TRANSCRIPTIONAL AND GENOME-WIDE EPIGENETIC CHANGES UPON ENVIRONMENTAL STRESS

PhD thesis

Borbála Vető

Molecular Medicine Doctoral School

Semmelweis University

Supervisor:	Tamás Arányi, MD, Ph.D
Official reviewers:	Zsófia Nemoda, MD, Ph.D Máté Varga, Ph.D
Head of the Final Examination Committee:	János M. Réthelyi, MD, Ph.D
Members of the Final Examination Committee:	Krisztina Takács, Ph.D Attila Patócs, MD, Ph.D

Budapest

2018

INTRODUCTION

Eukaryotic gene expression is regulated by transcription. This step is controlled by proteins binding specific gene-regulatory sequences that modify the activity of the RNA polymerase. The tight regulation of gene expression is achieved by the intricate and complex network of these regulatory proteins. Mammalian gene expression is controlled by specific sequences: promoters and enhancers located near or far away from the transcription start site, respectively. The latter includes specific, conserved consensus sequences upstream from the TATA box, such as CCAAT sequences, which bind C/EBP. A multitude of TFs can bind to *cis*-acting promoters and enhancers and they can interact with the RNA polymerase and the proteins of the transcriptional machinery by DNA looping. The basic function of TFs is binding DNA. It is carried out by the DNA-binding domain (DBD). In addition, TFs can be categorized based on their DBD.

One major and evolutionary highly conserved family of transcription factors is encoded by various homeobox genes. Hox genes play major roles in the formation of main body parts during early embryonic development. The PRD class is the second largest of the homeobox gene classes, the name of which is originating from the Drosophila Paired (Prd) gene. A 'new' group of PRD class homeodomains – Argfx, Dprx, Leutx and Tprx – arose from tandem duplication from the Crx (Cone-rod homeobox) gene in a diversity of placental mammals. Furthermore, some were transposed along (Dprx and Leutx) or to another chromosome (Argfx) throughout asymmetric evolution. It has never been proved that these proteins are actual transcription factors localized in the nucleus and they were never endowed with regulatory roles in totipotent embryonic cells.

Members of the NR superfamily regulate transcription in response to binding their lipophilic ligands, which can be hormones (steroid, thyroid hormones), vitamin D, retinoic acid and also fatty acids and phospholipids. NRs comprise a ligand-binding and a DNA-binding domain. NRs also play an important role in responding to various environmental stimuli, thereby controlling development, homeostasis and metabolism, as well. HNF4α is required for liver development and liver cell differentiation. The interaction of HNF4α with different transcription factors, co-activators and co-repressors was discovered. Mass spectrometry analyses of post-translational modifications revealed that HNF4α can be phosphorylated at several sites in the

hepatocellular carcinoma HepG2 cell line. The HNF4α protein was shown to be phosphorylated by different kinase cascades at specific residues.

TFs profoundly influence gene expression. However, the structure of DNA (e.g. DNA loops) and its surrounding genomic and chromatin environment in the nuclei of cells are also fundamental. Gene expression is regulated at two interconnected levels: transcriptional and epigenetic. Epigenetics is the science of the hereditary information located in the nucleus but not encoded in the genome. Covalent, epigenetic modifications (histone modifications, DNA methylation) affect changes in both chromatin structure and gene transcription. DNA methylation is the covalent modification of cytosines by the addition of a methyl group. Methylation of a regulatory region implies gene silencing. DNA methylation is a reversible modification, thus methylation patterns can be established *de novo* and modified genome-wide or at specific loci, reflecting environmental conditions. DNA methylation is implemented by DNMT enzymes. Several anti-cancer drugs have been developed; HDAC inhibitors or DNMT inhibitors, for instance Decitabine.

5-hydroxymethyl-cytosine (5hmC) is an epigenetic modification of cytosines formed by Ten-eleven translocation (TET), thereby taking part in demethylation. It has been reported that 5hmC does not have an obvious inhibitory effect on transcription. Global DNA methylation can be accurately quantified by the highly reliable and simple LC-MS/MS method. TETs are 2-oxoglutarate, oxygen- and iron- dependent demethylases, so α-ketoglutarate, a key player of the citric acid cycle can have epigenetic effect on genome activity and therefore influence health and disease. The activity of chromatin modifications are influenced by key metabolites and cofactors originating from intermediary metabolism, for instance acetyl-CoA, NAD+, S-adenosyl-methionine (SAM) and α-ketoglutarate (α-KG).

In vivo studies demonstrating the dynamic changes of chromatin modifications upon environmental challenges are getting more and more frequent. The importance of the maternal diet (e.g. folate and other methyl donors) during pregnancy has an effect on DNA methylation patterns and the metabolism of the offspring. Epigenetics seems to be the link between early life nutrition with later health. We have seen from the above investigations, that nutritional challenges are often accompanied by general disturbances of covalent chromatin modifications. The changes in TFs and enzymes

related to metabolism are quite well-studied upon fasting for different, but short durations. Fasting induces gluconeogenesis (via the action of PEPCK), glycogenolysis, β-oxidation and ketogenesis. Refeeding immediately blocked hepatic ketogenesis. The pancreatic hormones regulate the response of the liver to metabolic stress. The insulin/glucagon ratio is decreased upon fasting and restored upon refeeding.

Glucose is an essential energy-providing source for most mammalian tissues. Upon fasting, glycogenolysis and gluconeogenesis are activated in order to maintain glucose levels *via* its production and release. The fasting-related TF families are reported in the literature, which are the following: nuclear receptors (e.g. HNF4α, GR, PPARα, RXR), CREB1, FoxOs and CEBPα, β. Essential co-activators of gluconeogenesis are CBP/p300 and PGC-1α. The maintenance of glucose homeostasis upon fasting and feeding (or hormone) is tightly regulated, already at the transcriptional level, for instance through the action of HNF4α. PGC-1α interacts with HNF4α on the response element of PEPCK promoter. During refeeding, insulin represses glucagon secretion in the pancreas and its action in the liver. Altogether, glycogenolysis, gluconeogenesis and fatty acid oxidation are inhibited, whereas glucose degradation, citric acid cycle, fatty acid and triglyceride synthetic pathways are activated. The DNA methylation changes, their role and the mechanism of their appearance have not been studied yet in detail during the fasting-refeeding cycle.

OBJECTIVES

I intended to answer the following questions during my PhD:

I. What is the subcellular localisation and the role during human embryonic development of the group of human PRD-class proteins (Argfx, Dprx, Leutx and Tprx)?

II. Does ERK1 phosphorylate HNF4α? If yes, what is the result of ERK1/2-phosphorylated HNF4α on target gene transcription and target gene DNA-binding?

III. What are the macroscopic, protein level and methylation changes occurring upon short-term fasting and refeeding *in vivo*?

IV. How does treatment with DNA methylation inhibitor or ascorbate affect genomic 5mC and 5hmC levels in cell lines?

METHODS

Cell culture

A375 melanoma, A2058 melanoma, HepG2 hepatocarcinoma, HeLa cervix carcinoma, MES-SA uterine sarcoma, H1650 bronchoalveolar carcinoma and HTR8 placenta cell lines were purchased from ATCC and cultured in Dulbecco's modified Eagle's medium or Advanced MEM or RPMI-1640 supplemented with 10% fetal bovine serum (FBS), 1% glutamine, 1% penicillin/streptomycin. DT40 cells were cultured in RPMI-1640 medium supplemented with 7% FBS, 3% chicken serum, 50 μM β-mercaptoethanol and penicillin/streptomycin.

Treatments

Treatment was performed in serum-free medium for 30 minutes with human recombinant epidermal growth factor at 100 ng/ml final concentration.

Cells were treated with 5-aza-2′-deoxycytidine in DMSO at a final concentration of 1 μg/ml or with vehicle for 48 hours.

Transfection and luciferase experiments

PcDNA5-FRT/TO expression vector containing wild-type, full length human HNF4α gene was obtained from Addgene. Mutations were created for serine or threonine phosphorylation sites in order to have phosphomimetic (glutamate or aspartate) or neutral (alanine) mutations. Gene synthesis and site-directed mutagenesis were performed by GenScript. Triple co-transfection was performed in HeLa cells with the phACCC6(-332/+72)Luc construct containing the *ABCC6* promoter fragment (-

332/+72) in pGL3-Basic vector (Promega)), pcDNA5-FRT/TO plasmid encoding HNF4α mutants (GenScript) and pRL-TK Renilla luciferase Control Reporter Vector. Luciferase activity was determined by luminometric plate reader utilizing the DualGlo Luciferase system (Roche). Results were normalized firstly for the background noise, then for transfection efficiency by the co-transfected Renilla control reporter vector.

ChIP (chromatin immunoprecipitation) assay

1% formaldehyde was added to the culture media of HepG2 cells. After 10 min incubation at room temperature, fixation was stopped by 125 mM ice-cold glycine, then washed with ice-cold phosphate buffered saline (PBS). Cells were harvested and lysed. Sonication was done set for 6 pulses of 15s ON and 30s OFF to obtain 500 bp fragments. Immunoprecipitation was performed with Dynabeads A and G and 2 μg anti-HNF4α mouse monoclonal or H3K27ac antibody was added. Beads were washed and samples were eluted. Reverse crosslinking was performed overnight at 65°C. Samples were treated with RNAse then with proteinase K. DNA was purified using High pure PCR template preparation kit.

ChIP-qPCR

The primers used were designed with BiSearch. QPCR was performed with SYBR green mix. Plates were run in a LightCycler 480 real-time PCR machine. Standard curves were generated from DNA at different dilutions, and their relative amounts were calculated by extrapolating from the dilution curves. Enrichment of a given DNA fragment was calculated by comparing its relative concentration to concentration to the input.

The Reduced Representation Bisulfite Sequencing (RRBS) method

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) from freshly obtained mouse liver. The RRBS kit and protocol is designed by Diagenode. 100 ng genomic double-stranded DNA was digested by MspI enzyme. Unmethylated and methylated spike-in controls were added to the samples. It was followed by adaptor ligation and size selection performed by AMPure XP Beads. The range of the fragments was size selected to be between 200-1200 bp. Sample concentrations were quantified by

quantitative PCR, which permitted them to be pooled together by comparing their relative concentrations to each other. Bisulfite conversion was performed overnight. Amplification was conducted with MethylTaq Plus Polymerase. Samples were sequenced with Illumina HiSeq2000 platform with single end 50 bp reads resulting in 120M reads per pool.

RT-PCR

RNA was isolated using the miRNeasy Micro Kit (Qiagen) from freshly obtained mouse liver. Reverse transcription was performed from 1 μg RNA (RevertAid, ThermoFisher). Published primer sequences were used for expression analysis. QPCR was performed with SYBR green mix in a BioRad CFX96 PCR machine. Standard curves were generated from cDNA at different dilutions, and their relative amounts were calculated by extrapolating from the dilution curves. Enrichment of a given cDNA fragment was calculated by comparing its relative concentration to concentration to the housekeeping 18S RNA.

RESULTS

I. *What is the subcellular localisation and the role during human embryonic development of the group of human PRD-class proteins (Argfx, Dprx, Leutx and Tprx)?*

We have shown the proteins Argfx, Dprx, Leutx and Tprx containing homeodomain localise predominantly to the nucleus in HeLa cells. V5-tagged constructs were also transfected to primary human fibroblasts. Confocal microscopy revealed that the proteins in question are also located in the nucleus. This prominent nuclear localisation is a characteristic of transcription factors. RNA-seq revealed a great number of significantly up- and down-regulated genes. These genes are expressed between the oocyte and blastocyst stages in human embryonic development. Argfx, Leutx and Tprx1 exhibited a sharp transition from low or zero expression until the 4-cell stage to a high expression at 8-cell and morula stages with a steep decline before the blastocyst stage. It

clearly shows that these genes are characterized with a sharp switch-on and off expression pattern and they are expressed immediately before cell fate determination.

II. *Does ERK1 phosphorylate HNF4α? If yes, what is the result of ERK1/2-phosphorylated HNF4α on target gene transcription and target gene DNA-binding?*

We performed *in vitro* phosphorylation assay of ERK1 on HNF4α. We found that ERK1 kinase can phosphorylate the human recombinant HNF4α protein. We intended to identify the phosphorylated serine/threonine residues. Mass spectrometry analysis was performed. We have found numerous phosphorylated amino acid residues. ERK1/2 can indeed phosphorylate HNF4α at a number of previously described sites (S138/T139, S142/S143, S147/S148, S151, T166/S167, S313) and new ones discovered by us (S95, S262/S265, S451, T457/T459). Furthermore, the ERK1 targets the same positions as other kinases, for example PKA, p38 and AMPK. Next, we were interested which phosphorylation site might have an effect on target gene transcription. Mutations were designed for either serine or threonine phosphorylation sites resulting in phosphomimetic (glutamate or aspartate) or neutral (alanine) mutants: S87D, T166A/S167D, S313D, S451E, T457A/T459E and S451E/T457A/T459E triple mutant. Luciferase assay revealed that wild type HNF4α shows similar activity to T166A/S167D, S451E, T457A/T459E and S451E/T457A/T459E. In contrast, both S87D (positive control) and S313D have significant effect on *ABCC6* promoter activity: they inhibit its activity. In summary, both ERK1 and AMPK target the phosphorylation site S313, which has an inhibitory effect on target gene transcription. In order to detect the active HNF4α binding sites at genomic level and select some target loci to examine the effect of ERK1, ChIP was followed by next generation sequencing. The overlap between loci bound by both HNF4α and H3K27ac was remarkable, suggesting that a great number of HNF4α sites are active regulatory regions. Many of the HNF4α TFBSs are located near genes associated with PPAR and insulin signalling, fatty acid metabolism and ABC-transporters. In order to investigate the effect of ERK activation on TF binding, I examined HNF4α binding to several selected genomic target regions upon ERK1/2 induction in HepG2 cells. In conclusion, ERK1/2 phosphorylates HNF4α, resulting in reduced HNF4α DNA-binding capacity to target sequences.

III. *What are the macroscopic, protein level and methylation changes occurring upon short-term fasting and refeeding in vivo?*

Fasting for 8 hours did not, but fasting for 16 and 24 hours lowered body weight in mice compared to their control group. Refeeding for 8 hours could not restore body weight. In conclusion, both fasting and refeeding have a dramatic effect on the body weight of the mice. Furthermore, fasting (8h, 16h, 24h) drastically lowered blood glucose levels. Refeeding for 8 hours daytime could restore blood glucose levels. We were interested in different proteins playing an important role in metabolic adaptation of the liver to acute environmental stress, for instance short-term fasting. Fasting is known to disrupt glucose homeostasis, since HNF4α has a prominent role in glucose metabolism. The protein level of HNF4α does not change significantly upon acute metabolic stress, i.e. fasting. Furthermore, PCK1 is a well-known enzyme and key player in gluconeogenesis. Both CEBPα and PCK1 protein levels were elevated upon fasting. We have investigated the methylation changes occurring upon 16 hours' fasting and 16 hours' fasting followed by 8 hours' refeeding in order to explore the effect of short-term nutritional challenge on global and site/region-specific methylation levels. Altogether, there is no significant difference of average methylation among the groups, but we intended to capture the most important methylation changes. 16 hours' overnight fasting resulted in global hypermethylation, while 16 hours' overnight fasting followed by 8 hours' refeeding lead to global hypomethylation. Hypermethylation occurs mainly outside CpG islands and shores, whereas hypomethylation greatly affects CpG shores. Concerning the distribution of hypomethylation, DNA methylation changes primarily affect promoters compared to the other regions. In addition, the intergenic regions within distal promoters are important targets of DNA methylation change. Next, we were interested which genes are the ones that have a specific CpG changed in one direction upon fasting and this exact CpG changed in the opposite direction upon refeeding. Pathways which are implicated in methylation changes include metabolic pathways, for instance cholesterol, fatty acid, phospholipid, amino acid and carbohydrate (gluconeogenic and glycogen metabolic) metabolic processes. With the DMR analysis we also found several metabolic genes which change their methylation in

their promoter. RT-PCR was also performed to measure the mRNA expression of several metabolic genes, which undergo significant methylation change. Then we examined if the methylation changes observed by the RRBS technique are CpG-specific and enriched at CpG-rich regions or they characterize any Cs in the genome. The LC-MS/MS measurement revealed that acute metabolic stress does not cause quantitative changes in the 5mC and 5hmC levels when the cytosines of the whole genome are investigated.

IV. *How does treatment with DNA methylation inhibitor or ascorbate affect genomic 5mC and 5hmC levels in cell lines?*

We developed a new liquid chromatography mass spectrometry (LC-MS/MS), which can accurately detect genomic 5mC and 5hmC. We explored genomic 5mC and 5hmC levels in different cell types and tissue types. 5mC levels are quite stable among different tissues, but the 5hmC is changeable. Very high 5mC/C ratios were measured in the mouse forebrain. Cell lines had lower 5mC levels. The 5mC level of the sample with the lowest methylation was approximately three times less than that of the highest level. The 5hmC level of the sample with the lowest methylation was approximately hundred times less than that of the highest level. Mouse forebrain showed high 5hmC level. Mouse liver and primary hepatocytes contain similar 5hmC levels. Cell lines had very low 5hmC levels. In conclusion, the newly developed MS technique is sensitive and simple enough to measure even very low levels of 5hmC. 5azadC, or decitabine inhibits DNA methylation. Recent reports have shown paradoxical increase of the 5hmC level upon 5azadC treatment in HL60 human promyeloblast cells. Our aim was to investigate other cell types, as well. Interestingly, the surprising effect of 5azadC on 5hmC levels was only found in hematopoietic cells. Ascorbate is a cofactor regulating DNA methylation and it is required for the full catalytic activity of TET dioxygenases. 5hmC levels were also increased upon 5azadC treatment in HeLa cells, but only upon co-treatment with ascorbate. 5hmC levels were significantly increased upon decitabine and ascorbate treatment in HL60 cells. Our results indicate that some cell lines lack functionally fully active TeT enzymes and possibly are deficient in ascorbate, therefore decitabine treatment does not have an effect on their 5hmC levels.

CONCLUSIONS

In the present PhD dissertation, I intended to shed light on four instances where epigenetics plays an essential role. I have postulated that transcription and epigenetics are closely related. Moreover, environmental stimuli also influence the transcriptional and/or epigenetic status of the cells, organs or organisms.

Firstly, we could see that the homeodomain-containing genes Argfx, Dprx, Leutx and Tprx encode transcription factors having essential roles in animal development, which derives from their highly conserved homeodomain-containing DNA sequences. I have shown that these proteins are located in the nucleus. Our experiments have revealed that these TFs are responsible for driving totipotency to cell fate decisions. More importantly, a downstream effector has been found, which is the HIST1H2BD histone H2 variant. Therefore, it can be postulated that developmental changes are accompanied by changes in both epigenetics and transcription. Indeed, these two processes depend on each other, where most probably epigenetic marks regulate gene expression.

Secondly, we have investigated the transcriptional role of HNF4α - required for hepatocyte differentiation and development in the liver - under physiological conditions and upon extracellular stimuli and its concomitant epigenetic changes. We have revealed that ERK1 can phosphorylate HNF4α. Moreover, I have shown that ERK activation results in lowered DNA binding capacity of the TF to target hepatic DNA sequences in HepG2 cells. In our ChIP-Seq experiments, 8748 transcription factor binding sites (TFBSs) could be identified for HNF4α. We have also detected actively transcribed genes regulated by HNF4α; they were also marked by the epigenetic modification H3K27ac. The overlap between loci bound by HNF4α and H3K27ac was remarkable, suggesting that a great number of HNF4α sites are active regulatory regions. The HNF4α-target regions, which were occupied by expansive H3K27ac signal, showed diminished HNF4α-binding in ChIP-qPCR experiments suggesting that ERK1 activation might not only change TF-binding, but epigenetic mechanisms might also be involved.

Thirdly, we were interested if nutritional stress (short-term fasting and refeeding) influences the methylation status of hepatocyte DNA *in vivo*. We have observed enormous DNA methylation changes, fasting lead to global CpG-hypomethylation,

whereas refeeding resulted in global CpG-hypermethylation. These changes affected several metabolic pathways, for instance cholesterol, fatty acid, phospholipid, amino acid and carbohydrate (gluconeogenic and glycogen metabolic) metabolic processes. It is clear that CpG methylation influences gene expression. We investigated these methylation changes with several methods, and we could couple them with changes in blood glucose, body weight, protein and mRNA expression. It seems that rapid and vast DNA methylation changes occur due to its reversibility and its responsiveness to environmental conditions.

Last but not least, we have also seen that despite DNA methylation changes at CpG islands, genome-wide DNA methylation of cytosines is not affected when animal livers undergo nutritional stress conditions. It seems that the mouse genome is protected from vast methylation changes which might be too drastic. However, transcription and gene expression seem to be altered and the underlying mechanisms are partly epigenetic. We have set up a method to measure 5-methylcytosine and 5-hydroxymethylcytosine levels in the genome by LC-MS/MS. We demonstrated in several experiments that DNA methylation inhibition leads to the elevation of 5hmC levels in hematopoietic cell lines. Moreover, adding ascorbate – a cofactor of the TeT enzymes – further enhances the formation of 5hmC in hematopoietic cell lines.

In conclusion, all the four projects have clear epigenetic implications and raise very interesting further questions which would be worth investigating.

BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

Publications related to the thesis

1. Vető B, Szabo P, Bacquet C, Apro A, Hathy E, Kiss J, Rethelyi JM, Szeri F, Szuts D, Aranyi T. (2018) Inhibition of DNA methyltransferase leads to increased genomic 5-hydroxymethylcytosine levels in hematopoietic cells. FEBS Open Bio, 8(4): 584-592. IF: 2.143

2. Maeso I, Dunwell TL, Wyatt CD, Marletaz F, Vető B, Bernal JA, Quah S, Irimia M, Holland PW. (2016) Evolutionary origin and functional divergence of totipotent cell

homeobox genes in eutherian mammals. BMC Biol, 14: 45. IF: 6.779

3. Vető B, Bojcsuk D, Bacquet C, Kiss J, Sipeki S, Martin L, Buday L, Balint BL, Aranyi T. (2017) The transcriptional activity of hepatocyte nuclear factor 4 alpha is inhibited via phosphorylation by ERK1/2. PLoS One, 12(2): e0172020. IF: 2.806

Publications not directly related to the thesis

1. Pagliaroli L, Vető B, Aranyi T, Barta C. (2016) From Genetics to Epigenetics: New Perspectives in Tourette Syndrome Research. Front Neurosci, 10: 277. IF: 3.556