includes the main objectives of the work, a description of the used research methods and brief conclusions of the obtained results. A clearly written and comprehensive introduction with well-chosen figures and a chapter presenting detailed objectives of the study are preceded by subsequent parts of the work including description of materials and methods and the most extensive part of the work - the obtained results. The dissertation is summarized by a rather brief but conclusive discussion chapter, followed by a part containing informative summary and conclusions, a literature list and an appendix with additional tables with the results. Longer sections of the paper are divided into shorter subsections, which contain very well done and clear figures and tables. I particularly liked the idea of creating a graphical version of the research objectives that were undertaken. The dissertation is written comprehensively at an appropriate level of detail.

Merit of the dissertation

Introduction of the thesis

The introduction of the dissertation provides a very good overview of knowledge important for the scope of the work. It succinctly and clearly presents the most important issues that are crucial for the

thesis. It covers the characteristics and classification of gliomas, the problem of tumor heterogeneity, and the importance of the microenvironment of glioblastomas and outlines the knowledge of the molecular basis of the tumor under study. The introduction provides a concise characterization of epigenetic phenomena, chromatin structure, and transcription factors and the links between disorders of these regulatory elements and tumorigenesis. The subject presented in the introduction is very broad and it was not easy to select the most important facts. In my opinion, the PhD Student has managed this problem very well. The introduction is supported by a very rich and recent body of literature.

Materials and Methods

To achieve His research goals, the PhD Student had to use many of the methods mentioned earlier in this review. First of all, these were methods involving high-throughput analysis of nucleic acids and bioinformatics analysis of the collected data. The Author presented the methods and bioinformatics pathways in a sufficiently clear manner. The bioinformatics workflows were supported by very good graphics, which made it much easier to follow.
I do not understand why the description of the ChIPseq method and DNA methylation sequencing appeared in this chapter. While reading the paper, I understood that the results of these analyses were derived from previously collected data.
The work presented is multithreaded and required collaboration of many people. In the „Acknowledgements" part of the Dissertation there are acknowledgements to researchers who collaborated with the Author on specific tasks. In my opinion, the Materials and Methods section is a good place to indicate who performed or co-performed a particular experiment. This would certainly make it easier to assess the enormous involvement of the PhD student in the presented work.

Results

During the course of the dissertation, the Student pursued tree major research objectives.

Objective 1: involved the characterization of genomic and transcriptomic profiles in HGG after relapse.

In this task, the PhD student performed comparisons at the molecular level of primary and recurrent tumors: at the genome level using targeted sequencing of 700 cancer-related genes and at the transcriptome level using the RNA -seq method.
Sequence analysis of the defined panel of genes showed that only a small number of pathogenic mutations were common to both primary and recurrent HGG. This is in agreement with previous data and the well-known issue of glioma heterogeneity. Detailed mutation summary revealed the presence of previously known somatic mutations in genes such as *TP53, PTEN, PIK3R1, IDH1, ATRX* and *PIK3CA*, the most frequently altered genes in both primary and recurrent HGG. A previously undescribed frame-shift insertion was discovered in the *ZNF384* gene, which encodes a C2H2-type zinc finger protein that functions as a transcription factor, possibly regulating extracellular matrix genes. Copy number variation (CNA) analysis based on sequencing of DNA isolated from blood (reference) and from primary and recurrent tumors of the same patient revealed frequent duplications in the region of chromosome 7 involving the *EGFR* gene and deletions on chromosome 10 involving the *PTEN* gene. Similar changes at the level of CNAs were identified in both primary and recurrent HGG. The data obtained are consistent with the results collected within the GLASS consortium.
In turn, comparative transcriptomic analysis of recurrent and primary GBAs followed by functional studies identified significantly altered signaling processes and pathways. Genes belonging to the IFN signaling, phosphatidylinositol (PI) signaling, and PI metabolism pathways, indicating remodeling of

the tumor microenvironment (TME), increasing migration, and invasion of tumor cells, respectively, were among the genes with elevated expression.
Next, using a list of differentially expressed genes (DEGs) in recurrent and primary HGGs and the xCell webtool allowing for in silico cell type enrichment analysis based on hematopoietic cell transcriptomic signatures, the PhD Student demonstrated major differences between the compared tumor stages. Significant accumulation of pro-tumorigenic macrophages and immunosuppressive dendritic cells was found in recurrent GMBs, as confirmed by immunocytochemical staining of tumor samples. This finding indicates the important role of the tumor microenvironment in tumor aggressiveness.

The results of this part of the work have already been published in 2021 in the Journal of Molecular Medicine.
Roura AJ, Gielniewski B, Pilanc P, Szadkowska P, Maleszewska M, Krol SK, Czepko R, Kaspera W, Wojtas B, Kaminska B. Identification of the immune gene expression signature associated with recurrence of high-grade gliomas. J Mol Med (Berl). 2021 Feb;99(2):241-255. doi: 10.1007/s00109-020-02005-7. Epub 2020 Nov 19. PMID: 33215304.

Objective 2. Identification of mechanisms of transcription deregulation and chromatin remodeling in HGGs when compared to benign gliomas

Transcription factors (TF) are commonly deregulated in cancerous cells and have a major contribution to the pathogenesis of human cancer.
In presented work to identify key transcription factors for HGGs, the ATAC-seq was performer. This modern technology serves to map the accessible chromatin regions and provides information about spatial changes in chromatin structure and regions of transcription factors binding associated with gene expression.
The following approach adopted in this work allowed for identification of the potential TF targeting open chromatin regions in GBMs.
First ATAC-seq was performed for both LN18 and LN229 lines and two GBM specimens. Chromatin accessibility data were obtained for all cases. Based on the these data and using the HOCOMOCO human motif database, prediction of transcription factor binding sites in GMBs was made.
Further analyzes were performer only for predictions of transcription factor binding sites (TFBS) identified in both cell lines. Next predicted TFBS were integrated with upregulated genes in GBMs samples (grade IV glioma) or LGGs (grade II glioma). Comparison of GMBs and LGGs transcriptomes was performed on data deposited in TCGA. Data integration led to the discovery of specific transcription factors for overexpressed genes in grade II or grade IV glioma. Stage dependent pattern of transcription of identified GBMs and LGGs-specific TF was established. Further analysis indicated that *c-JUN* , a very well-known proto-oncogene, is the transcription factor bound to promoter regions of highly overexpressed genes in GBM with biological functions involved in glioma progression. Characterization of *c-JUN* transcript level in different WHO glioma grades showed positive association of increasing c-*JUN* expression with increasing tumor malignancy confirming the importance of this TF in gliomas. Additionally the significance of the expression of all 16 genes potentially regulated by c-JUN for the survival of patients with LGGs and GBMs was examined. In all cases, increased expression levels indicated poor prognosis, but only for patients with LGGs and not GBMs.

(Specific questions - I would like to know the number of genes over-expressed in GBM matched with the number of potential binding sites in the HGGs genome for the transcription factor c-JUN).

Next the work was focused on the identification of enhancers - critical elements for transcriptional regulation - harboring c-JUN binding motifs. To reach this goal previously collected data in the laboratory concerning active enhancers in different glioma staged were integrated with information about c-JUN motifs present in open chromatin in GMBs. The analysis gave 94 specific binding sites for c-JUN.

Last decade showed the importance of 3D genome structure in gene regulation. Now it is accepted that physical contact between enhancer and its target promoter may be critical for gene activation.
To discover possible connections of enhancers harboring c-JUN motives and promoters of protein coding genes, the available Hi-C chromatin data from a human brain development conformation analysis were employed giving several candidates as c-JUN regulated genes.
At this point DNA methylation was additionally considered as an important element of gene expression and is was examined if methylation level of c-JUN regulated promoters might be involved. The differences in DNA methylation were observed when comparing GII/GIII –IDHmut gliomas with GIV glioma stage, leading to the conclusion that this modification could be a part of regulatory mechanism of c-JUN controlled genes.
This part of the PhD work was finalized with electrophoretic mobility shift assay (EMSA) experiment confirming c-JUN binding to *VIMENTIN* promoter, that was computationally established in this work as one of main genes regulated by c-JUN. Vimentin is known as an intermediate filament protein expressed in all mesenchymal cells and linked to cancer.

(Given the complexity of how c-JUN factor works - as described in your discussion - which is that c-JUN is a component of the dimeric transcription factor AP1 and can form dimers with various TFs, additionally interacts with factors outside of AP1, additionally can recruit the SWI/SNF remodeling complex and is a subject to post-translational modifications, not to forget DNA metylation. And all of this may affect the specificity of the interaction of this factor with DNA - how would you confirm your results determining the sites of interaction in the genome for c-JUN and the genes regulated by this TF?
Do you think that transcription factors (including c-JUN) could be a therapeutic targets, leading to better cancer treatments?)

Objective 3: Defining the role of SMARCA2/4 remodelers in chromatin regulation in gliomas

ATP-dependent chromatin remodeling is essential for nearly all aspects of DNA metabolism, including transcription. The SWI/SNF complex consisting of more than a dozen subunits was the first identified chromatin remodeling machinery. In fact one can consider SWI/SNF as a family of complexes whose enzymatic drivers are two mutually exclusive SNF2 family ATPases called SMARCA2 and SMARCA4, and whose other subunits are encoded by small gene families. SWI/SNF complexes in mammals are involved in chromatin remodeling in enhancers, promoters, and gene bodies and are associated with gene activation and repression. Genes encoding subunits of SWI/SNF complexes are mutated in approximately 20% of human cancers and in most of the cases there is a significant deregulation in expression of genes encoding sub-units of SWI/SNF.
In the presented work, based on data deposited at TCGA, it was found that m-RNA levels of both *SMARCA2* and *SMARCA4* is significantly down-regulated in high grade gliomas comparing to lower-grade tumors. To study the effect on the chromatin structure and gene expression of the lowered level of these ATP-ases in glioma cell, the siRNA method was employed to silence *SMARCA2* or *SMARC4* or both genes in LN18 glioma cells. Significant down-regulation of analyzed genes both on the mRNA and protein level contributed to the phenotypic change of the glioma cells. In all cases, the cells were characterized by reductions in viability and proliferation.
(How would you explain this result, it seems contradictory with the lower level SMARCA2 and SMARCA4 in HGG?)
ATAC-seq analysis was performed for *SMARCA4* and *SMARCA2/4* downregulated variants. In particular, a difference in chromatin accessibility was observed in the cells with reduced expression of both enzymes, mainly it was higher chromatin accessibility in promoter regions in a number of protein coding genes, however there were also identified regions with the decreased chromatin accessibility.
(How would you explain the appearance of regions with increased chromatin accessibility in cells with reduced level of chromatin remodeler ?)

Examining the genes with changed chromatin status in their promoters in *SMARCA2/4* depleted cells versus controls provided information about biological pathways disturbed in these cells and possibly regulated by SWI/SNF. Genes with promoters located in open chromatin were involved for example in pathways related to cell projection organization and neuron projection development, while genes with promoter deceased chromatin accessibility in pathways related to protein localization, signal transduction and cell communication.
Promoters of *SMAD1* and *BMPR1A* genes involved in the TGF-β pathway were also located in chromatin open regions in *SMARCA2/4* depleted cells. To confirm the upregulation of these genes and examine changes in the TGF-β pathway the Western blotting and (GAGA)-dependent luciferase reporter were performed.
(Specific questions - I am wondering if the level of SMAD1 and BMPR1A was validated on mRNA. On the western blot for BMPR1A one can observe higher level of this protein also in control cells – what could be the reason, please comment on that)

Discussion and Summary

In the Discussion chapter, the PhD Student summarized and interpreted the results obtained in relation to current knowledge. He also presented open research paths.
The dissertation ends with a clear summary and the most important conclusions of the work.

Summary

The presented work is very extensive and rich in results, therefore I did not refer to all of its aspects, I chose the most important ones. The work contains an impressive amount of results. The Doctoral Student applied many modern research methods, conducted complex bioinformatics analyses and integrated data from many different experiments. Thanks to this approach, He obtained very interesting results, which He presented and described in a clear way. The presented work, like other very good ones, opens up the field for further questions. I am convinced that after publication of the results obtained during this PhD, the work will find many interested readers and will have an impact on the HGG research area.

I positively evaluate the dissertation submitted to me for review by Mr. ADRIA-JAUME ROURA CANALDA and conclude that it meets the conditions specified in Article 187 of the Act of 20 July 2018. Law on Higher Education and Science (Journal of Laws of 2021r pos.478, 619, 1630). Therefore, I request the Scientific Council of the Nencki Institute of Experimental Biology of the Polish Academy of Sciences to acknowledge the thesis as meeting the requirements for doctoral dissertations and to admit ADRIA-JAUME ROURA CANALDA to further stages of the doctoral program.

Due to the high level of the obtained results I apply for awarding with appropriate distinction.

[Pozytywnie oceniam przedłożoną mi do recenzji rozprawę doktorską Pana ADRIA-JAUME ROURA CANALDA i stwierdzam, że spełnia ona warunki określone w art. 187 Ustawy z dnia 20 lipca 2018r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021r poz.478, 619, 1630). W związku z tym zwracam się do Rady Naukowej Instytutu Biologii Doświadczalnej im. Marcelego Nenckiego PAN o uznanie pracy za odpowiadającej wymogom stawianym pracom doktorskim i dopuszczenie ADRIA-JAUME ROURA CANALDA do dalszych etapów przewodu doktorskiego.

Ze względu na wysoki poziom uzyskanych wyników zwracam się o przyznanie odpowiedniego wyróżnienia.]

Dr Hab. Marta Koblowska, prof. ucz

Review of the doctoral thesis of Adria-Jaume Roura Canalda „**A multi-omics evaluation of somatic mutations, transcriptomic dysregulation, chromatin accessibility and remodeling in High-Grade Gliomas**" conducted under the supervision of Professor Bożena Kamińska and Professor Bartosz Wojtaś at the Laboratory of Molecular Neurobiology of the Nencki Institute of Experimental Biology Polish Academy of Sciences.

Since the publication of the first draft of the human genome, we have witnessed an unprecedented surge of genomic information in the biomedical sciences due to the introduction and improvements of sequencing technologies and related applications. This progress wouldn't be possible without the subsequent development of new bioinformatics tools to manage and analyze the large amounts of data generated by modern biology. New genomic applications have accelerated the study of the genetic and molecular basis of cancer development. For a more complete understanding of cancer pathogenesis and the search for new therapies or treatment regimens, multicenter pan-cancer studies have been initiated (eg. the TCGA initiative) to catalog somatic DNA variants, RNA expression, and subsequent, gene expression-related biochemical and structural changes in chromatin. Cancer resistance to the treatment and progression determine the fatal outcome of this devastating disease. The likelihood of tumor recurrence after initial multimodal treatment is especially very high for high-grade gliomas (HGGs), therefore deciphering the genetic and molecular background of recurrent HGGs could define the molecular mechanism behind the rapid progression. This problem is comprehensively undertaken in Adria's dissertation where several state of- art genomic applications and big-data analysis approaches are employed to define genomic, transcriptomic, and epigenetic alterations associated with HGG progression and recurrence. The issues that I would like Adria to address during the defense are highlighted in bold text.

The dissertation layout includes the table of contents, a list of abbreviations, figures, and tables legends followed by abstracts in English and Polish. The dissertation contains all the required elements of introduction, description of research material and methods, presentation of results, and their discussion. The dissertation ends with a list of 336 references (almost all of them from the present millennium).

In the introduction, the opening section gives a general overview of gliomas and current treatment options at the same time acknowledging the limitations of gliomas histological and molecular classification due to their intra-tumoral heterogeneity and the influence of tumor's microenvironment. The following section briefly presents the WHO gliomas histological classification underlying the role of common genetic alternations in this classification. The next three sections in more detail describe the role of genetically diverse multiple cancer cell subpopulations as well the role of the tumor microenvironment as the source of glioblastoma's heterogeneity ultimately responsible for its aggressiveness, treatment failure, and recurrence. In the next section, the current state of knowledge on changes in the epigenome of gliomas is extensively presented providing in the subsequent subsections relevant examples of DNA methylation deregulation, histone code, and chromatin structure alternation, and finally the role

of chromatin remodeling complexes with the emphasis on the SWI/SNF complex. The final section of the introduction familiarizes the readers with the role of transcriptional factors (TFs) in tumorigenesis and provides several examples of their deregulation observed in gliomas. In conclusion, the introduction to the research topic and the definition of the scientific problem have been described comprehensively. The extent of the individual sections of the introduction is adequate, without going unnecessarily beyond the scope of the dissertation. Overall the introduction lays the solid ground for the rationale to undertake the studies presented in the dissertation. Since most glioma patients develop recurrent disease and still the knowledge on the molecular patterns of glioblastoma recurrence, especially on epigenetic level, is limited therefore there is a pressure need to further explore this important scientific question. The main aim and specific aims of the thesis are clearly articulated.

The methods chapter is very well written. It contains sufficient descriptions of the materials and protocols used. The details on experimental procedures are appropriate and easy to follow. The processing steps of raw data from sequencing and microarray applications, the use of external data, and the data analysis pipelines are described thoroughly. **Could it be explained why two different protocols of DNA and RNA isolation from gliomas are included in subsubsections 2.1.2 and 2.2.2? Also in the 2.1.2 section, the concentration unit of Proteinase K appears to be incorrect.**

The first section of results starts with a brief description of the patients' group and provides a general overview of the somatic mutations landscape detected with a custom-made 700 gene panel. In concordance with previous reports TP53, PTEN, PIK3R1, IDH1, ATRX, EGFR, and PIK3CA were the most frequently altered genes in both primary and recurrent gliomas. Importantly, more variants were unique for either primary or recurrent samples than shared between them. This is an interesting observation that later in a discussion is interpreted as driven by the emergence of a clone that survived the treatment pressure and initiated progression. Additionally, the recurrent frameshift (Q515fs) mutation in ZNF384 gene was identified and designated as frequently mutated in primary gliomas by the specific mutation clustering approach. The same method confirmed that mutational hot-spots in TP53, IDH1, and PIK3R1 genes were similar in primary and recurrent tumors. **The 3.1.3 subsection on gene splicing deregulation would fit better in the 4.1 section as it relates to RNA-Seq data.** Nevertheless, the increase in novel splicing events in recurrent gliomas is an important observation that was additionally supported with the functional analyses of transcriptomic data where down-regulated genes involved in mRNA splicing, among the cell cycle, and DNA repair were significantly overrepresented in the recurrent tumors. On the other hand, the upregulated genes after relapse were overrepresented in the functional categories related to interferon signaling and immune response. This observation prompted further exploration of bulk RNA-Seq data to find the expressional signatures of immature dendritic cells and M2 macrophages, an observation that was then validated using immunohistochemistry. The data presented in the first section of results were published in the Journal of Molecular Medicine and were featured with an editorial that undoubtedly underlines the importance of this research for understanding gliomas relapse. This

part of the results brings some additional questions that I'd like Adria to elaborate on. **Would it be possible to calculate the tumor mutational burden (TMB) for this genetic custom panel to compare it between primary and recurrent gliomas? Since the signature of immune response was dominant in recurrent gliomas if there is a relation of this signature to the TMB as recently noted by Gromeier et al. (Nat Commun . 2021 Jan 13;12(1):352)?**

The next section of the result is a demonstration of Adria's proficiency in handling and visualizing large-scale datasets. By using chromatin accessibility (ATAC-seq) data generated for LN18 and LN229 cell lines and a pair of patients' tumor samples the glioma grade-specific TFs binding sites were identified and then overlaid on on-house generated and external datasets, including RNA-Seq, DNA methylation, Chip-Seq, and Hi-C data. This analysis ultimately identified c-Jun protein as associated with gliomas progression. Interestingly, c-Jun expression and phosphorylation positively correlated with mRNA expression of its targets, including FOSL2 and VIM gene, in the cohort of glioma patients from the TCGA project. Finally, the observation of c-Jun recruitment to VIM promoter region was experimentally confirmed in four cell lines using EMSA assay and the highest c-Jun occupancy was observed in the cells lines derived from glioblastoma. **I wonder why the EMSA was used instead of a ChIP-qPCR to detect c-Jun recruitment to the VIM promoter?** In sum, this section of results advocates that Adria has mastered working with scientific big data, that he can ask relevant scientific questions, and proposes experiments confirming in-silico predictions.

In the last section of the results, the role of SMARCA2/4 genes in LN18 glioma cell line was investigated by measuring the consequences of their depletion alone and in combination on chromatin accessibility. The motivation was the observation of progressive SMARCA2/4 mRNA decrease in glioma malignancy grades in the TCGA dataset. **With this regards the SMARCA4 expression in GBM is a bit controversial. While the thesis shows a progressive decrease in SMARCA4 expressions with glioma staging the analysis of TCGA dataset by Peng et al. (Front. Immunol., 04 October 2021) and by Wang et al. (J Cell Mol Med. 2021 Mar;25(6):2956-2966) show increase in SMARCA4 expression GMB when compared to NB. Also, the REMBRANDT GMB dataset shows SMARCA4 increase in GMB. How these discrepancies in SMARCA4 mRNA expression could be explained?** The qPCR and Western blot (WB) data convincingly confirmed the decrease of SMARCA2/4 expression, although the proteins depletion was not complete. Despite that, it was possible to observe substantial changes in LN18 cells chromatin accessibility that encompassed regulatory elements near genes belonging to the TGF-b pathway. **Why the RNA-Seq measurements were not performed?** The impact of SMARCA2/4 depletion on abundances several constituents of TGF-b pathway, including SMAD1, SMAD3 and SMAD6, JAK1, BMPR1A, and TGFBR2 was independently validated by WB. Overall, the observation on TGF-b pathway being under the control of SWI/SNF in glioma cells is of great importance. Such dependencies have already been observed in tumors of epithelial origin upon epithelial-to-mesenchymal transition (Mol Cell Biol. 2013 Aug; 33(15): 3011–3025), and are worth further investigation in gliomas. In conclusion, the dissertation fulfills the objectives of the work.

The discussion is very well written. The first section briefly summarizes the contribution of conducted studies to glioma biology. This summary is compelling based on the presented data. In the subsequent sections, Adria has comprehensively examined the results against the scientific literature. **In the last sentence on page 126, the deleterious mutations in SMARCA2/4 are mentioned and linked to gliomas, however, I'm not aware of such for this cancer.** The conclusions correspond with the aims, and what is important, Adria presents a vision for further research and proposes further experiments to deepen the presented results.

My final evaluation of Adria's doctoral thesis is very positive. I believe that the dissertation handed to me for evaluation constitutes an interesting and extensive scientific study of great value. The work presented is multidisciplinary, as it combines expertise in cell culture, working with human specimens, next-generation sequencing libraries preparations, and most importantly big data handling and visualization. Therefore, I can consider Adria, without a doubt, as an experienced data scientist. I hope that he will continue his scientific career with passion.

The dissertation meets the requirements set out in the art. 187 of the Act from July 20, 2018 Law on Higher Education and Science (Dz. U. z 2021 r. poz. 478, 619, 1630). Therefore, I recommend to the Scientific Council of the Nencki Institute of Experimental Biology Polish Academy of Sciences to allow Adria-Jaume Roura Canalda to advance to the further stages of his doctoral proceedings. Given that part of the described studies already has been published, I also propose to award the dissertation.

dr hab. n med. Michał Mikula Warszawa, 13.02.2022

Laboratory of Bioinformatics and Computational Genomics LB!GO
Faculty of Mathematics and Information Science, Warsaw University of Technology
ul. Koszykowa 75, 00-662 Warsaw, Poland

Laboratory of Functional and Structural Genomics LFSG
Centre of New Technologies, University of Warsaw
Banacha 2c Street, 02-097 Warsaw, Poland

mobile: +48504726203, e-mail: Dariusz.Plewczynski@pw.edu.pl, www: https://plewczynski-lab.org

Warszawa,
22.02.2022

Warsaw, 22/02/2022

Prof. dr hab. Dariusz Plewczyński
Laboratory of Bioinformatics and Computational Genomics,
Faculty of Mathematics and Information Science,
Warsaw University of Technology
Laboratory of Functional and Structural Genomics
Centre of New Technologies
University of Warsaw

REVIEW of PhD dissertation of ADRIÀ-JAUME ROURA CANALDA, MSc

A multi-omics evaluation of somatic mutations, transcriptomic dysregulation, chromatin accessibility and remodeling in High-Grade Gliomas

Completed at the Laboratory of Molecular Neurobiology
Nencki Institute of Experimental Biology
Polish Academy of Sciences

under the supervision of
Prof. dr hab. Bożena Kamińska-Kaczmarek
and
Dr hab. Bartosz Wojtas

in the field of biological sciences
in the discipline of biology

The work presented to me for review is the result of a successful biological analysis in the multi-omics paradigm. The author managed to creatively combine the innovative experimental methodology with a biologically significant research problem focusing on a deeper understanding of the transcriptional regulation of the High-Grade Gliomas (HGGs).

Genomics, and in particular the molecular mechanics of chromatin regulation in cancers, is a highly interdisciplinary research field in which, to obtain a description of a biological phenomenon, we must use advanced high-throughput experimental approaches and tailored computational pipelines to identify the correct observables for the phenomenon and remove erroneous readings resulting from an experimental noise. In my opinion only by applying the orthogonal techniques, observing the biological system at multiple spatial and temporal scales using omics approaches, we can propose the significant features for the descriptive model. In the case study of high-grade gliomas we can actually identify the set of regulatory layers that are exploited by cancer to hijack the normal processes. The stage of the aggressiveness of a tumor is defined by DNA sequence mutations, methylation, affected transcription factors binding profiles, chromatin spatial reorganization – all contributing to the dysregulations of transcription.

In particular, mgr. Adrii Jaume Roura Canaldy in her thesis entitled "*A multi-omics evaluation of somatic mutations, transcriptomic dysregulation, chromatin accessibility and remodelling in High-Grade Gliomas*" focuses not only on the difference between normal tissues / cells and the tumor ones, but also about recurrences of HGGs. This allows to better understand the alternative pathways of oncogenesis and how cancer respond to applied treatments. Author also presents the analysis of the open chromatin regions using ATAC-seq experiments combined with transcriptomics that explain multiple routes of brain cancer progression.

The subject of my assessment, i.e. the doctoral dissertation, in my opinion is in the full accordance with the conditions set out in Art. 187 of the Act of July 20, 2018 Law on Higher Education and Science (Journal of Laws of 2021, items 478, 619, 1630). It presents the originality of the solved scientific problem, general knowledge theoretical candidate in the field of molecular biology, as well as the ability to conduct scientific work.

The doctoral dissertation of mgr. Adria Jaume Roura Canaldy was prepared in the Laboratory of Molecular Neurobiology at the Nencki Institute of Experimental Biology, Polish Academy of Sciences, under the supervision of prof. dr hab. Bozena Kaminska-Kaczmarek and dr hab. Bartosz Wojtas. The study was carried within TEAM-TECH Core Facility project funded by the Foundation for Polish Science and research grant by Polish National Science Centre Symfonia grant. The title of the TEAM-TECH project, i.e. "*NGS platform for comprehensive diagnostics and personalized therapy in neuro-oncology*" precisely matches the scope of the doctoral thesis. Similarly, the participation in the "*Atlas of regulatory regions specific to the human brain - a new tool to discover the pathways causing selected brain diseases*" grant equipped PhD candidate with the unique and useful set of concepts, tools, and techniques.

The author focuses on the three major topics relevant to the cancer: the recurrence, the role of SMARCA remodeler and the identification of TFBS in open chromatin regions. All studies were done on the either glioma samples collected from hospitals (including relapses), or cell lines acquired commercially from the American Type Culture Collection (ATCC).

The first research challenge is addressed in the context of somatic SNPs and short indels, the analysis of RNA-seq transcription profiles, and the dissection of tumor microenvironment. It would be exciting to see the more precise analysis of causative eQTLs, structural variants (DNA sequence modifications exceeding 50bp in length) and long-read (like ONT) splicing variants analysis of the collected transcripts. Yet, it requires much more extensive work that would exceed the time frame of PhD studies.

The second axis of the problem is directed toward SWI/SNF-dependent chromatin remodeling complex that bind to chromatin and cause nucleosomes displacement, which improve access to DNA sequence, and in turn initiate the transcriptional process. This protein-protein complex prevents the accumulation of Polycomb proteins performing tumor-suppressive function in the fundamental way by sustaining the balance between differentiation and self-renewal. Reviewer believes that it would be interesting to identify more precisely the structural role of SWI/SNF complex in shaping the high-resolution chromatin conformation, especially in the context of transcription activation by DNA looping, and not only at the scale of open/close chromatin analysis linked basically to the increase/decrease of the accessibility of target genes promoters.

Finally, the third component of the tumorigenesis is related to the specific transcription factors (TFs) that are observed in open-chromatin regions, which includes prediction of their binding sites in cis-regulatory regions. The study also includes methylation analysis of promoter regions. Reviewer is curious about the role of trans-regulatory regions that are mediated by the long-range chromatin loops initiated by CTCF, cohesin or RNAPOL2 proteins. How would it impact the regulatory program leading to oncogenesis? Nevertheless, mostly the cis-regulatory analysis of transcription factor binding sites predicted by bioinformatics algorithms is proven to be effective way to identify local promoter P and enhancer E localized TFs, which contributes to the protein-protein mediated micro chromatin loops forming EP contacts.

The formulation of a new holistic approach for studying cancer allows for detailed study of transcriptomics deregulation in cancer cells. PhD candidate successfully identified the most important transcription factors responsible for this deregulation that leads to tumorigenesis in High-Grade Gliomas. Moreover, he reached further by analyzing tumor immune microenvironment and clonal evolution, especially in recurrent cases.

The work proposes a new omics research paradigm combining the carefully selected set of the next generation sequencing (NGS) based experiments, while referring readers to the bioinformatics algorithms and tools used by Author to process and interpret the results of those experiments. Moreover, bioinformatics predictions were validated experimentally (for example c-Jun binding testing by the electrophoretic mobility shift assay). The selected research problem has important impact on human society, health, and the quality of life. The proposed omics methodology allows to identify the potential treatment strategies, or protein targets for drug design and discovery.

A very important research problem is motivated not only by scientific curiosity but also by the need of the community of genomic medicine researchers trying to identify the statistically significant biomarkers of the pathogenic cellular state by computational comparison of different NGS-based experiments. It is challenging due to a very complex problem, a high level of noise in the experimental data as well as sometimes illusory nature of genomic features. For example, some researchers believe that the observed bulk behaviors in RNA-seq, HiC, ChIP-seq are not real at the single cell level and result only from statistical aggregation of data in population-type experiments. This is an interesting argument that needs further inquiry in the context of gliomas and other cancers, within the advent of single cell biology.

The author presents the main theses of the PhD in two pages summary of the dissertation. It emphasizes the importance of the selected experimental methods and accompanying bioinformatics pipelines. The methods and tools proposed by the author fit into new research trend, enabling rigorous and statistically significant comparisons of various tumor bio samples. PhD candidate covered most of the available levels of regulatory programming, although the connection between the expression and the 3D structure of the chromatin is rather sketchily sketched. The link between DNA somatic mutations, methylation signatures, affected TF binding profiles, DNA accessibility, chromatin local and global connectivity given by 3C-type experiments, and the molecular phenotype of cancer given by RNA-seq, is not fully established. Nevertheless, within available time and reasonable funding effort the scientific achievements of PhD candidate are remarkable and does not raise any doubts.

In the first chapter, the author presents the introduction to various gliomas, its classification, heterogeneity, relevant DNA mutations, epigenomic dysregulation. PhD candidate focuses on introducing the basic biological concepts used at work. He does it efficiently and without an overly elaborate description, nevertheless providing key references to the literature on the subject.

In the second chapter, the author presents methods used in the study.

First, it addresses in detail the computational processing applied to the recurrent high-grade gliomas. It covers the description of a study cohort, panel design, library preparation and NGS step. Two bioinformatics pipelines are used, the first to identify DNA mutations, the second for transcriptomics analysis. In silico methodology of cell type enrichment, immunohistochemical and immunofluorescent immune cells characterization was also described. Moreover, TCGA database was extensively used to support both claims and the results obtained within the project.

Secondly, multi-omics analysis of transcription factors detection is presented, especially in open chromatin regions. Both human glioma cell lines and surgically resected tumors were used as bio samples in the study. As I mentioned earlier, omics approaches covered DNA and RNA sequencing, ATAC-sequencing, the identification of the differentially expressed genes between glioma grades (the tumor progression), ChIP-seq with the focus on H3K27ac

histone modification, search for the glioma enhancers, DNA methylation, cis- and trans- intra-chromosomal contacts between enhancers and promoters.

Thirdly, the selected key protein factor - SMARCA2/SMARCA4 was silenced as knockdown in Glioma cells and omics analysis was performed in detail, both experimentally and using bioinformatics tools. The results were supported by TCGA data, rigorous statistical analysis.

The third chapter is focused on the presentation of results across the whole study. First, somatic mutational landscape in cancer progression is presented including copy number aberrations, gene splicing changes, transcriptional profiles rearrangement accompanied by the tumor microenvironment evolution. Multi-omics integration, open chromatin and transcription factor analysis leads to the identification of c-Jun and other genes associated with the glioma progression. Finally, chromatin accessibility changes by SMARCA2 and SMARCA4 knockdowns is observed in human glioblastoma cells, where some open chromatin regions are observed in SMARCA-depleted cells, which leads to the conclusion that open chromatin differences are associated with some transcriptional regulation programs.

The fourth chapter is dedicated to the discussion and future perspectives of the omics method applied to genomic medicine of cancer. PhD candidate describes the clonal evolution and molecular mechanics of transcriptomic deregulation, more specifically the identification of high grade gliomas specific transcription factors, promotor regions and involved enhancers. The knockdown of SMARCA proteins confirm their role as chromatin remodelers involved in the critical signaling pathways in gliomas.

To sum up - the author has demonstrated knowledge of experimental and bioinformatics methods in genomics. He used his technical skills and deep biological knowledge to propose and implement novel omics methodology and implement it for the model system of glioma cancer. PhD candidate presents the unique combination of the ability to design and perform experiments together with the application of computer algorithms and knowledge of statistical data analysis. Unfortunately, the software scripts developed by him have not been made available as packages or make public in web servers (such as github), which may be a bit surprising. At least reviewer was unable to find the programming codes being reported in the thesis and reported publication.

The results of the PhD thesis were already published in one paper and one BioRXiv preprint. The other four manuscripts are under preparation, hopefully published shortly. Three publications not directly related with the PhD are already published in internationally recognized journals.

The reported by author manuscripts related to the PhD thesis:

- [P1] **Adria-Jaume Roura**, Bartlomiej Gielniewski, Paulina Pilanc, Paulina Szadkowska, Marta Maleszewska, Sylwia K. Krol, Ryszard Czepko, Wojciech Kaspera, Bartosz Wojtas, and Bozena Kaminska. "*Identification of the immune gene expression signature associated with recurrence of High-Grade Gliomas*"***Journal of Molecular Medicine*** 99, no. 2 (2020): 241-55. doi:10.1007/s00109-020-02005-7;
- [P2] **Adria-Jaume Roura**, Paulina Szadkowska, Michal J. Dabrowski, Karolina Stepniak, Bartosz Wojtas, and Bozena Kaminska. "*The oncogenic transcription factor c-Jun regulates critical over-expressed genes in Glioblastoma and is widely involved in distal-regulatory glioma elements*";
- [P3] Chinchu Jayaprakash, **Adria-Jaume Roura**. Bartosz Wojtas, Bartek Gielniewski, Paulina Szadkowska, Sylwia K. Krol. "*Knockdown of SMARCA4 and SMARCA2, subunits of the SWI/SNF chromatin remodeling complex, deregulates open chromatin and transcription profiles in human gliomas*";

Other referenced works that are presently in preparation:

- [R1] Bartlomiej Gielniewski, Katarzyna Poleszak, **Adria-Jaume Roura**, Paulina Szadkowska, Sylwia K. Krol, Rafal Guzik, Paulina Wiechecka, Marta Maleszewska, Beata Kaza, Andrzej Marchel, Tomasz Czernicki, Andrzej Koziarski, Grzegorz Zielinski, Andrzej Styk, Maciej Kawecki, Cezary Szczylik, Ryszard Czepko, Mariusz Banach, Wojciech Kaspera, Wojciech Szopa, Mateusz Bujko, Bartosz Czapski, Miroslaw Zabek, Ewa Izycka-Swieszewska, Wojciech Kloc, Pawel Nauman, Joanna Cieslewicz, Bartosz Wojtas, and Bozena Kaminska. "*The Novel, Recurrent Mutation in the TOP2A Gene Results in the Enhanced Topoisomerase Activity and Transcription Deregulation in Glioblastoma.*" 2020. doi:10.1101/2020.06.17.158477 (bioRxiv, under revision in PLOS Genetics);
- [R2] Maria Banqueri, **Adria-Jaume Roura**, Anna Kiryk, Marie-Eve Tremblay, Bozena Kaminska. "*Transcriptomic responses of microglia to a chronic, unpredictable, mild stress in the prefrontal cortex and hippocampus in a murine model of depression*";
- [R3] Malgorzata Perycz, Marta Jardanowska, **Adria-Jaume Roura**, Bartlomiej Gielniewski, Karolina Stepniak, Michal J Dabrowski, Michal Draminski, Bozena Kaminska, Bartosz Wojtas. "*REST transcription factor holds the balance between the invasion and cell differentiation in IDH-mutant and IDH-wild type gliomas*".

Other published works that are reported but according to author not related directly with the PhD thesis:

- [P4] Aleksandra Ellert-Miklaszewska, Natalia Ochocka, Marta Maleszewska, Ling Ding, Erik Laurini, Yifan Jiang, **Adria-Jaume Roura**, Suzanne Giorgio, Bartlomiej Gielniewski, Sabrina Pricl, Ling Peng, and Bozena Kaminska. "*Efficient and Innocuous Delivery of Small Interfering RNA to Microglia Using an Amphiphilic Dendrimer Nanovector*" ***Nanomedicine*** 14, no. 18 (2019): 2441- 459. doi:10.2217/nnm-2019-0176;
- [P5] Ilona E. Grabowicz, Bartek Wilczyński, Bożena Kamińska, **Adria-Jaume Roura**, Bartosz Wojtaś, and Michał J. Dąbrowski. "*The Role of Epigenetic Modifications, Long-range Contacts, Enhancers and Topologically Associating Domains in the Regulation of Glioma Grade-specific Genes*" ***Scientific Reports*** 11, no. 1 (2021). doi:10.1038/s41598-021-95009-3;
- [P6] Paulina Pilanc, Kamil Wojnicki, **Adria-Jaume Roura**, Salwador Cyranowski, Aleksandra Ellert- Miklaszewska, Natalia Ochocka, Bartłomiej Gielniewski, Marcin M. Grzybowski, Roman Błaszczyk, Paulina S. Stańczak, Paweł Dobrzański, and Bozena Kaminska. "*A Novel Oral Arginase 1/2 Inhibitor Enhances the Antitumor Effect of PD-1 Inhibition in Murine Experimental Gliomas by Altering the Immunosuppressive Environment*" ***Frontiers in Oncology*** 11 (2021). doi:10.3389/fonc.2021.703465;

Below, I will try to briefly summarize the key points according to the author and reviewer, the PhD student's research achievements, including the proposed wet lab experiments and applied bioinformatics tools, as well as a general biological formalism related to the clinically relevant challenge of high grade gliomas characterization by omics paradigm exploiting contemporary genomics tools and experiments.

• Author identified omics signatures of sub-clone substitution during onco-progression. For example, the frame-shift insertion affecting the stability of ZNF384 protein, or inversely correlated copy number changes for the EGFR and PTEN genes;

• Recurrence of cancer was accompanied by the down-regulation of several genes coding for the components of the spliceosome machinery. Moreover, PhD candidate identified the immunosuppressive changes in the tumor microenvironment (the enrichment of M2 macrophages and immature dendritic cells);

• PhD candidate identified by binding motifs prediction the c-Jun transcription factor as the one of the key regulators in Glioblastomas. Moreover, the bioinformatics inquiry was further experimentally validated by analyzing the c-Jun binding to the VIMENTIN gene promoter;

• Going beyond the descriptive analysis PhD candidate performed knockdown of SMARCA family of chromatin remodelers and proved that they affect chromatin openness in glioblastoma cells, especially in the promotor regions of several master regulators of TGF-β signaling pathway. The silencing up-regulated the expression of TGFR2, SMAD1 and SMAD3 proteins;

• The characterization of epigenomic and transcriptomic profiles of cancer after recurrence was performed allowing for the identification of molecular mechanisms responsible of the pathological deregulation;

• The extensive integration of the high-throughput and the next-generation genomics data by applying advanced bioinformatics pipelines and carefully selected experimental methodologies supporting the biological questions;

• The identification of the pathogenic factors driving the progress of tumor invasion, which according to author allow to decipher of the future treatment options by proposing novel protein targets within chromatin remodellers and/or transcription factors;

• Finally, the identification of several gene regulatory networks that are up-regulated in tumor, affecting the homeostasis critical for the health, normal behavior of human cells. The findings were confirmed by the knockdown of SMARCA proteins that supported the equilibrium supporting bidirectional changes in the chromatin structure.

For the future reference it would be great to extensively analyze the wide networks of protein-protein interactions, signaling pathways, regulatory programs identified in the PhD study at the scale of whole cells computational modelling within Systems Biology paradigm. The present advances in live microscopy allow to study the cell migration, monitoring the invasion of pathogenicity into healthy tissue, and more precise descriptive data-driven models of tumor with the resolution of single cells.

Final conclusion

In the summary of my assessment of the doctoral dissertation of mgr. Adria-Jaume Roura Canalda under the title "*A multi-omics evaluation of somatic mutations, transcriptomic dysregulation, chromatin accessibility and remodelling in High-Grade Gliomas*", I say that I rate the presented work highly.

Taking into account the readability and scientific value of the doctoral dissertation, the successful combination of carefully described and computational tools, as well as biologically relevant experiments and fundamental research questions, I value the doctoral dissertation of mgr. Adria-Jaume Roura Canalda as an important contribution to molecular biology in the field of multi-omics and genomics of cancer.

The doctoral dissertation meets the conditions set out in Art. 187 of the Act of July 20, 2018 Law on Higher Education and Science (Journal of Laws of 2021, items 478, 619, 1630). Moreover, I believe that this dissertation exceeds all customary and statutory requirements for doctoral dissertations, constitutes an original solution to a scientific problem, demonstrates the candidate's general biological knowledge in cancer and techniques of molecular biology and demonstrates the ability to independently conduct scientific work.

Therefore, I am asking the Council of the Nencki Institute of Experimental Biology, Polish Academy of Sciences to admit mgr. Adria-Jaume Roura Canalda to the next stages of his doctoral dissertation.

In addition, considering the high substantive level of the dissertation, its careful preparation, transparent way of presenting the research topics, methodology and results, I would like to request that the dissertation be distinguished with an appropriate award.

Dariusz Plewczynski, PhD, Professor of Exact and Natural Sciences
Principal Investigator and Head
Phone: +48 22 554 36 54 or +48 22 234 7219, mobile: +48 504 726 203
e-mail: d.plewczynski@cent.uw.edu.pl or Dariusz.Plewczynski@pw.edu.pl www: https://plewczynski-lab.org

Laboratory of Functional and Structural Genomics LFSG
Centre of New Technologies, University of Warsaw
Banacha 2c Street, 02-097 Warsaw, Poland

Laboratory of Bioinformatics and Computational Genomics LB!GO
Faculty of Mathematics and Information Science, Warsaw University of Technology
ul. Koszykowa 75, 00-662 Warsaw, Poland