# Changes in heat shock duration influence regulatory schemes of HSF1 activity.

Iwanaszko Marta[1*], Janus Patryk[2], Tomasz Stokowy[1,3], Widłak Piotr[2], Kimmel Marek[1,4]

[1]Systems Engineering Group, Silesian University of Technology, Gliwice, Poland
[2]Center for Translational Research and Molecular Biology of Cancer, Maria Skłodowska-Curie Memorial Institute of Oncology, Gliwice, Poland
[3]University of Bergen, Department of Clinical Science, Bergen, Norway
[4]Department of Statistics, Rice University, Houston, USA

*corresponding author: marta.iwanaszko@polsl.pl

**Abstract.** The stress induced by thermal events at cellular and molecular level is known as the heat shock response and is one of the best conserved response systems in organisms. In eukaryotes gene regulation induced by heat shock occurs at both, transcriptional and post-transcriptional levels. Stress induced transcription is directed by the transcription factor family called Heat Shock Factors (HSF), of which the most prominent is HSF1. HSF1 recognizes and binds to DNA sequence, called Heat shock element (HSE) which can be found in the promoter of many, not only heat shock responsive genes. In this paper we show computational analysis of regulatory changes induced by different duration time of the heat shock condition in U2-OS wild type cells. U2-OS WT cells either not pretreated or subjected to heat shock for 10 or 20 minutes. Afterwards, the global analysis of HSF1 binding to DNA was made using next generation sequencing (ChIP-seq). Comparison of data from both experiments showed differences in ranking of top activated genes and shifts in HSF1 peaks in genes common for both datasets. This suggests different regulatory strategies in reaction to longer or shorter stress conditions, which we conclude to be an effect of transition between mild and severe stress. Additionally we analyzed occurrence and type of binding motifs found in promoter regions of genes with the strongest response to HS duration to define the most prominent HSF1 binding motifs. Heat shock proteins are known to play an important and positive role in a number of pathophysiological states including immunity against infection, ischemia and neural injury. HSF family also interacts with other crucial regulatory networks, thus knowledge of regulatory schemes used by HS response system may help in developing clinical protocols which utilize information about heat shock effects in diseases.

**Keywords:** HSF1, HSF, Heat shock, Gene regulation, Promoter region, Transcription factors.

# 1 Introduction

Regulation of gene transcription is a crucial component in the control of gene expression. The key to understanding gene regulation is identification of functional regulatory sequences, such as transcription factor binding sites (TFBS) in the promoter region of genes and other regulatory elements such as enhancers and silencers mainly present in noncoding DNA. Challenge for contemporary genomics is to understand how transcription is regulated.

Studies on the regulation of gene expression, conducted in recent years, allow us to explore the role of non-coding DNA in cellular response. Particularly interesting in this respect is the promoter region of the gene and its role in the process of transcription - a key step of gene expression. Cellular response to several types of suboptimal conditions that usually lead to denaturation of proteins, involves elevated expression of heat shock proteins (HSPs), and is usually called the heat shock response (HSR). Along with inducible HSPs, which are represented by the most abundant HSPA1, other members of the HSP family are expressed constitutively in the absence of stress. Constitutively expressed HSPs display distinct physiological functions for cellular maintenance and adaptation to stress, which is related to their roles as major molecular chaperones [1, 2]. Stress induced transcription is directed by the transcription factor family called Heat Shock Factors (HSF), of which the most prominent is HSF1. In response to different forms of cellular stress, HSF1 becomes trimerized and phosphorylated, and then binds to regulatory DNA elements (termed heat shock elements, HSE) [3, 4] present in HSP genes. In addition to regulation of HSP genes, HSF1 is involved in the regulation of several other genes coding for proteins partaking in various cellular processes, including cell signaling and maintenance of cell integrity [5, 6]. Furthermore heat shock proteins (HSPs) are known to play an important and positive role in a number of pathophysiological states including immunity against infection, ischemia, and neural injury [7]. HSF family also interacts with other crucial regulatory networks, such as NF-κB, thus knowledge of regulatory schemes used by HS response system may give better insight into cross-talk between major regulatory networks and help in developing clinical protocols which utilize information about heat shock effects in diseases.

In this paper we show computational analysis of regulatory changes induced by different duration time of the heat shock condition in U2-OS wild type cells. While this is work in progress, we present preliminary analysis, focused on promoter regions of genes with high response to HSF1 activation.

# 2 Methods

## 2.1 Experimental procedures

U2-OS WT (osteosarcoma cell line) cells were divided into three groups: two subjected to heat shock for 10 or 20 minutes accordingly, and not exposed control group. Cells were fixed with 1% formaldehyde and lysed. DNA was sonicated into fragments

of 100-500 base pairs (bp) of length and then ChIP assay was made using anti-HSF1 antibody or without antibody (mock). In the first stage, selected promoter regions of classical heat shock responsive genes, containing binding sites for HSF1, were analyzed by qPCR, using specific primers and ChIP-precipitated DNA. Ct values of all samples were first normalized according to their mock probes and then these ratios were referred to untreated control. Afterwards, the global analysis of HSF1 binding to DNA was made using next generation sequencing (ChIP-seq approach).

### 2.2 Analysis of High-Throughput Sequencing data

Raw Illumina sequencing reads were analyzed according to established standards of ChiP-Seq data analysis. Quality control of reads was performed with FastQC software [8]. Sequences with low quality (average phred<30) were filtered out. Reads accepted for the analysis were aligned to human genome (hg19) using bowtie2.0.4 [9]. Peak detection was carried with MACS [10], whereas the outcome was annotated with Homer [11]. Peak intersections and genomic coordinates handling was done by application of Bedtools [12]. Full dataset consist of three peak-lists: list of peaks after 10min exposition, 20 min exposition and control group. This dataset was reduced to two lists consisting of significant peaks after 10 min and 20 min expositions, which we denote as 'HS10' and 'HS20' accordingly. These two lists were subjects of further analyses. Analyzed peak sequences were located in promoter region, which has been defined as 1000 bp upstream and 500 bp downstream of transcription start site (TSS). From this selection we have analyzed top 100 peaks in both sets, ranked by peak score. Additional functional analysis was performed using Panther [13]. Analysis of regulatory elements was performed with aid of MEME-ChIP [14], TOMTOM [15], NucleoSeq [16] and R-based scripts. HSF1 binding motifs were identified using the most popular matrix presented in SwissRegulon [17] and two motif matrices presented in Pacholczyk et al. [18], these matrices were generated using knowledge based potentials.

## 3 Results

Main dataset shows differences in number and of location of detected binding sites. Dataset was reduced to only significant peaks in comparison with control data, after this step lists comprised of 19510 peaks in HS10 dataset and 17922 peaks in HS20 dataset. Detailed locations and counts of peaks are presented in Table 1. Analysis of peak scoring revealed that in case of HS20 set we see more high scored peaks in intergenic region and introns, than in case of HS10 set, if we count peaks that score more than 1000, in intergenic region we found 136 in HS20 set and only 64 in HS10, and in introns 127 in HS20 and 54 in HS10 set. In promoter and 5'-UTR regions count of the highest ranked peaks is greater in HS10 set. This may be due to the ongoing chromatin remodeling after longer exposure to the heat shock. This seems consistent with HSF1 need of active chromatin environment to bind HSEs [19]. The most significant differences in peak counts between HS10 and HS20 are in promoter region

(30.6 %) and 5'-UTR (28.8%), in favor of HS10 set. Taking this into consideration we decided to analyze peaks identified in promoter region, which we defined as 1000bp upstream and 500bp downstream of TSS. A list of 100 genes corresponding to the top scored peaks was chosen for further analysis. In this ranking we found 64 common genes between HS10 and HS20 datasets. We observed significant changes in ranking among genes in HS10 and HS20 datasets, and an offset between HS10 and HS20 peak location for given common gene. The only gene retaining its rank between both lists is HSPA6 which is also the highest ranked gene in main dataset. Analysis of genes which differ between HS10 and HS20 shows that time of heat shock exposure have impact on activation of different biological processes. In both groups 'Metabolic process' is the first ranked process with 35.1% in HS10 and 38.9% in HS20. In HS10 next ranked are 'Response to stimulus'(19%), 'immune system process' (14.3%) and 'developmental process' (11.9%). Next ranks in HS20 subset belong to: 'Cellular processes' (cell communication/cell cycle/cellular component movement; 19.4%), 'Biological regulation' (11.1%), 'Localization' (11%). Apart from that in HS10 we found 'Apoptotic process' (negative regulation) and 'Reproduction', which are not found among HS20 differential genes and 'Biological adhesion' only present in HS set.

| Region \ Exposure | HS10 | HS20 |
|---|---|---|
| Intergenic | 9440 | 8654 |
| Introns | 7466 | 7292 |
| Exons | 393 | 338 |
| Promoter* | **1606** | **1114** |
| TTS | 246 | 220 |
| 5'-UTR | **163** | **116** |
| 3'-UTR | 196 | 188 |
| Sum | 19510 | 17922 |

**Table 1.** Peak counts in genome locations. The biggest difference in count between data sets in bold. *Defined in initial annotation as 1k bp:TSS:+1k bp

We analyzed biological processes that are significantly overrepresented in HS10 and HS20 sets. Reference set contained 21804 biological processes described in Panther database, 87 genes were mapped in HS10 list and 84 genes were mapped in HS20 list. All significant processes are overrepresented in both lists; details are presented in Table 2. The most significant terms, for both HS10 and HS20 datasets, are 'protein folding' and 'response to stress'. For functional terms significantly overrepresented are distinctive to HS10 list, and are not significant in HS20 dataset, two of them are metabolic terms. Analysis of regulatory pathways associated with our dataset shows

| *Biological Process* | *Ref. list* | *HS10* | | | | *HS20* | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Counts* | *Expected counts* | *+/-** | *p-value* | *Counts* | *Expected counts* | *+/-** | *p-value* |
| **Protein folding** | 194 | 17 | .77 | + | **9.41E-16** | 11 | .75 | + | **7.07E-08** |
| **Response to stress** | 439 | 16 | 1.75 | + | **5.57E-09** | 9 | 1.69 | + | **1.28E-02** |
| **Protein metabolic process** | 2807 | 34 | 11.20 | + | **1.94E-07** | 25 | 10.81 | + | **9.31E-03** |
| **Protein complex biogenesis** | 79 | 7 | .32 | + | **9.29E-06** | 4 | .30 | + | 6.60E-02 |
| **Protein complex assembly** | 79 | 7 | .32 | + | **9.29E-06** | 4 | .30 | + | 6.60E-02 |
| **Cellular component biogenesis** | 114 | 7 | .45 | + | **1.08E-04** | 4 | .44 | + | 2.58E-01 |
| **Primary metabolic process** | 7177 | 48 | 28.64 | + | **3.98E-03** | 38 | 27.65 | + | 1.00 |
| **Response to stimulus** | 1671 | 18 | 6.67 | + | **2.31E-02** | 11 | 6.44 | + | 1.00 |
| Metabolic process | 8613 | 50 | 34.37 | + | 1.32E-01 | 44 | 33.18 | + | 1.00 |
| Tricarboxylicacid cycle | 23 | 2 | .09 | + | 0.98 | 1 | .09 | + | 1.00 |
| Developmental process | 2846 | 14 | 11.36 | + | 1.00 | 11 | 10.96 | + | 1.00 |
| Multicellular organismal process | 1798 | 6 | 7.17 | - | 1.00 | 5 | 6.93 | - | 1.00 |
| Mytokine production | 2 | 0 | .01 | - | 1.00 | 0 | .01 | - | 1.00 |
| Lipid metabolic process | 902 | 2 | 3.60 | - | 1.00 | 4 | 3.47 | + | 1.00 |

**Table2**: Top 15 functional terms ranked by adjusted p-value (multiple testing correction: Bonferroni correction method). Significant processes are in **bold.** Statistically different processes between dataset are underlined. Number of Homo Sapiens Reference IDs: 21 804; Mapped IDs in HS10: 87; Mapped IDs in HS20: 84. *Over- (+) or underrepresentation (-)in dataset in comparison to reference.

that the most significant is 'Apoptosis signaling pathway' (HS10: 8.7%, HS20: 8.2%), the following are 'Gonadotropin releasing hormone receptor pathway' (5.8%) and 'Angiogenesis' (4.3%) in HS10, while in HS20 it is 'Angiogenesis' (4.1%) and 'Nicotinic acetylcholine receptor signaling pathway' (4.1%).

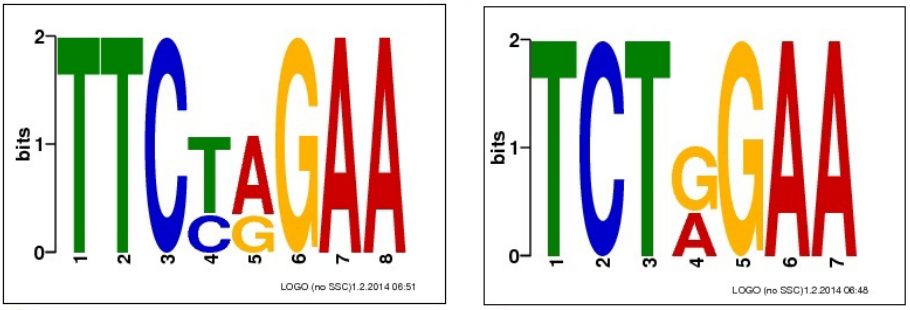
We identified 178 up-regulated genes and 119 down-regulated genes in HS10 with bound HSF1, and 180 up-regulated and 93 down-regulated genes in HS20 with bound HSF1; 140 up-regulated and 46 down-regulated genes were common to both HS10 and HS20 sets. This is consistent with knowledge of HSF1 acting as a protecting agent under heat shock conditions.

To search for HSF1 binding motifs as the first step we have searched for overrepresented motifs in main dataset using MEME-ChIP. Results show slight difference between top ranked motifs, but we observed high resemblance to the commonly known HSF1 binding motif logo, the most relevant motifs are presented in Figure 1. In main dataset corresponding to HS10 next rank, after HSF family motifs, is motif corresponding to ERG transcription factor with JDP2 motif following. In main set corresponding to HS20 similarly the HSF family motif is top ranked as in other dataset, but the following motif belongs to JDP2, no ERG motif was overrepresented in HS20 set.



Proceedings IWBBIO 2014. Granada 7-9 April, 2014 711

**Fig.1.** Overrepresented motifs returned by MEME-ChIP for dataset after 10min heat shock exposure (A) and dataset after 20min.




A) B)

Analysis of HS10 and HS20 sets show more differences between overrepresented motifs. In HS10 top 4 overrepresented motifs are associated with SP1/SP3, NFATC1/MEF2A, GLIS/ZIC3/ and HSF4/HSF3/HSF1. In HS20 we found 3 overrepresented motifs associated with ZIC3/ZIC4, MEF2A/FOXD2 and ZNF75A/SP1. In the next step we have analyzed sequences corresponding to HS10 and HS20 using position 3 weight matrices (PWM) associated with HSF1: the most common matrix [17] and two position weight matrices generated using knowledge based method [18]. In general offset between HS10 and HS20 peaks in range of the same gene was not connected to significant changes in count of computationally found binding sites.

## 4 Conclusion

Study of HSF1 activity in stress has brought many insights into the combinatorial control of transcription factors that operate with HSF1 in a stage and tissue-dependent manner. However some steps of heat shock response control are still unknown and analysis of data from different time exposures brings new insight in HSF1 activity. Comparing results from both experiments we observed differences in group of top activated genes and shifts in HSF1 peaks, what may be a sign of different regulatory strategies in reaction to longer or shorter stress conditions. Change in ranking of genes bound by HSF1 also show activity in different processes, such as preventing apoptosis in HS10 set and high response to stress activity in comparison to HS20 data. Differences in the most active pathways and differentiating processes suggest transition from mild stress condition (10 min exposure) to severe stress condition (20 min exposure), which can occur in borderline temperature of 43$^{o}$C [20]. It is interesting if such strategy is also presented in different tissues, healthy and cancer cells, therefore we would like to attend to this problem when more data will be available. As it is known HSF1 are highly expressed in human cancer cells of various origins [21, 22] where play critical role in proliferation and preservation of cancer. Having that in

mind, it seems that targeting HSF family with knowledge of HSF1 response schemes may help in developing clinical protocols which utilize information about heat shock effects in diseases.

## Footnotes

This work was funded by Polish Ministry of Science grant nr DEC-2012/04/A/ST7/00353 (grant to M.K.)

## References

1. Ciocca, D., and Calderwood SK, *Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications.* . Cell. Stress Chaperones, 2005. **10**(2): p. 86-103.
2. Daugaard, M., Rohde M, Jäättelä M, *The heat shock protein 70 family: highly homologous proteins with overlapping and distinct functions.* FEBS Lett. , 2007. **581**.
3. Fernandes, M., O'Brien T, Lis JT, *Structure and regulation of heat shock gene promoters.* , in *The biology of heat shock proteins and molecular chaperones.* , T.r.A. Morimoto RI, Georgopoulos C , Editor. 1994, Cold Spring Harbor Laboratory Press. : Cold Spring Harbor NY. p. 375–394.
4. Amin, J., Ananthan J, Voellmy R, *Key features of heat shock regulatory elements.* Mol Cell Biol 1988. **8**: p. 3761–69.
5. Gonsalves, S., Moses AM, Razak Z, Robert F, Westwood JT *Whole-Genome Analysis Reveals That Active Heat Shock Factor Binding Sites Are Mostly Associated with Non-Heat Shock Genes in Drosophila melanogaster.* . PLoS Comput Biol, 2011. **6**(1).
6. Pirkkala, L., Nykänen P, Sistonen L., *Roles of the heat shock transcription factors in regulation of the heat shock response and beyond.* . FASEB J. , 2001. **15**(7): p. 1118-31.
7. Kumar, S., Tomar MS, Acharya A, *HSF1-mediated regulation of tumor cell apoptosis: a novel target for cancer therapeutics.* . Future Oncol. , 2013. **9**(10): p. 1573-86.
8. Andrews, S. [cited 2013; Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
9. Langmead, B., Salzberg SL, *Fast gapped-read alignment with Bowtie 2.* Nat. Methods, 2012. **9**(4): p. 357-359.
10. Feng, J., Liu T, Zhang Y, *Using MACS to identify peaks from ChIP-Seq data.* Curr Protoc Bioinformatics, 2011. **2**.
11. Heinz, S., Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK, *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol. Cell, 2010. **38**(4): p. 576–589.
12. Quinlan, A., Hall IM, *BEDTools: a flexible suite of utilities for comparing genomic features* Bioinformatics, 2010. **26**(6): p. 841-842.
13. Mi, H., Muruganujan A, Thomas PD, *PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees.* Nucl. Acids Res., 2012.
14. Machanick, P., and Bailey TL, *MEME-ChIP: motif analysis of large DNA datasets.* Bioinformatics 2011. **27**(12): p. 1696-1697.
15. Gupta, S., Stamatoyannopolous JA, Bailey T, Stafford Noble W, *Quantifying similarity between motifs.* Genome Biology, , 2007. **8**(2): p. R24.

16. Jaksik, R ., R zeszowska-Wolny J , *The distribution of GC nucleotides and regulatory sequence motifs in genes and their adjacent sequences.* *Gene* 2012. **492**(2).
17. Pachkov, M ., E rb I , Molina N , v an N imwegen E , *SwissRegulon: a database of genome-wide annotations of regulatory sites.* Nucl. A cids R es., 2007. **35(Database issue):**: p . D127-31.
18. Pacholczyk, M., Smolińska K, Iwanaszko M , K immel M , *Computational approach for modeling and testing NF-κB binding sites*, in *IWBBIO 2014: INTERNATIONAL WORK-CONFERENCE ON BIOINFORMATICS AND BIOMEDICAL ENGINEERING*. 2 014: Granada.
19. Guertin, M ., L is J T *Chromatin landscape dictates HSF binding to target DNA elements.* PLoS Genetics, 2010. **6**(9).
20. Cates, J ., G raham J C, O mattage N , P avesich E , S etliff I , S haw J, S mith C L, Lipan O , *Sensing the Heat Stress by Mammalian Cells.* BMC Biophysics 2011. **4**(16).
21. Hoang, A., Huang J, Rudra-Ganguly N, et al., *A novel association between the human heat shock transcription factor 1 (HSF1) and prostate adenocarcinoma.* Am J Pathol. , 2 000. **156**: p. 857–864.
22. Ciocca, D., Arrigo AP and Calderwood SK *Heat shock proteins and heat shock factor 1 in carcinogenesis and tumor development: an update.* Arch Toxicol., 2013. **87**: p. 19–48.