



Choosing the target loci:

Heat shock factors HSF1 and HSF2 as regulators of  
cell stress and development

Anniina Vihervaara

Department of Biosciences, Åbo Akademi University  
Turku Centre for Biotechnology, University of Turku and Åbo Akademi University  
Turku Doctoral Program of Biomedical Sciences  
Finland  
2014

## ABBREVIATIONS

|               |   |
|---------------|---|
| ACRV1         | Acrosomal vesicle protein 1   |
| AD            | Activation domain   |
| AHA1/AHSA1    | Activator of heat shock 90kDa protein ATPase                        |
| ARHGEF        | Rho guanine nucleotide exchange factor                              |
| BANF1/BAF1    | Barrier to autointegration factor 1                                 |
| bp            | Base pair   |
| CCT           | Chaperonin containing TCP   |
| CDC37         | Cell division cycle 37  |
| CDK           | Cyclin-dependent kinase   |
| ChIP          | Chromatin immunoprecipitation                                       |
| ChIP-chip     | Chromatin immunoprecipitation on microarray                         |
| ChIP-seq      | Chromatin immunoprecipitation coupled to high-throughput sequencing |
| CHRNA         | Cholinergic receptor, nicotinic, alpha                              |
| CTCF          | CCCTC-binding factor  |
| CTD           | Carboxyterminal domain of RNA polymerase II                         |
| DARS          | Aspartyl-tRNA synthetase  |
| DBD           | DNA-binding domain  |
| DNAJ          | J-domain containing chaperone                                       |
| DTT           | Dithiotreitol   |
| DZ            | Dark zone of mouse seminiferous tubule                              |
| eEF           | Eukaryote elongation factor   |
| ENCODE        | Encyclopedia of DNA elements  |
| FTMT          | Mitochondrial ferritin  |
| GAF           | GAGA-binding factor   |
| GAPDH         | Glyceraldehyde-3-phosphate dehydrogenase                            |
| GBA           | Glucosidase beta, acid  |
| GEO           | Gene expression omnibus   |
| $\gamma$ H2AX | Histone 2A X, phosphorylated at serine (S) 139                      |
| HAT           | Histone acetyl transferase  |
| HDAC          | Histone deacetylase   |
| HOX           | Homeobox protein  |
| HR-A/B/C      | Heptad repeat A/B/C   |
| HSE           | Heat shock element  |
| HSF           | Heat shock factor   |
| HSP           | Heat shock protein  |
| HSR           | Heat shock response   |
| IL            | Interleukin   |
| KCNN1         | Calcium-activated channel N1  |
| LDLR          | Low density lipoprotein receptor                                    |
| LMNB          | Lamin, beta   |
| lncRNA/LINC   | Long non-coding RNA   |
| ME1           | Malic enzyme 1, NADP(+)-dependent, cytosolic                        |
| miRNA         | MicroRNA  |

## *Abbreviations*

---

|               |  |
|---------------|--|
| MLL/KMT2A     | Myeloid/lymphoid or mixed-lineage leukemia/Lysine (K)-specific methyl-transferase 2A |
| mRNA          | Messenger RNA  |
| MSCI          | Meiotic sex-chromosome inactivation  |
| MRPS          | Mitochondrial ribosomal protein  |
| NEAT1         | Nuclear paraspeckle assembly transcript 1  |
| NLS           | Nuclear localization signal  |
| NUDC          | Nuclear distribution C homolog   |
| PAR           | Pseudoautosomal region   |
| PcG           | Polycomb group protein   |
| PDSM          | Phosphorylation-dependent sumoylation motif  |
| PFN3          | Profilin 3   |
| PIC           | Preinitiation complex  |
| PRKC/PKC      | Protein kinase C   |
| PRM           | Protamine  |
| PTGES3/p23    | Prostaglandin E synthase   |
| PTM           | Post-translational modification  |
| PZ            | Pale zone of mouse seminiferous tubule   |
| RD            | Regulatory domain  |
| RNAi          | RNA interference   |
| RNP           | RNA polymerase   |
| rRNA          | Ribosomal RNA  |
| shRNA         | Short hairpin RNA  |
| SLX           | Sycp3 like X-linked  |
| SLY           | Sycp3 like Y-linked  |
| SNAP29        | Synaptonemal-associated protein 29 kDa   |
| SPEER         | Spermatogenesis associated glutamate (E)-rich protein                                |
| SPT           | Suppressor of Ty   |
| SRSY          | Serine-rich secreted, Y-linked   |
| SRY           | Sex determining region Y   |
| SS            | Strong spot of mouse seminiferous tubule   |
| SSTY          | Spermiogenesis-specific transcript on the Y  |
| SWI/SNF       | Switch/sucrose nonfermenting   |
| TAF           | TBP-associated factor  |
| TBP           | TATA-box binding protein   |
| TF            | General transcription factor   |
| TNP           | Transition protein   |
| tRNA          | Transfer RNA   |
| trxG          | Trithorax protein  |
| TSS           | Transcriptional start site   |
| TUBE          | Tandem ubiquitin binding entity  |
| Ub            | Ubiquitin  |
| WS            | Weak spot of mouse seminiferous tubule   |
| Wt            | Wild type  |
| Yp            | Short arm of the mouse Y chromosome  |
| Yq            | Long arm of the mouse Y chromosome   |
| ZFAND2A/AIRAP | Zinc finger AN1-type domain 2A/Arsenite-induced RNA associated                       |

## ABSTRACT

Heat shock factors (HSFs) are an evolutionarily well conserved family of transcription factors that coordinate stress-induced gene expression and direct versatile physiological processes in eukaryote organisms. The essentiality of HSFs for cellular homeostasis has been well demonstrated, mainly through HSF1-induced transcription of heat shock protein (HSP) genes. HSFs are important regulators of many fundamental processes such as gametogenesis, metabolic control and aging, and are involved in pathological conditions including cancer progression and neurodegenerative diseases. In each of the HSF-mediated processes, however, the detailed mechanisms of HSF family members and their complete set of target genes have remained unknown. Recently, rapid advances in chromatin studies have enabled genome-wide characterization of protein binding sites in a high resolution and in an unbiased manner. In this PhD thesis, these novel methods that base on chromatin immunoprecipitation (ChIP) are utilized and the genome-wide target loci for HSF1 and HSF2 are identified in cellular stress responses and in developmental processes. The thesis and its original publications characterize the individual and shared target genes of HSF1 and HSF2, describe HSF1 as a potent transactivator, and discover HSF2 as an epigenetic regulator that coordinates gene expression throughout the cell cycle progression. In male gametogenesis, novel physiological functions for HSF1 and HSF2 are revealed and HSFs are demonstrated to control the expression of X- and Y-chromosomal multicopy genes in a silenced chromatin environment. In stressed human cells, HSF1 and HSF2 are shown to coordinate the expression of a wide variety of genes including genes for chaperone machinery, ubiquitin, regulators of cell cycle progression and signaling. These results highlight the importance of cell type and cell cycle phase in transcriptional responses, reveal the myriad of processes that are adjusted in a stressed cell and describe novel mechanisms that maintain transcriptional memory in mitotic cell division.

# 1 INTRODUCTION

The cell is the smallest entity of life and the structural and functional foundation of all organisms. To understand nature and the mechanisms of life, a comprehensive appreciation of cellular processes and cellular networks in complex organisms is required. Inside each cell, hereditary information carries the instructions for differentiation, cell type-specific functions and division of labor between tissues. Coordinated execution of cellular processes involves sensing external and internal conditions, organizing signaling cascades and changing the cell's structure or behavior. Signals that reach the genome adjust the expression of genes, the instruction entities for synthesis of cellular components. Consequently, the gene expression programs in individual cells determine the molecular constituents and the possibilities for cellular responses, making the regulation of gene expression one of the most fundamental processes in all living organisms. In this PhD thesis, the coordination of gene expression is investigated in development and in response to acute, protein-damaging stress.

Heat shock factors (HSFs) are an evolutionarily well conserved protein family that coordinates gene expression in a variety of physiological processes. HSFs are best characterized as rapid activators of gene expression upon protein-damaging stress when the overall gene activity in the cell is silenced. Beyond stress, HSFs are crucial regulators of developmental processes and aging, and involved in several pathological conditions such as neurodegenerative diseases and cancer. In this thesis, I have investigated the versatile roles of HSF1 and HSF2 in development and in cellular stress responses, addressing how HSFs interact with the dynamic chromatin environment in different cell types and cellular conditions. As a model system for development, I have used male gametogenesis which consists of strikingly complex and well coordinated changes in the chromatin landscape. The developing gametes undergo clonal expansion via mitosis, reorganization and reduction division of chromosomes during meiosis and a profound morphological differentiation during the haploid phase of spermatogenesis. The rapidly provoked gene expression in response to stress provides a model system where the molecular mechanisms of HSF-mediated transcriptional activation and the cellular processes that maintain homeostasis can be studied. To elucidate how HSF1 and HSF2 interact with chromatin in distinct states, I have investigated the transcriptional programming and the cellular survival mechanisms in freely cycling cells and in cells that undergo mitotic division. In each of these studies, the genome-wide target sites for HSF1 and HSF2 have been characterized using advanced techniques that base on chromatin immunoprecipitation (ChIP). The genome-wide analyses have been elaborated with computational data mining, biochemical characterization of gene-specific regulatory mechanisms for HSF1 and HSF2 and by investigating the biological significance of the transcriptional reprogramming.

The results presented in this thesis and its original publications reveal the importance of the cell type, developmental state and the chromatin environment for transcriptional responses. In male gametogenesis, HSFs are uncovered to control the X- and Y-chromosomal multicopy genes, which are crucial for chromatin compaction in the sperm head and for correct sperm morphology. HSF1 is shown to localize to silenced X- and Y-chromatin in pre- and post-meiotic germ cells whereas HSF2 resides at the dividing genome in meiosis I and II. To date, HSFs remain the only

transcription factors that have been shown to occupy the meiotic sex-chromatin and to regulate the expression of the sex-linked multicopy genes. In cellular stress responses, HSFs are revealed to coordinate the expression of whole chaperone machinery, including protein foldases, disaggregases, inhibitors of aggregate formation and cochaperones. Furthermore, HSFs induce the expression of ubiquitin and regulate the expression of translational components, mediators of cell cycle progression, metabolic processes and signaling cascades. The highly intertwined functions of HSF1 and HSF2 are contrasted with their profoundly distinct mechanisms on the genome, particularly in dividing cells. While HSF1 is efficiently excluded from the mitotic and meiotic chomatin, HSF2 avidly interacts with the condensed genome. HSF2, however, does not compensate for the lack of HSF1 at the stress responsive genes, but instead, is involved in reactivationont of transcription in post-mitotic cells.

## 2 REVIEW OF THE LITERATURE

### 2.1 Transcriptional regulation of cellular functions

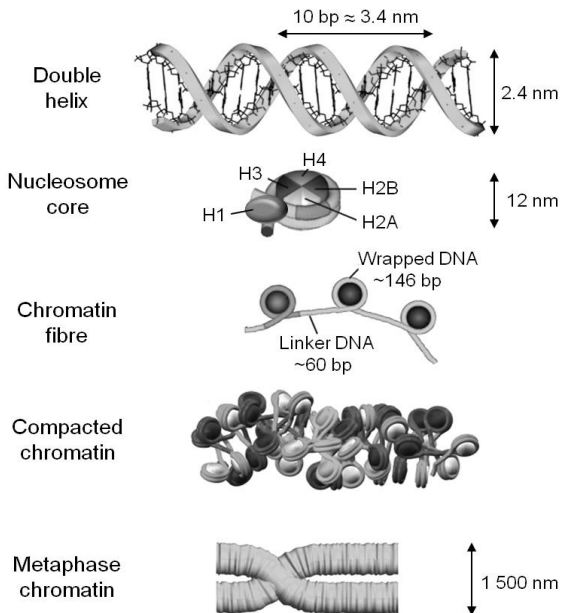
Faithful implementation and propagation of hereditary information is a prerequisite for the execution of biological processes and development of eukaryotic organisms. How the genome is utilized in individual cells dictates the synthesis of cellular components and coordinates differentiation, cell type-specific functions and cellular responses. In a process called transcription, the genetic information is used as a template to build structural and functional molecules of the cell. The delicate coordination of transcriptional programs is essential for physiological processes and is coordinated by regulatory factors that organize the genetic information and direct the gene expression in response to internal and external stimuli.

#### 2.1.1 Chromatin structure and dynamics

The hereditary information is encoded in the sequence of complementary deoxyribose nucleic acid (DNA) polymers that form a double helix (Figure 1). In eukaryotes, the DNA is organized inside a membrane enclosed nuclear compartment and associated with proteins to form a structure called chromatin (reviewed in Felsenfeld and Groudine, 2003; Schlick *et al.*, 2012). The basal constituents of the chromatin are nucleosomes in which 146 base pairs (bps) of DNA encircle a histone octamer. The octamer, in turn, is composed of two sets of histones H2A, H2B, H3 and H4, and the nucleosome structure is stabilized by histone H1 that contacts the DNA at the site where it enters and exits the histone core (Figure 1). The nucleosomes are connected with linker DNA sequences and the chromatin fiber is further organized by scaffold proteins into a higher-order structure (Figure 1; reviewed in Woodcock and Ghosh, 2010).

Organization of genetic material into chromatin enables efficient condensation of the DNA, but also dynamic regulation of the accessibility of distinct genomic regions to transcriptional regulators, replication factors and DNA repair machinery. As a result of different grades of packaging and associating factors, the molecular composition of the chromatin fiber is highly diverse along the length of the chromosomes, making chromosomes among the most complex entities in the cell. The first layer of chromatin condensation is conducted by the histone molecules, which can occur in different variants and undergo post-translational modifications (PTMs) (reviewed in Zentner and Henikoff, 2013). The histone proteins associate with each other via globular core domains, whereas their highly dynamic tails can undergo a wide range of modifications including acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ribosylation. The combination of the histone modifications affects the compactness of the chromatin and profoundly contributes to the accessibility of the underlying DNA (reviewed in Felsenfeld and Groudine, 2003; Zentner and Henikoff, 2013). The histone modifications are catalyzed by enzymes such as acetyl transferases (HATs), deacetylases (HDACs), methyl transferases, demethylases, ligases and proteases (Brownell and Allis, 1996; Peterson and Laniel, 2004; Shilatifard, 2006). The chromatin is, furthermore, targeted by remodeling factors and architectural proteins that can change the position of histones or organize the genome into a more

or less compacted higher-order structure (reviewed in Woodcock and Ghosh, 2010). In general, chromatin regions that are actively encoded are kept within an open conformation (euchromatin) and are marked by hyperacetylation of histones H3 and H4, as well as by trimethylation of histone H3 at lysine (K) 4 (H3K4me3). Instead, silent chromatin regions reside in a closed state (heterochromatin) and are typically characterized by H3K9me3 and H3K27me3, and heterochromatin-associated proteins (Zentner and Henikoff, 2013). In non-dividing cells, the DNA is relatively loosely packed and display region-specific patterning of condensation. However, at every cell division, the chromatin undergoes profound condensation as the nuclear membrane breaks down and the duplicated metaphase chromosomes are separated into the daughter cells (Figure 1).



**Figure 1. Chromatin structure.** DNA double helix is wrapped around a histone octamer and the chromatin fibre is further organized into a higher-order structure. Adapted from Schlick *et al.*, *J. Biol. Chem.* 2012, reprinted with permission from ASBMB.

### 2.1.2 Organization of the genetic information

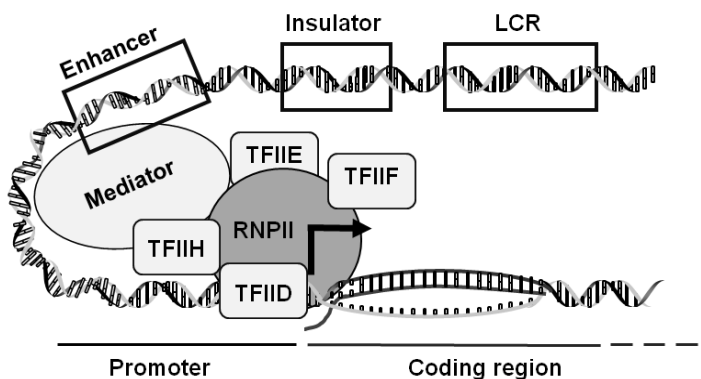
Information in the DNA is transcribed into single-stranded sequences of ribonucleic acid (RNA) polymers. Several RNA species exist in the cells, carrying versatile functions. Some of the RNA molecules contain information that is translated into amino acid sequence of proteins, and together, the RNA and protein molecules synthesize other cellular components such as lipids and define the structural and functional constituents of the cell. Traditionally, the sequences of DNA that code for proteins are termed genes, but genome-wide analyses have revealed the multitude of genomic regions that encode RNA as an end product (reviewed by Pennisi, 2012; Qu and Fang, 2013). In this thesis, the term “gene” refers to a DNA sequence that encodes a transcript. When specification is needed, “protein-coding” and “RNA-coding” define the end product of the gene. In human, the 20 687 protein-coding genes identified to date are sparsely distributed along the chromosomes and, due to the many non-coding sequences inside and between the genes, less than 3% of the human genome codes for amino acid sequences (Encyclopedia of DNA Elements, ENCODE, Consortium, 2013). The RNA-coding genes include a variety of small RNA species as well as 18 400 genes for RNA with over 200 bp. Consequently, recent estimates suggest that the majority of the human genome (~76%) is transcribed (The ENCODE Consortium, 2013; reviewed by Pennisi, 2012; Qu and Fang, 2013). A major challenge, however, for interpretation of the current genome-wide analyses is the presence of a number of repetitive sequences in the human genome, which routinely are neglected from the sequence-based analyses (de Koning *et*



*et al.*, 2011; reviewed in Treangen and Salzberg, 2011). It is also worth noting that the genome is complex and certain loci encode several transcripts. Moreover, an RNA molecule might carry a regulatory or structural function besides serving as a template for an amino acid sequence (Salmena *et al.*, 2011; Taulli *et al.*, 2013).

The genome contains a myriad of non-coding sequences including regulatory elements that contribute to the coordinated expression of the genes (Figure 2; reviewed by Dekker *et al.*, 2013; Maston *et al.*, 2006; Riethoven, 2010). Core and control promoters locate to the vicinity of transcriptional start site (TSS) and serve as an assembly platform for the protein complexes that synthesize RNA. Enhancers, silencers and insulators are so called distal regulatory elements that can occur either upstream or downstream of the gene and regulate the gene activity also from a great distance (Figure 2; reviewed by Maston *et al.*, 2006; Riethoven, 2010). Both enhancers and silencers contain binding sites for activatory and inhibitory factors, and typically organize the spatio-temporal gene expression in different tissues or in response to distinct stimuli. While insulators are able to confine the gene activity and chromatin landscape to a given region, locus control elements (LCRs) contribute to the coordinated activity of an entire locus or a gene cluster (Maston *et al.*, 2006; Riethoven, 2010). The genome also contains information for its three dimensional organization (reviewed by Dekker *et al.*, 2013; Gibcus and Dekker, 2013). For example, certain regions are targeted by lamins, which are the primary protein constituents of the nuclear envelope and can position genomic regions to the vicinity of the nuclear membrane (reviewed in Andrés and González, 2009). Importantly, the lamins are in direct contact with nuclear actin (Holaska *et al.*, 2004; Simon *et al.*, 2010) which has been coupled to gene expression, chromatin remodeling and processing of transcripts (reviewed in Akhtar and Gasser, 2007; Grosse and Vartiainen, 2013; de Lanerolle and Serebryanny, 2011). A line of evidence also points to formation of transcription and replication factories, as several genomic regions come together to utilize the same machinery for a cellular process (Iborra *et al.*, 1996; Mitchell and Fraser, 2008; Osborne *et al.*, 2004; reviewed in Cook, 1999; Pope *et al.*, 2013; Sutherland and Bickmore, 2009). All-in-all, the organization of the genome in space and time is a multi-dimensional task where a number of *cis*-acting DNA sequence elements and *trans*-acting factors interact for coordinated implementation of the hereditary information.

**Figure 2. Control regions at an RNA polymerase II (RNPII) transcribed gene.** General transcription factors (TFIIs) recognize DNA elements at the promoter and direct the binding and assembly of RNPII at the transcriptional start site (arrow). Transcriptional regulators, including the mediator complex, relay signals from distal control regions such as enhancers, insulators and locus control elements (LCRs) to the transcriptional machinery. Transcription is also controlled by site-specific transcription factors, cofactors and chromatin modifying enzymes (not shown).



### 2.1.3 Initiation of transcription

Expression of RNA and protein is regulated at multiple steps including transcriptional initiation, elongation and termination, as well as the stability, processing and localization of the produced RNA and protein. Since correct assembly of the transcriptional apparatus at the right genes at the right time is a prerequisite for correct gene expression, the initiation of transcription is a key regulatory step in the coordination of genome-wide transcriptional programming. Gene promoters harbor well conserved DNA elements such as TATA-box, GC-box, BRE-element or initiator that are recognized by general transcription factors (TFs) (reviewed in Kadonaga, 2012). The transcriptional activation is initiated by an ordered assembly of TFs at the promoter to form a preinitiation complex (PIC). PIC, in turn, directs the binding and correct positioning of RNA polymerase (RNP), the enzyme that catalyzes the synthesis of RNA using DNA as a template (Figure 2; reviewed in Cramer *et al.*, 2008; Maston *et al.*, 2006). In eukaryotes, three RNPs (RNPI-III) exist, each of which associates with a distinct set of TFs (TFI-III, respectively) and encodes a specified set of genes (reviewed in Hamperl *et al.*, 2013; Vannini and Cramer, 2012). RNPI catalyzes the synthesis of 45S ribosomal RNA (rRNA) that is processed into 5.8S, 18S and 28S structural elements of protein translating ribosomes. Instead, RNPIII catalyzes 5S rRNA and all the transfer RNA (tRNA) species that interact with ribosomes and recruit amino acids to the growing polypeptide chain. Transcription that is mediated by RNPI and RNPIII localizes to nucleoli, which are subnuclear compartments that contain clustered ribosomal gene copies and, therefore, are prominent transcription-organizing structures in the eukaryote cell (reviewed in Gibcus and Dekker, 2013). The best studied RNP in eukaryotes is, however, RNPII which synthesizes all the protein-coding genes, as well as most microRNA (miRNA) and long non-coding RNA (lncRNA) species.

After the assembly of PIC and recruitment of RNP, several signals confer to the release of the RNP to its elongation mode. At certain gene promoters, the PIC and RNPII are assembled and disassembled in a cyclinc manner (reviewed in Metiviér *et al.*, 2006). However, approximately 30% of human gene promoters harbor RNPII that is engaged in transcription but kept paused by negative elongation factors (reviewed in Adelman and Lis, 2012). Indeed, the release of paused RNPII to elongation has emerged as an efficient means to coordinate gene expression programs, particularly during development and in response to activating signals. Efficient transcription also involves melting of the DNA strands and clearance of the gene body from obstructing proteins and DNA coils (reviewed in Feklistov, 2013; Fuda *et al.*, 2009; Selth *et al.*, 2010). The recruitment and assembly of PIC and RNP, as well as the following steps of promoter escape, elongation, termination, and reinitiation depend on a synergistic action of transcriptional regulators and chromatin modifying enzymes.

### 2.1.4 Transcription factors

The assembly of PIC and RNP is directed by sequence-specific transcriptional regulators called transcription factors. Transcription factors are characterized by a DNA-binding domain (DBD) that recognizes short, generally in the range of 6-12 bp, DNA sequences (reviewed in Maston *et*

*al.*, 2006; Pabo and Sauer, 1992). Many transcription factors form homo- or hetero-oligomers, which is reflected in the consensus DNA-binding element, often composed of two half-sites. A transcription factor can recognize several variants of its consensus DNA element, but the precise sequence can dictate the regulatory impact, for example, by directing the oligomerization partner or by conferring affinity advantage for certain loci over the others (reviewed in Maston *et al.*, 2006). Transcriptional regulators act in a complex chromatin environment where, besides the underlying *cis*-elements, also the composition of other components at the target loci affects the binding ability and the regulatory output. Most transcription factors can directly or indirectly regulate the assembly of PIC, promote the escape of RNP, or recruit chromatin modifying enzymes that either enhance or inhibit the steps of gene expression. Transcriptional activation can also be influenced by cofactors which typically do not bind to DNA, but instead, target transcriptional regulators and modulate their transactivating capacity (reviewed in Roeder, 2005). Thus, the transcriptional control engages synergistic action from several transcriptional regulators and chromatin modifying enzymes that integrate the cellular and physiological signals to a coordinated behavior of cells and organisms.

#### 2.1.5 Transcription elongation, termination and reinitiation

Most of our knowledge on regulation of gene expression originates from RNPII-mediated transcription of protein-coding genes. For RNPII, the promoter escape is mediated by releasing inhibitory proteins, such as negative elongation factor (NELF), and by recruitment of activating factors such as positive transcription elongation factor b (P-TEFb). The P-TEFb complex mediates phosphorylation of serine (S) 2 at the heptad repeat of C-terminal domain (CTD) of RNPII, enabling RNPII to enter elongation. Transcriptional elongation is facilitated by chromatin remodeling factors that clear the gene body from obstructing proteins and by topoisomerases that cut and paste DNA strands to relieve coiling (reviewed in Feklistov, 2013; Fuda *et al.*, 2009; Selth *et al.*, 2010). At the 3' end of the gene, RNPII meets a polyA site and the transcription terminates. In many cases the components of RNP complex are recycled to the promoter for reinitiation of transcription (reviewed in Gilmour and Fan, 2008; Richard and Manley, 2009).

#### 2.1.6 Post-transcriptional processing of RNA and protein

Protein-coding genes contain exons that code for amino acid sequences and introns that do not code for proteins. Initially, the gene is transcribed as one unit termed pre-messenger RNA (pre-mRNA), but during a process called splicing, the introns are removed and a defined set of exons are united to form a mature mRNA (reviewed by Darnell, 2013; Kornblihtt, 2007). Splicing is an important step in controlling the transcript variant and provides a mechanism for producing several protein isoforms from a single gene. The processing of mRNA includes also methylation of the 5' end, as well as polyadenylation of the 3' end. The 3' polyA tail is bound by proteins that transport the mRNA to the cytosol, whereas the 5' region interacts with ribosomes to initiate translation (reviewed by Darnell, 2013; Kornblihtt, 2007).

The nucleotide sequence of an mRNA is translated into a sequence of amino acids that defines the structure and function of a protein. During translation, ribosomes position tRNA molecules to recognize nucleotide triplets, so called codons, of the mRNA and to transfer a corresponding amino acid to the growing polypeptide chain (reviewed by Schmeing and Ramakrishnan, 2009). Translation is initiated in the cytoplasm but signal sequences emerging at the polypeptide chain can direct the ribosome to a specific cellular location, e.g. to the membrane of endoplasmic reticulum (ER) where all the secreted and membrane-associated proteins are synthesized.

Proteins are versatile molecules whose proper biological functions depend on correct three-dimensional shape and conformational flexibility. While small proteins that comprise only one functional domain fold efficiently also *in vitro*, large proteins in the context of the crowded cellular environment require molecular chaperones for an efficient assembly. Besides assisting in *de novo* folding of newly synthesized proteins, molecular chaperones provide constant surveillance of the proteome and are integrated into the networks that coordinate protein-protein-interactions, localization, stability and degradation (reviewed in Hartl *et al.*, 2011).

## 2.2 Transcriptional regulation of cellular differentiation

Every mitotic cell division leads to symmetric division of the hereditary information. As a consequence, virtually every cell in an organism contains the same genetic information, which incites a question on how the distinct cell types and organs emerge? Despite the symmetrical distribution of the replicated genome, all the material in the cell is not equally divided between the daughter cells. For example at fertilization, the site of sperm entry and the position of polar bodies of the egg define an equatorial line for cleavage and initiate redistribution of maternal RNA and protein (Edwards and Beard, 1997; Kumano, 2012; Piotrowska and Zernicka-Goetz, 2002; Roegiers and Jan, 2004). This cellular polarization determines the dorso-ventral axis of the developing zygote at the very first division. Besides gaining a different set of regulatory molecules, the cells are directed by signals from their surroundings. Particularly during early development, positive and negative regulatory networks culminate on the genome and cause different transcriptional programs that direct the neighboring cells towards distinct lineages. Among the best studied examples of early transcriptional regulators are the homeobox (HOX) proteins, which activate and repress gene groups that coordinate the segmentation of the body (reviewed in Pearson *et al.*, 2005). The cell-specific expression patterns of *HOX* genes, in turn, are regulated by trithorax-group (trxG) proteins that maintain genes in an active state and by polycomb-group (PcG) proteins that can repress gene activity over many cell generations (Schuettengruber *et al.*, 2007).

### 2.2.1 Differentiation and commitment

Differentiation towards a committed cell type is accompanied with a genome-wide patterning of the chromatin. This epigenetic chromatin state is maintained over mitotic division and passed on to the daughter cells (reviewed in Delcuve *et al.*, 2008; Probst *et al.*, 2009). In pluripotent stem

cells, the loosely packed chromatin enables plasticity of gene expression and differentiation towards versatility of cell types (Fussner *et al.*, 2011). The lineage-specific genes are kept in a silent but transcriptionally available (poised) state, which is characterized by repressive H3K27me3 and activating H3K4me3 histone modifications. This bivalent chromatin enables rapid transcription when the cell becomes committed for differentiation (Bernstein *et al.*, 2006; reviewed in Tollervey and Lunyak, 2012).

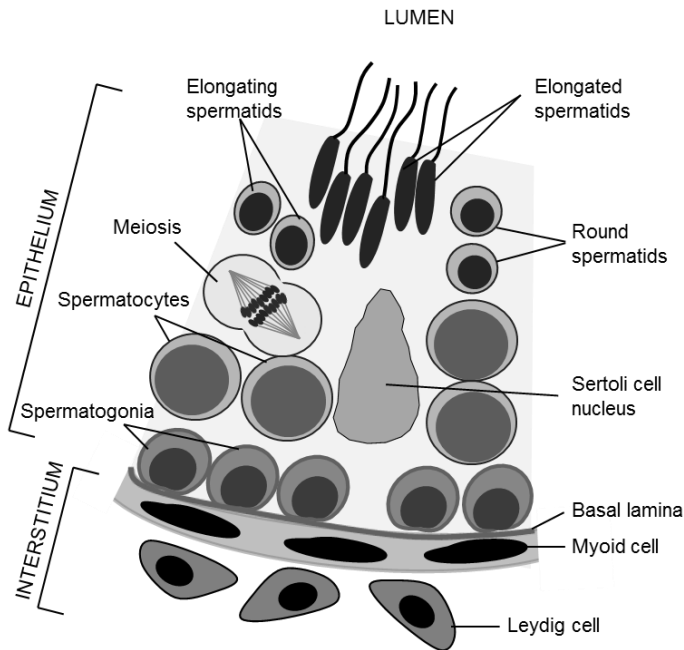
Majority of chromatin modifications in differentiating cells, however, is involved in silencing and compacting chromatin regions that are not utilized by the terminally differentiated cells. For example,  $\beta$ -globin is in an open state only in cells of erythrocyte lineage where its abundant expression generates components for oxygen transport (Reddy *et al.*, 1994). Especially in complex organisms, the billions of nucleotides in each cell create a daunting task for transcriptional regulators to localize a couple of nucleotides long DNA elements. Over the course of evolution, the expansion in the genome size has been accompanied with evolvement of mechanisms for chromatin compaction (reviewed in Mohn and Schübeler, 2009). Histones emerged in early eukaryote development, and through the versatile PTMs, create a specific code that directs transcriptional regulators (reviewed in Zentner and Henikoff, 2013). Moreover, the PcG proteins have undergone high diversification in multicellular organisms (Schuettengruber *et al.*, 2007; Whitcomb *et al.*, 2007) and, in vertebrates, genome-wide DNA methylation efficiently compacts and represses chromatin regions (Guibert *et al.*, 2009; Mohn and Schübeler, 2009). Although the epigenetic status of the chromatin can be modified even in terminally differentiated cells, the distinct features of the genomic regions define the chromatin landscape for transcriptional responses and direct the ability of cells to respond to the guidance cues. As a result, different cell types have profoundly distinct appearances, express different sets of molecules and carry a wide versatility of functions.

During the life cycle of a mammal, a global resetting of the epigenome occurs during gametogenesis and early embryogenesis. This reacquisition of totipotency is crucial for the ability of germ cells to develop into mature gametes, and for the fertilized egg to give rise to all the cell types of the new individual (reviewed in Cantone and Fisher, 2013). In mouse, the global resetting of the epigenome is initiated at E7.5-12.5 when the primordial germ cells migrate to the genital ridges (Cantone and Fisher, 2013). This epigenetic restoration involves incorporation of histone variants, RNA-mediated silencing of repetitive DNA elements, and establishment of germ line-specific pattern of DNA methylation (Hajkova *et al.*, 2008; 2010). After reaching the genital ridges, the germ cells undergo a mitotic or meiotic arrest that sustains until sexual maturity. In post-pubertal animals, the gametogenesis is reactivated and produces haploid germ cells that are able to generate individuals with a unique genetic composition. Upon fertilization, a second wave of epigenetic resetting takes place to reassure the capability of sperm and egg DNA to fuse and generate all the cell types of the new individual (Cantone and Fisher, 2013).

### 2.2.2 Spermatogenesis

Male gametogenesis is a remarkable process of cellular differentiation that produces millions of sperm cells daily. In testis of an adult organism, the germ cells undergo extensive clonal

expansion through mitosis, reduction division of the genome in meiosis and profound morphological differentiation of haploid gametes (reviewed in Rousseaux *et al.*, 2005). These carefully regulated processes of differentiation are spatio-temporally organized in the seminiferous epithelium of testis and include genome-wide transcriptional programming, chromatin reorganization and transition of the epigenetic state (reviewed in Meikar *et al.*, 2012). Spermatogenesis is under an endocrine control via pituitary-hypothalamus-axis and by testosterone that is secreted by Leydig cells in the testicular interstitium (Dohle *et al.*, 2003; Rousseaux *et al.*, 2005). Inside the seminiferous tubules, Sertoli cells are the only somatic cells and vital for gamete production. The Sertoli cells provide the developing germ cells both physical and nutritional support, and enable coordinated differentiation and migration of germ cells from the basal lamina to the lumen of the tubule (Figure 3). The mature spermatozoa are highly specialized cells that are released to the lumen and transported to the epididymis where they gain the capacity to move and fertilize an egg.



**Figure 3. Cells of spermatogenesis.**

Spermatogenesis takes place in the epithelium of seminiferous tubules of testis. The only somatic cells inside the seminiferous tubules are Sertoli cells which provide the developing germ cells with physical and nutritional support. Spermatogonia reside on the basal lamina and undergo mitotic cell divisions. Primary spermatocytes (leptotene, zygotene, pachytene, diplotene) synthesize DNA and undergo crossing over prior to the meiotic divisions. The haploid spermatids are initially small and round but during spermiogenesis they differentiate into spermatozoa which are released into the lumen of the tubule. Contractions of the Myoid cells beneath the basal lamina flush the immotile spermatozoa to the epididymis. Interstitial cells include Leydig cells which secrete testosterone.

#### 2.2.2.1 Clonal expansion and meiosis of spermatogenic cells

Germ line stem cells and spermatogonia reside attached to the basal lamina and, in a process called spermatocytogenesis, undergo clonal expansion to maintain the stem cell population and to give rise to a large pool of cells that are committed for differentiation. The following meiosis of spermatocytes takes 1.5-2 weeks in mammals, includes synthesis of DNA, active transcription, chromatin remodeling and reduction division of the hereditary material (reviewed in Kimmins *et al.*, 2004). The synthesis of DNA occurs in leptotene, and the subsequent pairing of homologous

chromosomes in the zygotene spermatocytes. The following pachytene is characterized by fully formed synaptonemal complexes and exchange of genetic material during crossing over (reviewed in Kimmins and Sassone-Corsi, 2005; Turner, 2007). The pachytene stage is, furthermore, an active phase of RNA and protein synthesis and includes exchange of histones to testis-specific variants (Kimmins and Sassone-Corsi, 2005; Gaucher *et al.*, 2010; Soumillon *et al.*, 2013). A specific feature of the male gametogenesis is the separation of the X and Y chromosomes, which contain only a small region that is capable for recombination and crossing over. As a consequence of no crossing over, the X and Y chromosomes undergo meiotic sex-chromosome inactivation (MSCI; Turner, 2007), which is mediated by incorporation of histone variants such as macroH2A and H3.3 and specific histone modifications (McKee and Handel, 1993; Khalil *et al.*, 2004; Solari, 1974; reviewed in Burgoyne *et al.*, 2009). Particularly, phosphorylation of H2AX at serine 139 (γH2AX) has been identified as a hallmark for the transcriptional silencing and sequestration of the X and Y chromosomes into the cellular periphery where they form a structure called sex-body (reviewed in Turner, 2007; Burgoyne *et al.*, 2009). In diplotene spermatocytes, the synaptonemal complexes dissolve, and meiosis I and II separate the homologous chromosomes and sister chromatids, respectively. The meiotic divisions are relatively fast processes (completed within 24h) and result in the formation of four haploid spermatids per primary spermatocyte (reviewed in Hess and de Franca, 2008).

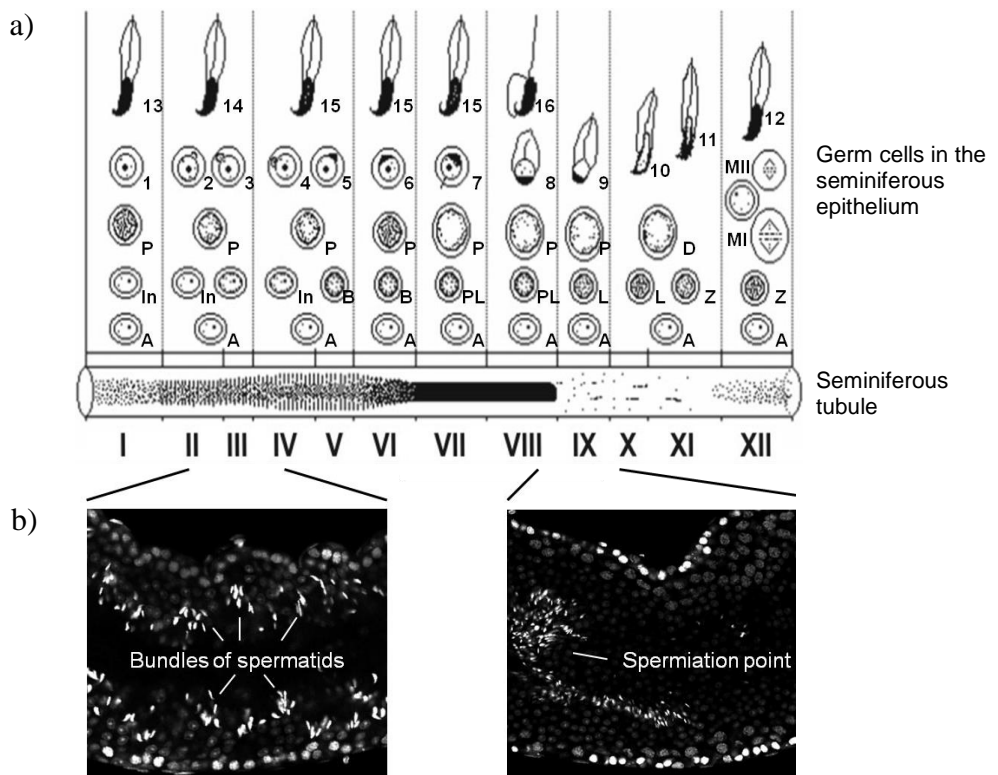
#### 2.2.2.2 Morphological differentiation of spermatids

Spermatids are initially small and round, but during a process called spermiogenesis, they develop into specialized spermatozoa with tightly packed DNA in the head, spirally organized mitochondria in the mid piece and a long tail which is composed of microtubules (Rousseaux *et al.*, 2005). In mouse, there are 16 steps of spermiogenesis during which spermatid-specific organelles form, the nuclei elongate and the DNA becomes tightly compacted (Oakberg, 1956; reviewed in Hess and de Franca, 2008). A remarkable phenomenon in step 2-6 spermatids is the high rate of transcription and the storage of RNA. The transcripts are stored in specialized organelles called chromatoid bodies and used later when the chromatin is tightly compacted and transcriptionally silent (reviewed in Kotaja and Sassone-Corsi, 2007; Parvinen 2005). Elongation of DNA is initiated in step 9 spermatids and the consequent shift from transcriptional to translational control is reflected in the many RNA species and RNA-binding proteins that are present in elongating spermatids (reviewed in Kleene, 2003; Paronetto and Sette, 2010). During steps 9-11, the DNA-packaging histones are changed to transition proteins (TNPs), which in turn are replaced by protamines (PRMs) in late spermatids. PRMs are small and positively charged proteins, the incorporation of which causes extreme compaction of chromatin and inactivation of transcription (Pogany *et al.*, 1981; reviewed in Braun, 2001; Miller *et al.*, 2010).

#### 2.2.2.3 Organization of spermatogenic cells in the seminiferous epithelium

During spermatogenesis, every mitotic and later meiotic cell division is followed by an incomplete cytokinesis that results in the formation of cytoplasmic bridges between the germ

cells (Erickson, 1973). These cytoplasmic channels unite germ cells that originate from one progenitor stem cell, and enable exchange of material and information between different developmental states (Ventelä *et al.*, 2003). Sharing material is especially crucial for the haploid spermatids that, besides becoming transcriptionally silent, contain only one of the two sex chromosomes. Remarkably, the chromatoid bodies have been shown to translocate via cytoplasmic bridges to other cells in the same cyncytium (reviewed in Parvinen, 2005). Together with Sertoli cells, the cytoplasmic bridges enable synchronized development of germ cells, which gives rise to defined cell associations termed stages. In mouse, there are twelve (I–XII) stages that follow each other as a wave of seminiferous cycle (Figure 4; Oakberg, 1956). Each stage contains a specified set of cells that give characteristic appearance for the tubule region (Figure 4; Kotaja *et al.*, 2004; Parvinen and Hecht, 1981). For a given tubule region it takes eight days to undergo all the twelve stages, whereas 4.5 cycles (~36 days) are required for a type A spermatogonia to develop into spermatozoa. Under a dissection microscope, the individual stages are challenging to



**Figure 4. Stages of seminiferous cycle in mouse.** a) Each stage (vertical columns, Roman numerals) contains a specific association of germ cells, which gives a specific light-absorbing pattern to the tubule. A: type A spermatogonium; In: intermediate spermatogonium; B: type B spermatogonium; PL: preleptotene spermatocyte; L: leptotene spermatocyte; Z: zygotene spermatocyte; P: pachytene spermatocyte; D: diplotene spermatocyte; MI: meiosis I; MII: meiosis II. The numbers indicate steps of spermiogenesis. 1–8: round spermatids; 9–12 elongating spermatids; 13–16 elongated spermatids. b) Confocal images of seminiferous tubules showing DNA in white.



identify, but four defined regions are easily recognized: Dark zone (DZ) contains stages VII-VIII and gets its light-absorbing appearance from the nearly mature spermatids that pack along the lumen (Figure 4). The abrupt release of the spermatozoa to the lumen is detected as a spermiation point that marks the transition from DZ to pale zone (PZ) (see Figure 4). PZ contains stages IX-XI and its most mature cells are step 9-11 spermatids that are sparsely distributed in the seminiferous epithelium. In the following weak spot (WS) the spermatids form bundles which are detected as small light-absorbing spots in stages XII-I. As the clusters of spermatids grow in size and their DNA condenses, the tubule gains a characteristic brush-like appearance that indicates stages II-VI of the strong spot (SS). Although the stages are identified mainly according to the morphology and associations of elongated spermatids, each stage contains also a defined set of spermatocytes and spermatids (Figure 4; Russell *et al.*, 1990). For example, DNA condensation in leptotene spermatocytes occurs in stages IX-X, pachytene phase extends over stages I-IX, meiotic divisions take place in stage XII and round spermatids reside in stages I-VIII (Figure 4).

### 2.3 The heat shock response

Cells are constantly challenged by external and internal conditions that can cause disrupted homeostasis. Such conditions include elevated temperatures, toxic compounds and cancer progression which all have deleterious effects on cellular organization and affect a broad range of structures from membranes to cytoskeleton and from organelles to the DNA. Mild protein-damaging stress causes microfilaments to reorganize, whereas severe stress leads to collapsed cytoskeletal networks (reviewed in Toivola *et al.*, 2010). Moreover, upon stress, the fluidity of plasma membrane increases, nucleoli swell, Golgi and ER become fragmented and mitochondria loose the correct structure (reviewed in Boulon *et al.*, 2010; Szalay *et al.*, 2007; Vigh *et al.*, 2007). As a result, the cellular transport system becomes defected and the production of ATP severely impaired. If not mitigated, the stress eventually causes apoptotic and necrotic death of the cells (reviewed in Richter *et al.*, 2010).

The cell's response to protein-damaging conditions is called heat shock response (HSR), a rapid and evolutionarily conserved mechanism that adapts elemental cellular processes to the harmful conditions. The HSR was discovered in 1962 by Ferruccio Ritossa who observed chromosomal puffs in heat-treated salivary gland cells of *D. melanogaster* larvae (Ritossa, 1962). These puffs appeared within minutes of the stress stimuli, coincided repeatedly at specific chromosomal loci and were already at the time known to be hallmarks of active transcription. It, however, took until 1974 before Alfred Tissières and others showed stress-induced production of RNA from these loci and correlated the transcriptional activation to the production of heat shock proteins (HSPs). Importantly, the onset of thermal stress was also detected to halt the production of constitutively expressed proteins owing to global silencing of transcription and translation (Lewis *et al.*, 1975; Spradling *et al.*, 1975; Tissières *et al.*, 1974; reviewed in Lindquist, 1981).

The HSPs are molecular chaperones and proteases that enforce correct folding, assembly, translocation and degradation of nascent and denatured proteins both in the cytosol and organelles (reviewed by Hartl *et al.*, 2011). Hence, the first line of defence against protein-

damaging conditions is the rapid production of proteins that maintain homeostasis (reviewed in Morimoto, 1998). HSPs are divided into families based on their molecular size (Kampinga *et al.*, 2009), and in human, HSPB1/HSP27, DNAJB1/HSP40, HSPA1A/HSP70.1, and HSPC/HSP90 have been shown to be the main stress-induced HSPs. The members of HSP70 and HSP90 chaperone families are ATP-dependent protein foldases that recognize exposed hydrophobic amino acid residues and provide an environment for an efficient refolding. Small HSPs (sHSPs) and type I and II chaperonins (HSPD1/HSP60 and CCTs, respectively) work independently of ATP by attaching to misfolded proteins and holding them until cleared by proteasomal degradation or refolded by the HSP40-HSP70 and HSP90 machineries (reviewed in Hartl *et al.*, 2011; Richter *et al.*, 2010). Besides leading to induced expression of molecular chaperones, heat stress has been detected to cause a stagnation of cell cycle progression and increased expression of metabolic enzymes, membrane modulating proteins, transcription factors and components involved in nucleic acid repair (Gasch *et al.*, 2000; Hahn *et al.*, 2004; Kühl and Rensing, 2000; Trinklein *et al.*, 2004; reviewed in Richter *et al.*, 2010).

## 2.4 The family of heat shock factors

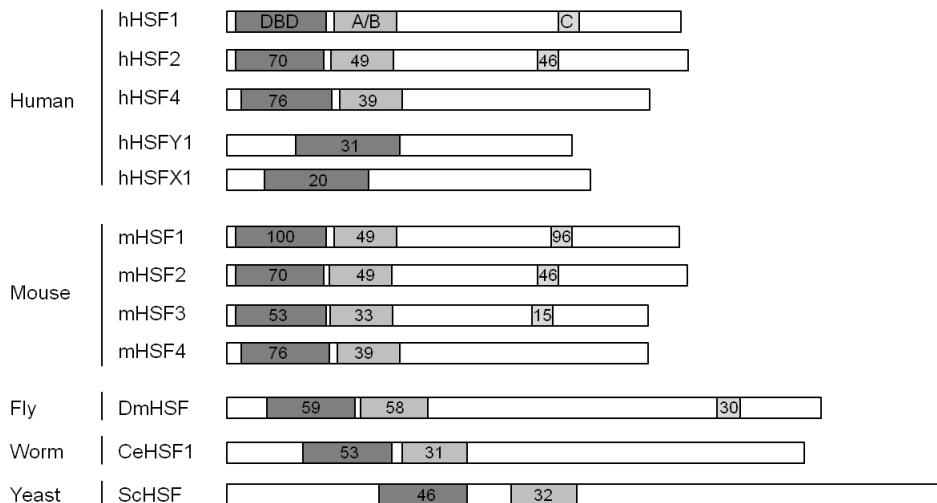
Cloning and deletion mapping of *HSP* genes enabled identification of the promoter element that mediates the heat-induced activation of transcription. This element consisted of inverted nGAAn pentamers and was named the heat shock element (HSE) (Amin *et al.*, 1988; Pelham, 1982; Sorger and Pelham, 1988). The identification of HSE, in turn, permitted purification and characterization of the HSE-binding protein which was named the heat shock factor (HSF) (Topol *et al.*, 1985; Wiederrecht *et al.*, 1987; Wu *et al.*, 1987). Since the initial discovery in yeast and fruit fly, the HSFs have been identified as a conserved family of transcription factors that orchestrate gene expression in eukaryote species. Besides providing stressed cells with chaperones that safeguard homeostasis, HSFs act on a plethora of physiological pathways that range from developmental processes to aging and immune responses (reviewed in Åkerfelt *et al.*, 2010; Fujimoto and Nakai, 2010). The recently discovered importance of HSFs for a wide range of core physiological processes is highlighted in severe pathologies, e.g. Alzheimer's and other neurodegenerative diseases, cancers, infertility and cataract that have been associated with disrupted activity of HSFs (Christians *et al.*, 2000; Dai *et al.*, 2007; Fujimoto *et al.*, 2004; 2005; Inouye *et al.*, 2004; Metchat *et al.*, 2009; Santagata *et al.*, 2011; 2013; Takaki *et al.*, 2006).

A single HSF is expressed in yeasts and invertebrates, whereas two whole-genome duplications during chordate/vertebrate development have led to the existence of four HSFs (HSF1-4) in mammals (reviewed in Fujimoto and Nakai, 2010). HSF1 is a homolog for the single HSF in yeasts and invertebrates, and the major director of stress responses. No other vertebrate HSF can compensate HSF1 for the heat-induced activation of *HSP* gene transcription (McMillan *et al.*, 1998; reviewed in Vihervaara and Sistonen, 2014). Also HSF2 is ubiquitously expressed in vertebrate species and coupled to responses to acute and chronic stress. However, the role of HSF2 as a regulator of stress responses has remained enigmatic and it has mainly been considered as a developmental factor and a modulator of HSF1-driven transcription (Chang *et al.*, 2006; Kallio *et al.*, 2002; Östling *et al.*, 2007; Shinkawa *et al.*, 2011; Sistonen *et al.*, 1992).

Similarly to HSF1, HSF3 is a stress-responsive factor, but it induces the expression of *HSPs* only in avian species (Fujimoto *et al.*, 2010). Also mice harbor a functional HSF3, but it is involved in the expression of non-classical heat shock genes (Fujimoto *et al.*, 2010). In humans, HSF3 is likely a pseudogene since no HSF3 transcript has been identified. HSF2 and HSF4 are mainly associated with developmental processes and, intriguingly, no stress-related function for HSF4 has been detected, indicating functional diversity of the HSF family members (Fujimoto *et al.*, 2004, reviewed in Fujimoto and Nakai, 2010). Both HSF2 and HSF4 have been shown to interplay with HSF1 at the target gene promoters, albeit at distinct tissues, physiological pathways and through distinct mechanisms. While HSF1 and HSF2 form heterotrimers and collaborate during stress responses and gametogenesis (Loison *et al.*, 2006; Östling *et al.*, 2007; Sandqvist *et al.*, 2009; Wang *et al.*, 2004), HSF4 competes with HSF1 for the same promoters in sensory plaque (Fujimoto *et al.*, 2004; Takaki *et al.*, 2006). Consequently, appropriate cooperation of HSF1 and HSF4 coordinates the spatio-temporal gene transcription and drives development and maintenance of lens and olfactory epithelium. The HSF family also includes HSF<sub>X</sub> and HSF<sub>Y</sub>, which are of sex chromosomal origin. Both HSF<sub>X</sub> and HSF<sub>Y</sub> are highly expressed in testis and coupled to defective spermatogenesis (Bhowmick *et al.*, 2007; Kinoshita *et al.*, 2006; Tessari *et al.*, 2004). However, the cellular functions and molecular mechanisms of these non-classical HSFs are still undetermined.

#### 2.4.1 The functional domains of HSFs

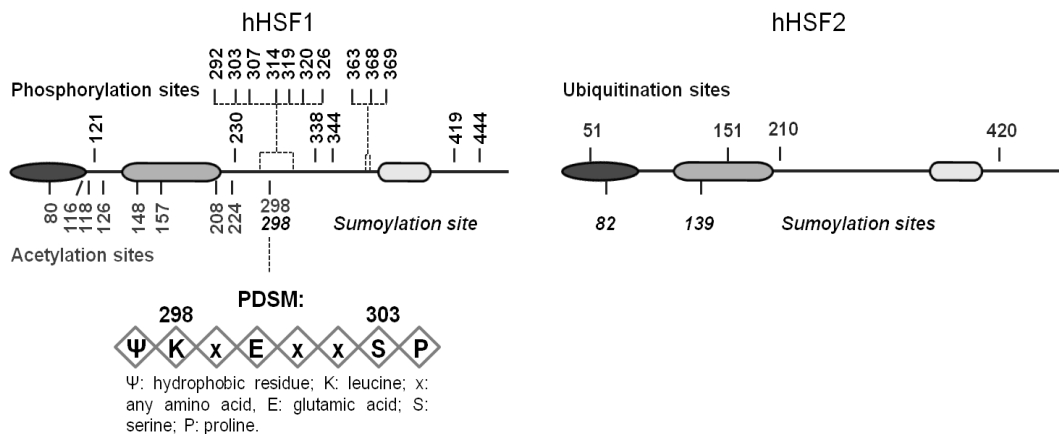
HSFs are characterized by a conserved N-terminal DBD that contains a looped helix-turn-helix structure (Figure 5; Littlefield and Nelson, 1999; Sorger and Pelham, 1988). Unlike most



**Figure 5. Functional domains and evolutionary conservation of HSFs.** Schematic presentation of HSF family members. The numbers in specified functional domains indicate percentage of amino acid similarity to human HSF1 and are according to Fujimoto and Nakai (2010). DBD: DNA-binding domain; A/B: oligomerization domain; C: hydrophobic heptad repeat C; h: *Homo sapiens*; m: *Mus musculus*; Dm: *Drosophila melanogaster*; Ce: *Caenorhabditis elegans*; Sc: *Saccharomyces cerevisiae*.

transcription factors, HSFs bind to the DNA as trimers and the oligomerization is mediated by a conserved HR-A/B domain that is constituted of heptad repeats (Peteranderl *et al.*, 1992; Rabindran *et al.*, 1993; Sorger and Nelson, 1989). Hydrophobic amino acid residues of the HR-A/B enable formation of a coiled-coil structure between HSFs and allow for the trimerization. Spontaneous HSF activation is thought to be inhibited by a C-terminal heptad repeat (HR-C) that forms intramolecular contacts with the HR-A/B and, thereby, allosterically blocks the oligomerization (Rabindran *et al.*, 1993). This hypothesis is supported by deletion studies where lack of HR-C rendered HSF1 constitutively active (Rabindran *et al.*, 1993; Zuo *et al.*, 1994; Nakai *et al.*, 2000), and by yeast HSF and mammalian HSF4, both of which lack the HR-C (Figure 5) and are constitutively bound to the DNA (Jakobsen and Pelham, 1988; Nakai *et al.*, 1997). HSFs also contain regulatory domains (RDs), nuclear localization signals (NLSs) and transactivation domains (ADs), which are targeted by several protein complexes that coordinate the steps and the magnitude of HSF activation (reviewed in Anckar and Sistonen, 2011). However, the number, location and amino acid sequences of these regulatory regions vary between distinct HSFs.

All HSFs recognize *cis*-acting HSEs, but show preferences over the precise sequence and architecture of the nGAAn repeats (Kroeger and Morimoto, 1994; Yamamoto *et al.*, 2009). Intriguingly, increasing number of nGAAn pentamers have been shown to mediate cooperative binding of HSFs, and the exact composition of the HSE to affect the transcriptional activation of target genes (Hashikawa *et al.*, 2007; Perisic *et al.*, 1991; Sakurai *et al.*, 2010; Xiao *et al.*, 1991).



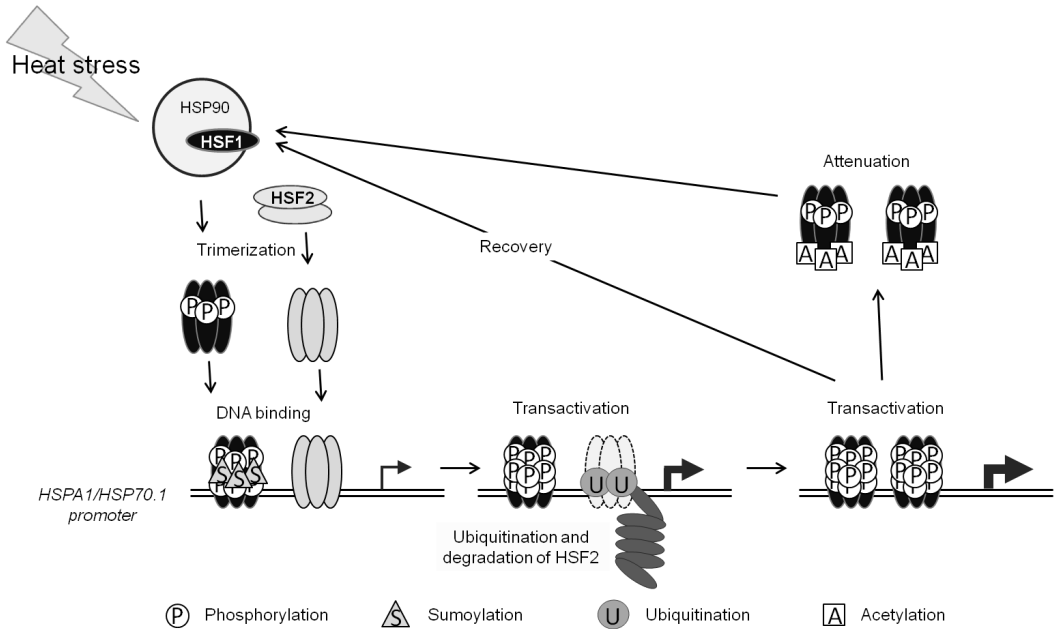
**Figure 6. Post-translational modifications of human HSF1 and HSF2.** HSF1 (left) is phosphorylated at multiple amino acids (indicated by numbers above the schematic HSF1) and subjected to acetylation (gray, below HSF1) as well as sumoylation (italicized). Only a few amino acids of HSF2 (right) are ubiquitinated (above HSF2) or sumoylated (italicized). The augmented regions in the HSFs denote DBD, HR-A/B and HR-C and are indicated as in Figure 5. The RD of HSF1 locates between the HR-A/B and HR-C domains (not shown) PDSM: phosphorylation-dependent sumoylation motif.

#### 2.4.2 Regulation of HSF1 and HSF2

The molecular structure and stress-sensitivity of HSF1 is highly conserved in eukaryote species. The need for rapid and delicate control of HSF1 is highlighted by the multitude of protein-protein interactions and PTMs that coordinate its activity (Figures 6 and 7; reviewed in Anckar and Sistonen, 2011; Vihervaara and Sistonen, 2014). According to the current model, HSPs can directly bind to and inhibit HSF1, which creates an autoregulatory mechanism that senses protein folding in the cell and fine-tunes the intensity of the stress responses (Baler *et al.*, 1992; Shi *et al.*, 1998). While HSP90 has been shown to restrain HSF1 in a monomeric state, HSP40 and HSP70 interact with trimeric HSF1 and dampen its transactivating potential (Abravaya *et al.*, 1992; Guo *et al.*, 2001; Zou *et al.*, 1998). Upon heat stress, HSF1 dissociates from the HSPs, trimerizes and binds to the DNA (Figure 7). The stress-induced activation of HSF1 is accompanied with hyperphosphorylation of several amino acid residues at the RD that is localized between HR-A/B and HR-C domains (Figure 6; reviewed in Anckar and Sistonen, 2011). Although the hyperphosphorylation of HSF1 is associated with its transactivating capacity, several amino acids are phosphorylated also in the absence of stress and repress the activation (Chu *et al.*, 1996; Kline and Morimoto, 1997). Intriguingly, of the multitude of PTM events that occur on HSF1, only phosphorylations of S230 and S326 have been coupled to increased transactivation (Holmberg *et al.*, 2001; Guettouche *et al.*, 2005). Instead, several PTMs contribute to defining the extent and duration of HSF1-mediated transcription. At so called phosphorylation-dependent sumoylation motif (PDSM), phosphorylation of S303 primes K298 for sumoylation, which in turn, restrains HSF1-mediated transcription (Hietakangas *et al.*, 2003; 2006). The PDSM was first characterized on HSF1 but it is now recognized as a conserved mechanism that controls a range of transcription factors (Anckar and Sistonen, 2007). During prolonged stress, HSF1 activity attenuates and the *HSP70* transcription ceases. The removal of HSF1 from the chromatin is facilitated by acetylation of K80, a site that directly contacts the DNA (Figures 6 and 7, Westerheide *et al.*, 2009). The HSF1 occupancy on DNA, and thus the transcription of *HSP70*, can be prolonged by Sirtuin 1 (SIRT1), which is a deacetylase, nutrient sensor and longevity factor that has been suggested to link HSF1 to the metabolic state of the organism (Westerheide *et al.*, 2009).

In striking contrast to the multitude of PTMs on HSF1, HSF2 is mainly regulated at the level of its expression (Ahlskog *et al.*, 2010; Björk *et al.*, 2010; reviewed in Björk and Sistonen, 2010). To date, no phosphorylation of HSF2 has been reported and the only PTMs detected on HSF2 are ubiquitination of S51, K151, K210 and K420, as well as sumoylation of K82 and K139 (Figure 6; Xu, *et al.*, 2012). Also the activation of HSF1 and HSF2 are controlled by distinct mechanisms: Although HSF2 binds to the DNA as a trimer, in unstressed conditions it exists mainly as a dimer (Figure 7; Sistonen *et al.*, 1994). Moreover, the mere increase in HSF2 concentration has been shown to cause its translocation to the nucleus and binding to the DNA (Sandqvist *et al.*, 2009). HSF1 and HSF2 co-occupy *HSP* promoters, but since HSF2 is a poor activator of *HSP* genes (Kroeger *et al.*, 1993; Sarge *et al.*, 1993; Yoshima *et al.*, 1998), these factors are likely to display diverse mechanisms at the target genes. The factor-specific control and different transcriptional contributions are partially explained by diverging amino acid sequences of HSF1 and HSF2. While the DBD, HR-A/B and HR-C are well conserved between HSF1 and HSF2 (Figure 5),

their overall amino acid similarity is only 35% (reviewed in Pirkkala *et al.*, 2001). The distinct regulation of HSF1 and HSF2 are highlighted by the selective removal of HSF2 from the *HSP70* promoter during heat stress (see Figure 7; Ahlskog *et al.*, 2010). The detailed mechanisms of regulation and transcriptional contributions of HSFs remain to be elucidated, but the highly specific coordination of HSF1 *versus* HSF2 demonstrates that cells have both the need and the means to delicately control the interplay of HSFs (Figure 7).



**Figure 7. Activation-attenuation cycle of human HSF1 and HSF2 upon heat stress.** Heat stress induces trimerization of HSF1 and HSF2 and leads to their binding to *HSP70.1/HSPA1A* promoter. The activity of HSF1 is controlled by HSPs and a multitude of PTMs, whereas HSF2 is mainly controlled at the level of expression. From the HSF1-HSF2 complex at the *HSP70.1* promoter, HSF2 is selectively degraded during the heat stress.

### 2.4.3 HSF1-mediated transcriptional initiation

The rapid HSF-mediated increase in *HSP* expression during stress has provided a robust model for studying transcriptional responses (reviewed in Guertin *et al.*, 2010). Particularly the polytene chromosomes in fruit fly *D. melanogaster* have been exploited for detailed investigations on the dynamic recruitment of transcription factors to the *Hsp70* loci (Yao *et al.*, 2006; Zobeck *et al.*, 2010). The *Hsp70* promoter is primed for activation by GAGA binding factor (GAF) and transcriptionally engaged RNPII that is paused by DRB sensitivity-inducing factor (DSIF) and NELF (Rougvie and Lis, 1988; Rasmussen and Lis, 1993; Wu *et al.*, 2003). Pausing of transcriptionally engaged RNPII was discovered at the *Hsp70* loci and is today acknowledged as a common mechanism that confers rapid gene activation and temporal coordination of transcriptional programs (Rasmussen and Lis., 1993; Rougvie and Lis, 1988; reviewed in Adelman and Lis, 2012). Heat stress induces binding of HSF to the *Hsp70* promoter and leads to recruitment of the mediator complex as well as P-TEFb that phosphorylates the C-terminal region

of RNPII (Brès *et al.*, 2008; Lis *et al.*, 2000; Marshal *et al.*, 1996; Park *et al.*, 2001). The transcriptional elongation is coupled to chromatin remodeling via removal of histones by facilitates transcription (FACT) and suppressors of Ty (SPT5, SPT6) as well as relieving of DNA coils by topoisomerase I. Moreover, poly-ADP ribose polymerase (PARP) generates a transcriptional compartment in which transcriptional components are efficiently recycled (Andrulis *et al.*, 2000; Gilmour *et al.*, 1986; Kaplan *et al.*, 2000; Petesch and Lis, 2008; Zobeck *et al.*, 2010).

Also in mammals the recruitment of HSF1 leads to the release of paused RNPII and removal of nucleosomes along the *HSP70* (Brown *et al.*, 1996; Brown and Kinston, 2007). However, the highly conserved function of HSF1 as inducer of *HSP* expression has species-specific mechanisms. In mammals, HSF1 collaborates with replication factor A (RPA) to maintain the *HSP70* promoter in an accessible state and, upon stress, the nucleosomes are removed along the gene via switch/sucrose nonfermenting (SWI/SNF) chromatin remodeling complex (Brown *et al.*, 1996; Fujimoto *et al.*, 2012; Sullivan *et al.*, 2001). In mammals, HSF1 also interacts with HSF2 which has been suggested, promoter-specifically, to modulate the HSF1-driven gene expression (Östling *et al.*, 2007; Sandqvist *et al.*, 2009).

#### 2.4.4 HSF1 and HSF2 in developmental processes

The significance of HSFs for developmental processes was revealed by inactivation studies. In yeast, HSF is indispensable for growth and viability in non-stressed conditions, and in fruit fly, it is required for oogenesis and larval development (Gallo *et al.*, 1993; Jedlicka *et al.*, 1997; Sorger and Pelham, 1988). In mammals, HSFs are involved in a variety of developmental processes from embryogenesis to gametogenesis. Although *Hsf1*<sup>-/-</sup> mice are viable, they show growth retardation and female sterility due to placental insufficiencies (Christians *et al.*, 2000; Xiao *et al.*, 1999). Furthermore, HSF1 is required for correct oogenesis, IgG production and formation and maintenance of sensory epithelium (Fujimoto *et al.*, 2004; Inoye *et al.*, 2004; Le Masson *et al.*, 2011; Metchat *et al.*, 2009; Takaki *et al.*, 2006). Abundant HSF2 expression has been detected during embryogenesis, particularly in the developing nervous system (Min *et al.*, 2000; Rallu *et al.*, 1997), and consequently, HSF2 has been shown to be essential for correct cortical lamination (Chang *et al.*, 2006). Male gametogenesis has been reported to be intact (Izu *et al.*, 2004) or only slightly defected (Salmand *et al.*, 2008) in *Hsf1*<sup>-/-</sup> mice, whereas the deficiency of HSF2 causes impaired spermatogenesis due to increased apoptosis of meiotic spermatocytes (Kallio *et al.*, 2002; Wang *et al.*, 2003). The fine cooperation between HSF1 and HSF2 is manifested by *Hsf1*<sup>-/-</sup>*Hsf2*<sup>-/-</sup> double knockout mice where a total block in meiosis inhibits formation of post-meiotic germ cells (Wang *et al.*, 2004). Thus, knocking out both HSF1 and HSF2 intensifies the meiotic defects in *Hsf2*<sup>-/-</sup> mouse to sterility, although the lack of HSF1 or HSF2 alone does not seem to remarkably impair male germ cell production. Intriguingly, HSFs have not been coupled to HSP expression during development and are suggested to be activated by distinct mechanisms in stress and development (Jedlicka *et al.*, 1997, reviewed in Abane and Mezger, 2010).

#### 2.4.5 HSFs in cancer progression

The importance of HSFs for core cellular processes is highlighted in the central role they display in cancer progression and metastasis (Dai *et al.*, 2007; Mendillo *et al.*, 2012; Santagata *et al.*, 2011; 2013). A ground breaking study by Dai and coworkers revealed that mice devoid of HSF1 are protected from carcinogen-induced skin tumors (Dai *et al.*, 2007; reviewed in Solimini *et al.*, 2007). Later, large patient studies coupled high HSF1 activity to cancer metastasis and identified HSF1 as a major marker for poor prognosis (Mendillo *et al.*, 2012; Santagata *et al.*, 2011; 2013). HSF1 was, furthermore, shown to drive a complex transcriptional program in human breast cancer cell lines (Mendillo *et al.*, 2012). Since cancer cells have defective cell cycle control, are highly proliferative, mutation prone, and live in crowded and oxygen-deficient environments (Hanahan and Weinberg, 2011), a commonly held view is that HSF1 enables survival and metastasis by allowing cancer cells to adapt to the hostile conditions. To date, the role of HSF2 in cancer development remains uncharacterized. Consequently, elucidating the detailed mechanisms, the interplay and the complete set of target genes for HSF1 and HSF2 is required for understanding the physiological functions of HSFs in distinct conditions.