Review of the PhD thesis of Agata Sekrecka titled „The role of GAF, ISGF3 and IRF1 complexes in time-dependent IFNα and IFNγ activated transcriptional responses and functional overlap“

The PhD thesis completed at the Laboratory of Human Molecular Genetics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznań, under the supervision of prof. dr hab. Hans A.R. Bluyssen.

The dissertation is based on a hypothesis that differential binding (to specific regulatory sequences) of transcription factors activated after stimulation with IFNα and IFNγ regulates the positive feedback of the genes encoding these transcription factors (i.e. STAT1, STAT2, IRF9, and IRF1) and long-term response to interferons and explains functional overlap. Four objectives were specified. Their implementation was possible thanks to the integration of generated genomic data and their analysis for long-term transcriptional responses (both in wild type and relevant knockout cells), the genome-wide bindings, and the identification of regulatory elements in promoters of interferon-stimulated genes.

The **Introduction** section is strictly related to the topic of the dissertation. It starts from the general characteristics of interferons, their types, specificity, and production. Then, it concentrates on IFNα and IFNγ signaling describing the structure and mechanism of action of specific receptors, interacting proteins, downstream effectors (STAT and IRF families) and DNA binding preferences of their complexes, as well as possible crosstalk between signaling pathways. Current understanding of IFNα and IFNγ signaling pathways, their canonical and non-canonical functions was described in detail but also problems that need further investigation were signaled. The importance of maintaining a balance between induction and termination of the response to IFNs and the dysregulation of the IFN response observed in patients with severe Covid-19 was also briefly described. The Introduction is illustrated by seven clear and useful figures and four tables. It is logically organized, comprehensive, and well written. This part of the thesis, as well as two review papers by the PhD student, show her extensive knowledge in the field of IFN signaling. Nonetheless, there are some minor mistakes/inconsistencies in the text, for example: “the complex of cyclic GMP–AMP synthase (cGAS) with stimulator of interferon genes (STING)” (page 3). In practice, cGAS does not form the complex with STING. Upon binding DNA, cGAS triggers the reaction of GTP and ATP to form cyclic GMP-AMP (cGAMP) and cGAMP binds to STING. By the way, it seems to me that there is an incorrect citation here: Li et al., 2018 (no information about cGAS-

STING pathway). Also, information that unphosphorylated STATs can dimerize in “head-to-head” orientation while phosphorylated dimers are in “head-to-tail” orientation (page 9) is inconsistent with Figs 2, 6, 7.

The **Materials and Methods** section contains all the necessary information about the experiments and analyses carried out in the dissertation sufficient to assess the credibility of the results. Almost all the necessary details are provided along with an explanation of the criteria for some selection options. The experiments were conducted in compliance with the appropriate standards. CRISPR/Cas9 gene editing, cloning, mutagenesis, PCR primers, etc. were properly designed and necessary controls included. My objection to this section is that in the dissertation there is no reason to exclude tables and figures relating to methods from the main text and transfer them to the supplement.

Minor comments/questions to this part:

1. It is not clear how/if both gRNAs are expressed from one plasmid used for the generation of Huh7.5 CRISPR/Cas9 knockout cell lines (page 43).
2. There is no information as to whether RNA has been tested for DNA contamination. Although contamination is unlikely since RNA used for reverse transcription was digested with DNase during isolation (on-column DNA digestion) and then 1μg was taken for reverse transcription and digested again. On the other hand, there is a concern that the amount of template for cDNA synthesis may vary between samples.
3. For RT-qPCR, only one reference gene was used (more than one is commonly recommended).
4. Parameters (coverage etc) of sequencing are not provided.
5. I wonder if the double cross-linking (disuccinimidyl glutarate, DSG, together with formaldehyde) was first tested for effectiveness in this experimental model or if it was applied according to previous recommendations for STATs or IRFs.
6. Were all read numbers across 4 kb region centered on ChIP-seq peak summits computed and assigned to one peak in analyzes of binding profiles (page 60)? Or were read numbers computed for all peaks with summits located within the 4 kb region defined as a promoter?
7. Was the same definition of the promoter used for data integration with the BETA tool as for the binding profiles analyzes?

The **Results** section started from the presentation of the experimental model. Knockout (KO) human hepatocellular carcinoma Huh7.5 cells for STAT1, STAT2, IRF9 used in the study were obtained from prof. Kiyonao Sada from University of Fukui, Japan. They were generated using a lentivirus-based CRISPR/Cas9 system. IRF1 KO and double

IRF1/IRF9 KO cells were generated by the PhD student using a "double-nicking" strategy to minimize off-targets (a plasmid vector containing all components necessary for the genome editing and selection of edited cells was introduced using Lipofectamine3000). Several clones obtained after selection were first tested by western blot for the presence of IRF1 after IFNγ stimulation and then *IRF1* genomic locus was analyzed. It was assumed that the entire 55 bp sequence between nicking sites of Cas9D10A should be completely removed. Clones with indels were considered as improper ones (although functional KO was confirmed by western blot). I am impressed by the high effectiveness of the procedure (especially, I did not expect that the entire fragment between single-strand nicking sites generated by Cas9D10A would be removed). As differences between clones can often be observed after clonal selection, two clones for each KO were chosen for an initial comparison of the transcriptional response to IFNα and IFNγ. Finally, in further studies, one KO clone was used as representative for each: STAT1, STAT2, IRF1, IFR9, and IRF1/IRF9.

Characterization of time-dependent IFNα and IFNγ response (by western blot, RT-qPCR, RNA-seq, and ChIP-seq, then validation of some results by RT-qPCR, ChIP-qPCR, and luciferase-based assays combined with site-directed mutagenesis) was first performed in wild-type Huh7.5 cells, then in KO cells. The analysis and presentation of the results of such extensive genomic studies was a great challenge, especially as there are no established standards. The results are presented clearly and logically. Although the analysis of the results was very demanding, the overall message is generally consistent. RNA-seq analysis (focused on up-regulated genes) revealed that the transcriptional response to IFNα was stronger and faster than to IFNγ but was quenched earlier. This was also evidently visible when only overlapping genes (i.e. responding to both IFNs) were compared. Interestingly, RNA-seq and ChIP-seq data integration combined with the identification of GAS and ISRE motifs in promoters of interferon-stimulated genes showed that GAS motifs dominated in the group of early responding genes while ISRE motifs dominated in genes activated later after stimulation. Moreover, GAS motifs dominated in promoters rapidly interacting with pSTATs (especially pSTAT1) after stimulation (as suggested by the cluster and motifs analyses of time-dependent binding profiles of both pSTATs). The analysis focused on genes commonly up-regulated by IFNα and IFNγ allowed the assessment of their transcriptional overlap and function. To further study the dependence of the transcriptional response to both IFNs on STAT1, STAT2, IRF1, and IFR9, knockout cells characterized in chapter 4.1 were used. Analysis was focused on genes identified as up-regulated by both IFNs and with detected

binding in the promoter of at least one: pSTAT1, pSTAT2, IRF9, or IRF1 in WT cells. Presented heatmaps indicate that STAT1 or STAT2 KO did not have as strong effects on IFNα stimulated transcription as IRF9 KO (which resulted in a general drop in the expression). After treatment with IFNγ, gene expression was wholly abrogated in the STAT1 KO cells (although it seems that such conclusion was drawn from RT-qPCR analyses and RNA-seq was not performed). Further detailed analysis (connected with functional studies based on luciferase reporter assay) focused on individual genes with GAS, ISRE, or both regulatory sequences showed that GAS genes depend not only on STAT1 while composite and ISRE genes differentially depend on STAT1, STAT2, IRF9, and IRF1. More importantly, these analyses clearly showed that STAT1, STAT2, and IRF9 are part of a positive feedback loop that controls long-term IFN responses. Moreover, IRF1 controls basal expression of IFNα and IFNγ-induced genes.

In RNA-seq analyses, the smallest changes in the expression (log2 FC<0.5) were not taken into account. This is biologically reasonable. I wonder, however, whether this arbitrarily taken cut-off point is too low. It is reasonable to keep the same criteria when the number of up-regulated genes in each time-point are compared but the interpretation of heatmaps with z-score scaling can be misleading if only one sample per row meets this criterion. Nevertheless, the analysis allowed comparing the global transcriptional response in time (up to 72h of IFNs treatments) and the overlap between IFNα and IFNγ. Additionally, the results were validated by RT-qPCR on independent material. My main objection to this part is that no transcriptional data was shown after IFNγ stimulation in STAT1 KO cells (Fig. 29 et seq.) where gene expression was completely abrogated. Such data should not be presented only as supplementary (Figure S 5) but in the main text. Their absence makes it difficult (or even impossible) to draw correct conclusions directly from the figures. For ChIP-seq analysis, several time-points of IFN treatment were selected based on the results of the kinetics of STATs phosphorylation and IRF1 and IRF9 expression. I appreciate the scope and quality of ChIP-seq analyses (especially, the discarding of peaks with the lowest score is a good idea).

I am impressed with the amount of work, both experimental and computational, done in the dissertation. However, it should be mentioned that as a reviewer with a biological background, I am able to fully assess the methodology of experimental research but not bioinformatics analysis. I greatly appreciate the clarity and consistency of the data presented. Nevertheless, I have some specific comments and questions listed below.

Comments/questions related to the Results section:

1. It is not clear to me whether real WT cells were used for these analyses or so-called WT clones (i.e. v.1, v.2, v.3 clones treated with a non-targeting sgRNA). This is quite important since cells subjected to CRISPR/Cas9 editing (here using lentiviruses or transient transfection, then antibiotics and clonal selection) may substantially differ from parental cells.

2. Western blot (Fig. 16A) indicated that STATs (STAT1 and STAT2) phosphorylation and IRF9 expression after IFNα treatment were almost separated in time. If so, how ISGF3 may be active? Moreover, I would not say that the correlation between the max expression of *MX1* and *OAS2* (8-24h) and STAT1 and STAT2 phosphorylation (0.5-4h) exists (Fig. 16), as concluded on page 69. Also, the results of the western blot analysis (STATs phosphorylation, IRF1 expression) appeared to be inconsistent with the ChIP-seq data showing increased pSTAT binding after long-term treatment with IFN and IRF1 binding before its expression is detected by western blot. How can this be explained? In my opinion, a longer blot exposition for IRF1 should be shown (it is expressed in untreated cells as shown later). Maybe also for IRF9: ChIP-seq indicates that it binds to DNA at earlier time points than the protein is visible on western blot.

3. Although such data can be extracted from Fig. 20, it would be interesting to clearly show the percentage of binding sites without any motif. The cluster analyses shown in Figs 21 and 25 are very interesting but initially confusing and maybe some additional description would be useful. Otherwise, I only understand the general idea, but not the details of the analysis. I can only guess the meaning of lines between violin plots in Fig. 21.

4. Fig. 22 is aimed to show an overlap of IFNα and IFNγ stimulated genes after RNA-seq and ChIP-seq data integration. Thus, also GO terms enrichment analyses are restricted to genes upregulated by both IFNs. Genes containing GAS, ISRE3, or GAS+ISRE3 motifs in the promoter (in Fig. 22B-D) were analyzed separately. The results are very interesting. I wonder if they would be the same if not restricted to overlapping genes.

5. In Figs 23, 24 each line symbolizes the binding of one of the transcription factors. Thus, targets of all tested ones (shown in bright red color in Fig. 23) should be linked with STAT1, STAT2, IRF9, and IRF1. I can not see a connection to IRF9 for some of the genes in this group. Networks of integrated genes additionally showed that some genes identified (in RNA-seq and ChIP-seq data integration) as containing composite motifs in the promoter were bound only by one transcription factor. How to interpret this? Moreover, not all integrated genes seem to be included in networks. How were they selected?

6. Page 65: clones were analyzed by RT-PCR at the DNA level (?)

7. Page 66, line 3: c2 (WT) – probably v2 since c2 is not WT.

8. Fig. 18: why 318 genes, not 329?

9. Figs 25, 26: I wonder what the reverse analysis would look like? I.e. binding of STATs in time in GAS, ISRE, or composite containing genes. Why the analysis was restricted to the first time points?

10. Figs 27, 28, chapter 4.8: Based on the results of the western blot, I would not speculate about the possible functional role of protein complexes in GAS or ISRE binding.

11. Page 92: „The most substantial effect was connected with depletion of IRF1" (ISRE-containing group upon IFNγ stimulation). Based on the heatmap (Fig. 29B), I would say that the most substantial effect was connected with double KO (IRF9 and IRF1). It seems that dKO cells were not taken into consideration in the above analyses.

12. Page 93: "these genes" (which ones?) "… were not very sensitive to IRF1 or IRF9 protein depletion and preserved their expression to some extent in IRF9/IRF1 dKO cells." I can not see this on the heatmap (Fig. 29B).

13. Fig. 29. I would expect more information in the legend on which time point is missing in the analysis of RNA-seq in KO cells.

14. Page 94: "as assessed by RNA-seq and presented on the heatmaps (Figure 29), the expression of GAS-containing genes in KO vs WT cells marked their dependence on STAT1 and STAT2, but not IRF9 in response to IFNα" (I can not see this on the heatmap) "and STAT1, but not IRF9, upon IFNγ treatment." (STAT1 was not shown on the heatmap).

15. Fig. 31A, B, and others presenting peaks in IGV: How wide is the genomic region shown? Due to the too-small range of the scale, most of the pSTATs peaks are truncated, which makes comparative assessment difficult.

16. Fig. 32. Higher *APOL6* and *DTX3L* fold changes after stimulation in IRF1 KO suggest that IRF1 may be a suppressor of the transcription. Or that in untreated cells expression is much lower in IRF1 KO than in WT (as, in fact, was shown later in Fig. 38). Perhaps fold changes in all samples should be calculated vs WT untreated cells (or shown as normalized counts).

17. Fig. 33. In the case of *APOL6* and *DTX3L* both GAS and ISRE are in the same locus (detected by the same primers). It would be interesting to check the binding to motifs located separately, eg. in *ZC3HAV1*. No statistics are shown in D, so we can not say that ΔISRE responded more effectively to IFNγ than to IFNα (page 99).

18. Page 102: "components of ISGF3 (pSTAT1, pSTAT2 and IRF1)". I was sure that IRF9 is a component of ISGF3.

19. Fig. 35C. "ChIP-PCR validation of pSTAT1, pSTAT2, IRF9 and IRF1 binding to ISRE motif present in the promoter of *MX1* gene was performed". There are two ISRE motifs in this promoter, so which one was tested? Why such inconsistency with Figs 34 and 35D where other ISRE-containing genes are presented? More importantly, why was *MX1* (used as a model ISRE-containing gene) classified as a composite gene in Table S 4?

20. Page 104 (chapter 4.12). STAT1 and STAT2 were shown as ISRE only genes in Figs 23, 24, and in Table S 4 but not in Fig. 37. Why did RNA-seq and ChIP-seq data integration analyses give different results than those presented in Fig. 37 and Table 4?

21. Luciferase-based reporter assays would be also done in KO cells to assess the dependence on STATs and IRFs.

22. Lack of the protein molecular mass on western blots.

23. pSTAT1 and pST1 (or 2) are used interchangeably (I would prefer one style).

24. Fig. S 3. Should be: in response to IFNα and IFNγ.

25. Some mistakes in the description, e.g. Fig. S 10: G instead of H, NANO instead NANOG; Table S 1. Mistake in primer name from position 6.

26. Table S 4. Insufficient description in the table header. I have to guess what is shown inside.

The **Discussion** chapter summarizes the results obtained in the experimental part of the work in the context of the literature data (with the predominance of re-describing the results). In general, this part of the dissertation proves the candidate's knowledge of the topic and scientific maturity. However, I do not understand some interpretations of the data (e.g. the conclusion based on data presented in Figs 19 and 31 that GAS-binding sites occupied by GAF or GAF-like complexes are not recognized in the later time-points; page 112) or links between own and published data (e. g. these connected with COVID-19; page 123). The most important and valuable here is the final figure (Fig. 39) presenting the complex model of the time-dependent regulation of transcriptional responses induced by IFNα and IFNγ, which depend on unphosphorylated and phosphorylated ISGF3, GAF/GAF-like complexes, and IRF1. The thesis ends with conclusions (in eight points) and very accurate ideas for the future (in five points). The conclusions are correct, although in my opinion the first one is not fully documented (though it does not have to be; the conclusion is that the IFNs-induced accumulation of pSTATs, IRF9, and IRF1 leads to the formation of corresponding complexes, while it only enables; complex formation has not been studied). Furthermore, it would be interesting to have a more detailed discussion considering possible solutions to the questions enclosed in future ideas.

The organization of the dissertation is reader-friendly. In terms of graphics, the thesis is very nice and legible but contains editorial errors, e.g. some full stops in the wrong places, inconsistency in the spelling of gene/protein names, etc., which, however, do not affect the substantive message. The thesis is generally written logically and correctly in English, and interestingly, the weakest part is the summary in Polish (with grammatical errors and unacceptable phrases/mental shortcuts such as bardziej wystymulowane, kompozyt GAS+ISRE, miejsca kompozytowe, geny dla białek STAT1, e.t.c.).

Summary

The assumptions and goals of the study were clearly defined. The planned experiments were adequate to resolve the scientific problem and performed properly. The summary of the results sufficiently reflects the content of the dissertation. In general, the conclusions are correctly drawn and supported by the results. The PhD thesis proved the candidate's general theoretical knowledge in a discipline and the ability to independently conduct scientific work (although the analysis of genomic data had to be done in cooperation with bioinformaticians). Also, the subject of the PhD thesis is an original solution to a scientific problem. Therefore, the PhD thesis meets the conditions specified in Art. 13 section 1 of the Act of March 14, 2003, on academic degrees and academic title, as well as degrees and title in art (Journal of Laws of 2017, item 1789).

I am requesting for Agata Sekrecka to be admitted to the next stages of the procedure. In my opinion, the doctoral thesis is excellent (and has no serious weaknesses). Taking into account the wide scope of the research and its high quality, the logical layout of the work, and the importance of the research problem, the possibility of distinction can be considered.

Rozprawa doktorska spełnia warunki określone w art. 13 ust.1 Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki (Dz.U. z 2017 r. poz. 1789).

Wiesława Widłak

prof. dr hab. Wiesława Widłak

Narodowy Instytut Onkologii im. Marii Skłodowskiej-Curie,
Państwowy Instytut Badawczy, Oddział w Gliwicach

Gliwice, February 15, 2022