UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 403

Genetic and Genomic Analysis of Transcriptional Regulation in Human Cells

MEHDI MOTALLEBIPOUR

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2008

ISSN 1651-6206
ISBN 978-91-554-7355-6
urn:nbn:se:uu:diva-9407

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbecklaboratoriet, Uppsala, Friday, December 12, 2008 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract
Motallebipour, M. 2008. Genetic and Genomic Analysis of Transcriptional Regulation in Human Cells. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 403. 64 pp. Uppsala. ISBN 978-91-554-7355-6.

There are around 20.000 genes in the human genome all of which could potentially be expressed. However, it is obvious that not all of them can be active at the same time. Thus, there is a need for coordination achieved through the regulation of transcription. Transcriptional regulation is a crucial multi-component process involving genetic and epigenetic factors, which determine when and how genes are expressed. The aim of this thesis was to study two of these components, the transcription factors and the DNA sequence elements with which they interact.

In papers I and II, we tried to characterize the regulatory role of repeated elements in the regulatory sequences of nitric oxide synthase 2 gene. We found that this type of repeat is able to adopt non B-DNA conformations in vitro and that it binds nuclear factors, in addition to RNA polymerase II. Therefore it is probable that these types of repeats can participate in the regulation of genes.

In papers III-V, we intended to analyze the genome-wide binding sites for six transcription factors involved in fatty acid and cholesterol metabolism and the sites of an epigenetic mark in a human liver cell line. For this, we applied the chromatin immunoprecipitation (ChIP) method together with detection on microarrays (ChIP-chip) or by detection with the new generation massively parallel sequencers (ChIP-seq). We found that all of these transcription factors are involved in other liver-specific processes than metabolism, for example cell proliferation. We were also able to define two sets of transcription factors depending on the position of their binding relative to gene promoters. Finally, we demonstrated that the patterns of the epigenetic mark reflect the structure and transcriptional activity of the promoters.

In conclusion, this thesis presents experiments, which moves our view from genetics to genomics, from *in vitro* to *in vivo*, and from low resolution to high resolution analysis of transcriptional regulation.

Keywords: Transcription, ChIP-chip, ChIP-seq, genome-wide, transcription factors, microsatellite, epigenetic

Mehdi Motallebipour, Department of Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-75185 Uppsala, Sweden

ISSN 1651-6206
ISBN 978-91-554-7355-6
urn:nbn:se:uu:diva-9407 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-9407)

یک چند بکودکی باستاد شدیم

یک چند ز استادی خود شاد شدیم

پایان سخن شنو که ما را چه رسید

از خاک بر آمدیم بر باد شدیم

In childhood we strove to go to school,
Our turn to teach, joyous as a rule
The end of the story is sad and cruel
From dust we came, and gone with winds cool.
Translation by Shahriar Shahriari

Myself when young did eagerly frequent
Doctor and Saint, and heard great Argument
About it and about: but evermore
Came out by the same Door as in I went.
Translation by Edward Fitzgerald

Poem by Omar Khayyâm (1048-1122), Persian poet, mathematician, philosopher, and astronomer

To My Family

List of publications

I *The promoter of inducible nitric oxide synthase implicated in glaucoma based on genetic analysis and nuclear factor binding.*
Motallebipour M., Rada-Iglesias A., Jansson M., Wadelius C.
Molecular vision 2005 Nov 4; 11:950-7.

II *Two polypyrimidine tracts in the nitric oxide synthase 2 gene – similar regulatory sequences with different properties.*
Motallebipour M., Rada-Iglesias A., Westin G., and Wadelius C.
Manuscript

III *Novel genes in cell cycle control and lipid metabolism with dynamically regulated binding sites for SREBP-1 and RNA-polII in HepG2 cells detected by ChIP-chip.*
Motallebipour M., Enroth S., Punga T., Ameur A., Koch C., Dunham I., Komorowski J., Ericsson J., and Wadelius C.
Resubmitted to FEBS J.

IV *Genome-wide localization of hepatocytic nuclear factors HNF4α, FOXA2 and GABPα reveals many distal regulatory elements and bindings at novel TSSs.*
Wallerman O., **Motallebipour M.**, Enroth S., Patra K., Bysani M.S.R., Komorowski J., and Wadelius C.
Manuscript

V *Differential binding pattern of FOXA1 and FOXA3 and their relation to H3K4Me3 in HepG2 cells revealed by ChIP-seq mapping.*
Motallebipour M.*, Ameur A.*, Bysani M.S.R., Patra K., Komorowski J., and Wadelius C.
Manuscript

*These authors contributed equally to this work.

Additional publications

Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma.
Thorleifsson G., Magnusson K.P., Sulem P., Walters G.B., Gudbjartsson D.F., Stefansson H., Jonsson T., Jonasdottir A., Jonasdottir A., Stefansdottir G., Masson G., Hardarson G.A., Petursson H., Arnarsson A., **Motallebipour M.**, Wallerman O., Wadelius C., Gulcher J.R., Thorsteinsdottir U., Kong A., Jonasson F., Stefansson K.
Science. 2007 Sep 7;317(5843):1397-400.

Contents

Abbreviations

3C	Chromosome conformation capture
CAGE	Cap analysis of gene expression
ChIP	Chromatin immunoprecipitation
ChIP-chip	ChIP with detection on microarray
ChIP-seq	ChIP with detection on massively parallel sequencer
Co-IP	Co-immunoprecipitation
CTD	Carboxy-terminal domain
DBD	DNA binding domain
DEBTF	Distal element binding transcription factor
DLD-1	Human colonic adenocarcinoma cell line
DNA	Deoxyribonucleic acid
dsRNA	Double stranded ribonucleic acid
EMSA	Electrophoretic mobility shift assay
ENCODE	Encyclopedia of DNA elements
GABP	GA binding protein
GTF	General transcription factor
HAT	Histone acetyltransferase
HCFC1	Host cell factor C1
HCP	High CpG content
HDAC	Histone deacetylase
HDM	Histone demethylase
HeLa	Human cervix carcinoma cell line
HepG2	Human hepatocellular carcinoma cell line
HGP	Human genome project
HMT	Histone methyltransferase
HNF4α	Hepatocyte nuclear factor 4α
hnRNP	Heterogeneous nuclear ribonucleo-protein
ICP	Intermediate CpG content
IFN-γ	Interferon-γ
IL-1β	Interleukin-1β
iNOS	Inducible nitric oxide synthase

IOP	Intraocular pressure
IP	Immunoprecipitation/Immunoprecipitate
LCP	Low CpG content
LCR	Locus control region
Mb	Megabases
MCF7	Human breast adenocarcinoma cell line
MeDIP	Methylated DNA immunoprecipitation
MNase I	Micrococcal nuclease I
MPS	Massively parallel sequencer
NFR	Nucleosome free region
NO	Nitric oxide
NOS	NO synthase
NRF	Nuclear respiratory factor
PCG	Protein coding gene
PEBTF	Proximal element binding transcription factor
PIC	Preinitiation complex
POAG	Primary open angle glaucoma
PPu	Polypurine
PPy	Polypyrimidine
PRMT	Protein arginine methyltransferase
PTBP1	Polypyrimidine tract binding protein 1
PTM	Post-translational modification
RNA	Ribonucleic acid
RNA polII	RNA polymerase II
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
SREBP	Sterol regulatory element binding protein
TF	Transcription factor
TFBS	TF binding site
TNF-α	Tumor necrosis factor-α
TSS	Transcriptional start site
UES	Unique enriched spots
USF	Upstream stimulatory factor
UTR	Untranslated region

Introduction

Two scientific efforts during the past 55 years have been milestones in the field of biology. The first was the determination of the structure of deoxyribonucleic acid, DNA, by Watson and Crick[1]. In this work, the double helicity of the chains, the anti-parallelism of the strands, and the type of base pairs in the DNA was predicted. These findings paved the way for other contributions of major significance in the field of genetics. The second effort, which made the pace of progression in genetics to become faster, was sequencing of all the bases of the DNA in the human genome in the Human Genome Project (HGP)[2,3].

Around three billions of nucleotides constitute the human genome, out of which only 1-2% code for mRNAs. Until recently, only a small part of the remaining 98-99% of the "non-coding" sequences was considered to have a function and the rest were regarded as "junk DNA". This was specially thought about the highly abundant repeats. In spite of several studies suggesting a possible role for repeats, no larger effort has been initiated to increase the understanding of these parts of the human or mammalian genome. In fact, repeats, due to their nature and base composition, are technically considered as disturbing elements and therefore are omitted or masked in analyses that concern the genome.

Initiation of the HGP was also a start for refinement and development of molecular and technical tools. In the wake of the HGP, came the ENCODE (ENCyclopedia Of DNA Elements) project with the goal to determine all the functional elements in the human genome[4]. As a result, knowledge of composition and function of different parts of the human genome as well as the interest for exploration of these parts has increased greatly and rapidly. For instance, whole genome studies of gene expression with different methodological approaches have all come to the same conclusion, *i.e.* a much larger fraction of the genome is transcribed than thought before[5-7]. The new findings have also led the scientific community to, once again, rethink the general concepts, such as *gene* and *promoter*[5,8].

An organism is constantly under the influence of its changing environment, which has a direct effect on many internal systems. In order to maintain the inner balance, the organism must adjust the metabolism. In humans this is achieved through the activity of more than a 100.000 protein variants encoded from around 20.000 human genes. What essentially maintains the inner stability and allows this organism to be responsive to changes in its

environment is a homeostasis, which is achieved through different mechanisms, one of them being the regulation of expression of all these proteins. Regulation of expression is the result of equilibrium in an immensely complex network of transcription factors, themselves proteins, and their association with regulatory sequences of genes, together with the epigenetic modifications and spatial organization of chromatin. Hence, a change in the equilibrium in this network, results in a change in expression of a set of genes, and thereby a modification of the balance in levels of different proteins. It is therefore plausible to believe that transcriptional regulation is the prime process in an organism.

In this introduction, the genetic and epigenetic components of transcriptional regulation will be discussed together with a short general background on the studied transcription factors, followed by a presentation of non-B DNA structures and DNA polymorphisms.

Transcriptional regulation

Expression of RNA polymerase II (RNA polII) transcribed genes is regulated at several levels: transcriptional, post-transcriptional, translational, and post-translational. The transcriptional regulation can be divided in several steps: regulation of expression and activation of transcription factors (TFs) and cofactors activating or repressing the target gene, availability of the DNA and the regulatory sequences regulated by the epigenetic factors, the balance between inducers and inhibitors of transcription, and so on.

Structure of regulatory regions

The human genome can roughly be divided in two parts, namely coding regions and non-coding regions. The coding regions, as understood by the name, are nucleotides that code for the product of the gene, which are pre-mRNAs and functional RNAs (ribosomal RNA, transfer RNA, micro RNA, and small nucleolar RNA). A fraction of the non-coding sequences, constituted by promoters, enhancers, silencers, boundary elements, *etc.*, participate in transcriptional regulation of genes. These regulatory sequences, or cis-regulatory elements, can be described as follows:

Core promoter and regulatory sequences

The general concept of a promoter has been a sequence upstream of the first exon, which is necessary for initiation of the transcription of the gene. The promoter has then been divided in two parts, the core promoter and the proximal promoter[9,10]. The core promoter was described as the sequence bordering the transcribed region containing the transcriptional start site (TSS), of-

ten very short (-30 to +30 relative TSS), and constituted of elements that are necessary for assembly of the transcriptional machinery and initiation of transcription (Figure 1). The most necessary elements were usually considered to be the TATA-box – an AT-rich sequence – 20-30 bp upstream of an initiator sequence, the site of TSS. Hence, promoters independent of TATA-boxes were regarded to be in minority in the genome. Core promoters themselves are virtually silent and cannot initiate transcription and therefore need to cooperate with the proximal promoter. By this definition the proximal promoter starts from where the core promoter ends, can be up to 300 bp long, and contains the necessary elements for driving the basal transcription of the gene including binding sites for some TFs[11]. Parts of the non-coding sequences that participate in regulation of genes but cannot be defined as promoter can be categorized as regulatory sequences[9,10,12]. Here lie elements that bind TFs for regulating the expression of the gene and may contain enhancers and silencers that can be located at considerable distances from the regulated gene.

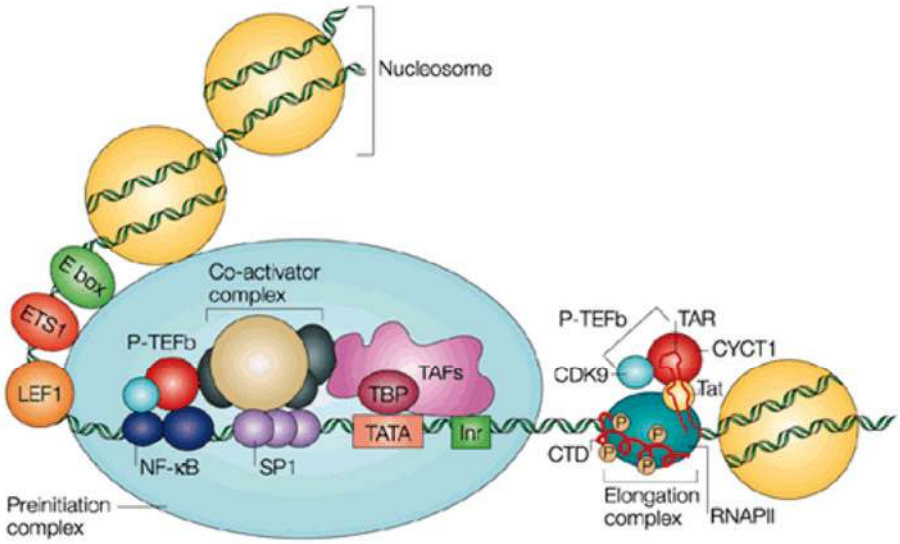

Figure 1. Structure of the promoter and the associated factors. The core promoter is constituted of the TATA-box and the TSS (Inr – initiator). In the core promoter, we also find TFs (SP1 and NF-κB) at their binding sites and in interaction with the GTFs and coactivators, together forming the preinitiation complex. Further downstream, the hyperphosphorylated RNA polII has left the TSS and initiated the transcription. As observed in the figure, the core promoter and the TSS are devoid of nucleosomes, which are instead positioned downstream of the elongating RNA polII and upstream of the proximal promoter. Adopted from Peterlin B. M. and Trono D. 2003. Nature Reviews Immunology[13].

Some of these hypotheses have been challenged by new findings. A study in 1% of the human genome in 16 cell lines demonstrated that only 16% of

the actively transcribed genes had promoters that harbored a TATA-box, and 19% a (CCAAT)-box[11]. This can be interpreted so that there are no single sequence elements essential for the transcription of all genes; rather diverse sets of genes employ different types of sequences depending on the function of those genes[14]. *E.g.* TATA-boxes are usually located upstream of tissue-specific genes, while CpG-islands are in vicinity of the ubiquitously expressed house-keeping genes. Most importantly, the notion of "single-gene – single-TSS" as a predominant structure was demonstrated to be misleading, when the human and mouse transcriptome was analyzed using the cap analysis of gene expression (CAGE) method[15,16]. Here, a larger fraction of the genes was found to contain multiple TSSs and thus four classes of promoters could be defined: 1) Promoters with a single dominant peak (SP), 2) Promoters with general broad distribution (BR), 3) Promoters with broad distribution with a dominant peak (PB), and 4) promoters with bi- or multimodal distribution (MU)[16]. While the promoters with a sharp TSS were strongly associated with the TATA-box, promoters with broad regions of TSS were associated with CpG islands. Additionally, CpG islands were also associated with bidirectional promoters.

Another important conclusion from the CAGE-tag studies was the abundance of TSSs within the gene body, especially in the 3'-UTR of protein coding genes (PCGs)[15,16]. This confirmed a similar finding made by localization studies of RNA polII in the ENCODE regions for the HeLa S3 cells[17]. In another study, treatment of MCF7 cells with 17beta-estradiol augmented the transcription from alternative TSSs, in particular those located in the 3'-UTR[18]. Although these data indicate the significance and wide usage of alternative TSSs, the function of these transcripts, specifically those starting in the 3'-UTR, is not completely known. Since some of these transcripts are antisense to the PCG they originate from, the hypothesis is that they participate in regulation of these genes[15].

One of the interesting features of regulatory regions is the bidirectionality of promoters, which is predicted to be fairly common. These are located between genes that are divergent or head-to-head with each of the genes encoded by one of the strands. An estimation based on 1% of the human genome calculated that 11% of the genes are arranged in this way and therefore share common regulatory elements[19], although a later whole genome study predicted this number to be 31% for PCGs[20]. Using the CAGE-tag data[16], which is a more complete and accurate map of all TSSs in the genome, the same study calculated that 78% of the CAGE-tags in the hepatocellular carcinoma cell line HepG2 were arranged in a bidirectional mode. Lin *et al.* could not identify any single common cis-regulatory element in all of the bidirectional promoters. However, they found binding motifs for NRF1, GABP, YY1, and NF-Y as the more frequent TFs, with GABP as the most frequent (83% of tested promoters)[21]. Moreover, 27% of the promoters were coregulated, while 1.7% were antiregulated, suggesting that the diver-

gent arrangement might be a way of coordinating expression of genes that participate in the same process. What unites all bidirectional promoters is that they are positioned in CpG-islands, as it has also been suggested by the CAGE-tag study[5,16,19,21].

The regulatory sequences contain the response elements for TFs, also known as TF binding sites (TFBS), which seem to be arranged in clusters[12]. Depending on the type of factors bound to them and hence their regulatory effect on genes, the clusters are called enhancers or silencers[22]. Enhancers and silencers not only affect the closest downstream gene, but may also regulate genes upstream or further downstream. Common for these two cis-regulatory sequences is that they contain more than one TFBS. Another element usually found in regulatory sequences is the boundary elements or the insulators. Insulators set the boundary between different domains of chromatin by limiting the effect of enhancers and silencers and also restrict the heterochromatin from spreading to neighboring regions[23]. The insulator function can be regulated through epigenetic mechanisms as well as binding of proteins. One of the widely abundant and well studied examples of an insulator binding protein is the CCCTC-binding factor (CTCF)[24]. The fourth element relevant for transcriptional regulation found in the regulatory sequences is the locus control region (LCR). LCRs are tissue-specific enhancers of expression for a group of genes or a locus and can be composed of several enhancers, insulators, and silencers[25]. LCRs are usually located at the DNAse I hypersensitive sites and must therefore be regions with an open chromatin formation for access of TFs.

Other non-coding sequences

Untranslated regions (UTRs) are parts of the gene that become transcribed to mRNA, but are not translated to proteins. They are located at the start (5'-UTR) and the end (3'-UTR) of the gene and thereby the resulting mRNA, and mostly participate in stability of the mRNA and regulation of translation. In addition, UTRs may also contain TFBS or be the site of TSS as previously mentioned.

Introns are the non-coding parts of a gene that are transcribed to mRNA and are spliced away before the mRNA is translated. The role of introns in the genome is not fully elucidated, but they contribute to differential splicing of transcripts and may also contain elements that participate in regulation of the gene they are contained in or genes at some distance.

Gene deserts are large areas of the genome that are void of any coding sequences. What function these sequences may have is largely unknown. In one study conservation of some sequences in these deserts was established, demonstrating that they can contain long-range enhancers[26]. Some of the gene deserts may also be the site of transcription for genes that have not yet been annotated, but which is now being revealed by the more sensitive techniques[5-7].

Transcription factors and the RNA polymerase II

Transcription factors or *trans*-acting factors are proteins that interact with DNA or other proteins to regulate the expression of genes. The sequence specific TFs are built of modules for DNA-binding, activation/repression, and regulation and may also contain a multimerization module[12]. There are around 1850[22] TFs in the human genome, which are categorized into families based on the DNA-binding motif, *e.g.* helix-turn-helix, homeodomain, zinc finger, helix-loop-helix, and leucine zipper[27]. The DNA-binding domain recognizes, with some specificity and varying affinity, short or long stretches of nucleotides, specific for a TF and which represents its consensus sequence[22]. A recent study corroborated that the DNA sequence is the determinant of TF binding and that the conservation of this sequence between species may play a minor role[28]. As the TF binds to its target sequence, the activation domain can interact with other factors and thereby exert an effect on the target gene. These other factors may be either additional activators or repressors, or another class of regulatory proteins, *i.e.* coactivators and corepressors. Cofactors not only act as mediators of interaction between different factors and the transcriptional machinery, but also function as chromatin and TF modifiers by regulating the post-translational modifications (PTMs) of the histones and the TFs[29-31]. One of the largest families of cofactors is the Mediator complex, demonstrated to be involved in regulation of various genes[32,33].

The biochemical models describe that transcription starts by binding of one or several TFs to their cognate TFBS and recruitment of the cofactors (Figure 2). The cofactors in turn modify the histones at the binding site and remodel the nucleosomes in the promoter to allow for binding of the transcriptional machinery. The cofactors at the same time contact and recruit the general transcription factors (GTFs) to the core promoter where they interact with the DNA and form the preinitiation complex (PIC). The PIC then directs the RNA polII to the TSS. At this point the carboxy-terminal domain (CTD) of the largest subunit of RNA polII is unphosphorylated, but upon recruitment of the transcription elongation factors, the CTD is hyperphosphorylated and transcription is initiated[12,22,34,35].

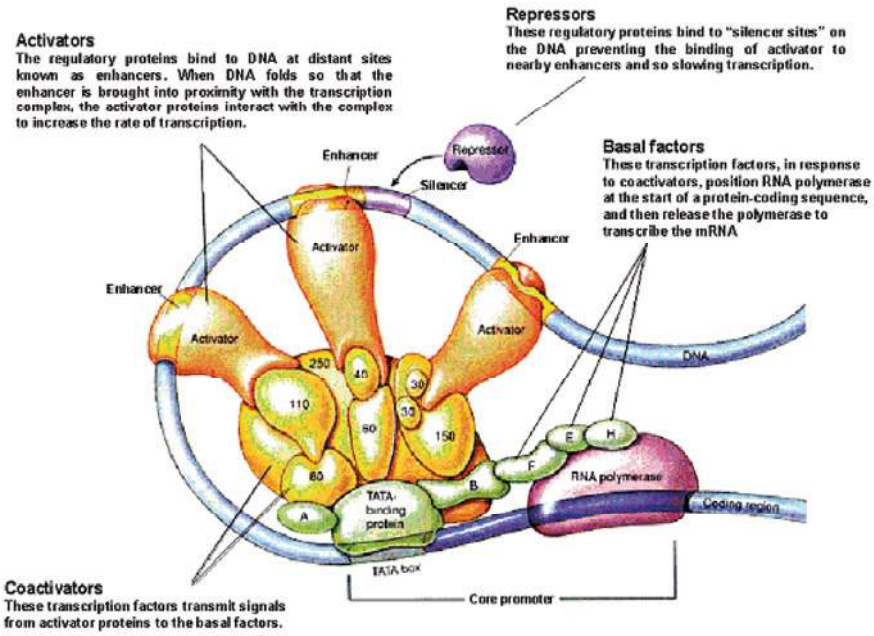

Figure 2. The transcriptional machinery. The figure is a representation of components in the transcriptional machinery. The GTFs are bound at the TATA-box in complex with the coactivators. The coactivators in turn are in interaction with the TFs (designated as activators in the figure) bound at their TFBSs in the enhancers. The RNA polII, bound at the TSS, is in contact with the GTFs and prepared for initiation of transcription. Adopted from *Regulation of gene expression in prokaryotes* by Donald P. Buckley.

Some examples of transcription factors

hnRNPs

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are a family of mRNA-binding proteins, that shuttle between the nucleus and the cytoplasm and are thought to participate in transcriptional regulation, telomere length maintenance, alternative pre-mRNA splicing, translation and turnover, *etc*[36,37]. Until now, two members of this family are indicated to interact with single stranded polypyrimidine sequences in DNA, namely hnRNPI and hnRNPK.

hnRNPI or polypyrimidine tract binding protein 1 (PTBP1) is proposed to participate in splicing and polyadenylation of mRNAs, however, it has also been demonstrated to interact with single stranded pyrimidine rich regions in DNA at gene promoters in a sequence specific manner and contribute to a higher promoter activity[38,39]. The target PPy-sequence has further been demonstrated to form triplex-DNA. PTBP1-binding and triplex formation was later confirmed *in vivo* in a mouse cell line, while a direct positive effect on the *Hmga2* could not be elucidated[40]. A study of PTBP1 in the ENCODE-

regions by chromatin immunoprecipitation (ChIP) with detection on microarrays (ChIP-chip) established that this protein bound throughout the gene body with a bias towards the promoter[41]. The binding in promoters often coincided with RNA polII. PTBP1 has been shown to interact with the 3'-UTR of both human and murine nitric oxide synthase 2 mRNA and regulate the stability at post-transcriptional level[42,43].

hnRNPK is suggested to be a TF that interacts with the transcriptional machinery, specifically the TATA-binding protein (TBP)[44]. One of hnRNPK target elements has been mapped to a PPy/PPu-sequence in the vicinity of the TSS for the c-myc promoter[45]. hnRNPK interacts with a diverse set of proteins in the cell, *e.g.* histone H4[46], chromatin modifying enzymes, protein kinases, GTFs, and RNA-processing factors, which is why it has been suggested that it functions as a docking platform for other proteins[47]. Additionally, both hnRNPK and PTBP1 have been demonstrated to cause bending and looping in their target DNA/RNA, which correlates well with their role in DNA tertiary structure binding[48,49].

SREBP

Sterol regulatory element binding proteins (SREBPs) belong to the basic helix-loop-helix (bHLH) leucine zipper family of TFs and are major regulators of the lipid- and cholesterol metabolic pathways[50-52]. Together with PPARγ and C/EBPα, they also participate in adipogenesis[53]. The three well known members of the SREBP-family are SREBP1a, -1c, and -2, although other splice variants have been reported[54,55]. SREBP1a and -1c are two isoforms encoded by the same gene, with dissimilar transactivation strength and different abundance in tissues and cells, as well as difference in recruited coactivators[56,57].

SREBPs are mainly regulated at the post-translational stage by sequestering the precursors in the endoplasmic reticulum (ER) membrane[52,58]. Lower level of lipids and cholesterol or induction by insulin signaling relieves the inhibitory interaction by a proteolytic cleavage in two consecutive steps, which release the active form of the factor. In the nucleus, the mature homodimer then interacts with the p300/CBP, the ARC/Mediator, and the SWI/SNF complex to regulate the transcription of target genes[59-61]. A genome-wide promoter study by ChIP-chip in insulin treated HepG2 cells revealed 1141 putative target genes for SREBP1[62]. Among these, genes involved in lipid and cholesterol metabolism, insulin signaling pathway and targets, cellular respiration, and cell cycle regulation could be found.

HNF4α

Hepatocyte nuclear factor 4 (HNF4) is an orphan nuclear receptor with a zinc finger as DNA binding domain (DBD)[63]. There are three main members in the family, α, β, and γ, as well as seven known splice variants. Although this receptor was originally categorized as orphan, later it was demonstrated

that long-chain fatty acids bound HNF4α as acyl-CoA esters[64]. Active HNF4α binds DNA as homodimer and interacts with GTFs, p300/CBP, and p/CAF for transcriptional regulation of target genes[63]. Genes in glucose, lipid, and drug metabolic pathways and blood coagulation are targets of this TF. Moreover, HNF4α is required for hepatocyte differentiation[65]. Mutations in the gene encoding this factor have been associated with maturity onset diabetes of the young type 1 (MODY1)[66].

GABP

GA binding protein (GABP), also known as nuclear respiratory factor 2 (NRF2; not to be confused with nuclear-factor-E2-related factor Nrf2) is a member of the Ets TFs with a winged-helix-turn-helix DNA binding motif (Figure 3)[67]. It forms a heterotetrameric complex when bound to DNA, which is constituted of two α and two β subunits. GABPα contains the DBD, while the transactivation domain is found in GABPβ. GABP interacts with Sp1, p300, HCFC1, and E2F1 and regulates the nuclear respiratory factor genes, in addition to genes regulating the cell cycle and apoptosis among others[67,68]. A genome-wide ChIP-seq study has identified 6442 putative binding sites in Jurkat (T-cell leukemia). As mentioned previously, GABP is a frequent binder in bidirectional promoters[21].

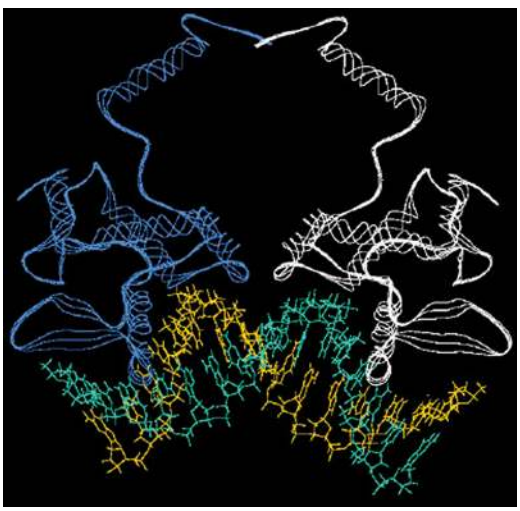

Figure 3. A transcription factor binding to DNA. The ternary complex of a member of the Ets-1 TF-family and DNA. The white and blue structures are the dimeric protein bound in the major groove of DNA (yellow and green). Adopted from Transcriptional regulation and Ets proteins by Marc Aumercier (http://www.ibl.fr/spip.php?rubrique12).

FOXAs

The FOXA subfamily belongs to the family of forkhead box/winged helix family of TFs. There are three FOXAs known today, FOXA1 (HNF3α), FOXA2 (HNF3β), and FOXA3 (HNF3γ)[63,69]. All three proteins share great homology in the DBD, but less outside of this domain. Additionally, they are predicted to bind DNA as monomers[70]. FOXAs interact with several other TFs for regulation of target genes, *e.g.* HNF1α, HNF4α, HNF6, USF1, and ER[71,72]. Since they also regulate some of these mentioned TFs at the gene level, they are proposed as master regulators of gene expression in the liver. Other targets of FOXAs are genes involved in metabolism and insulin signaling[69,73]. There have been a number of ChIP-chip and ChIP-seq studies of FOXA1 and FOXA2 in MCF7, mouse and human liver tissue, and HepG2, but none have covered all three factors together or all binding sites in the liver tissue[4,71,74,75].

The regulatory sequences of inactive genes are often in a state of compact chromatin and therefore not available for the TFs. However, there are examples of factors that can interact with the DNA in the compact chromatin and make it accessible for other factors. One example of these "pioneering factors" is the FOXAs, which have been found to resemble the histone H1 in structure and therefore are able to interact with the other histones in the histone core and reposition the nucleosomes[70,76,77]. This ability serves important purposes during the embryogenic development[69,78], and also constitutes an important link between the TF network and another imperative aspect of transcriptional regulation, namely the epigenetics.

Epigenetics

Epigenetics is defined as changes in the gene expression or cellular phenotype that can be inherited during cell division and that does not involve a change in the underlying DNA-sequence[79]. The molecular basis of epigenetics involves two main mechanisms, *i.e.* histone modifications and DNA methylation. It should be emphasized that these two processes are highly dependent on each other. Furthermore, these processes are highly dynamic, especially during the life time of the organism. The best example, which perhaps is also one of the most interesting aspects of the biology of human genome, is the finding that monozygotic twins have the same epigenetic pattern during the early years of life, but differ in both histone modification and DNA methylation later on, despite the common genotype[80].

Histones and their modifications

In the mammalian cell there are four types of histones in the histone core, H2A, H2B, H3, and H4 (Figure 4). Two copies of each of these histones together form an octameric complex, around which 147 bp of DNA is

wrapped and forms the smallest unit of the chromatin, the nucleosome. The fifth histone, histone H1 or linker histone, interacts with the nucleosome and seem to regulate the distance between the nucleosomes[81]. These together form the nucleosomal array in the chromatin, which can fold into the 30 nm fiber for higher order impaction of chromatin[82]. Other variants of the core histones have also been found, each with a specialized function[83]. Some examples are histone CENP-A, which marks the centromeric DNA in humans and histone H2A-Z found at the TSS of genes.

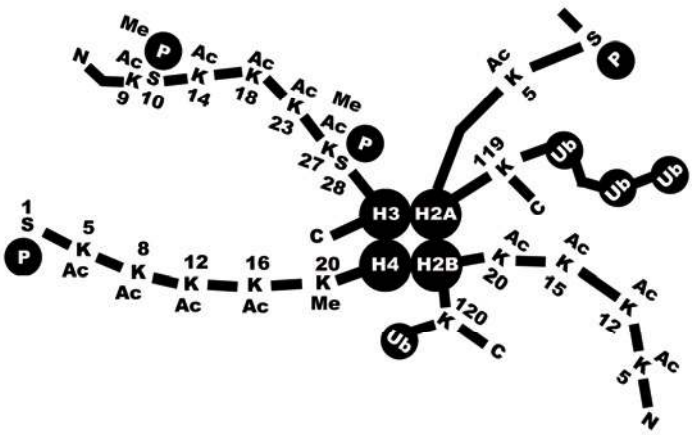

Figure 4. Configuration of a histone octamer. The histone octamer is constituted of two copies each of four types of histones, H2A, H2B, H3, and H4. The tail of each histone is modified at several positions, which together might constitute a "histone code". For simplicity, only one copy of each histone is shown.

The level of compaction of chromatin and thereby the accessibility of DNA is regulated by the type of modifications on the histone tails. There are eight types of modifications known today, acetylation, lysine methylation, arginine methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deimination, and proline isomerization[84,85]. The current knowledge is that acetylations, phosphorylations, and arginine methylations are positive marks of activity and sumoylation a mark of repression, while the other modifications can occur at both active and repressed genes. Modifications of the histone tails appear to change the net charge of the histones and thereby the accessibility of the DNA.

Table 1. *The known histone modifications and their effect on transcription*[84-87].

Histone	Residue	PTM	Function
H2A	K[a]5, 9	Ac[b]	Activation
H2A	K 119	Ub[c]	Repression
H2A	K126	Su[d]	Repression
H2A	S[e]/T[f]	Ph[g]	Activation
H2B	K5, K12, K20, K120	Ac	Activation
H2B	K5	Me[h] 1	Activation
H2B	S/T	Ph	Activation
H3	K4, 9, 14, 18, 23, 27, 36	Ac	Activation
H3	K4	Me[h] 1, 2, 3	Activation
H3	K9	Me 1	Activation
H3	K9	Me 2, 3	Repression
H3	K27	Me 1	Activation
H3	K27	Me 2	More active than silent
H3	K27	Me 3	Repression
H3	K36	Me 1	Activation
H3	K36	Me 3	Activation
H3	K79	Me 1, 2	No preference
H3	K79	Me 3	Activation
H3	R[i]2	Me 1, 2	Not apparent
H4	K5, 8, 12, 16, 91	Ac	Activation
H4	K20	Me 1	Activation
H4	K20	Me 3	Repression?
H4	R3	Me	No preference

a Lysine
b Acetylation
c Ubiquitylation
d Sumoylation
e Serine
f Threonine
g Phosphorylation
h Methylation
i Arginine

As presented in Table 1, there are many options of modification with each of them being associated with a state of transcription. Several of the residues on the same histone tail can be modified at the same time, some being mutually exclusive since they appear on the same residue, while others may be dependent on the modifications on other residues[84,85]. This cross-talk or combinatorial pattern was suggested already in 2000 by Strahl and Allis who coined the term "histone code" in parallel with the genetic code[88]. In 2008 this hypothesis was substantiated in Keji Zhao's lab. By genome-wide mapping of 19 methylation[86] and 18 acetylation[87] marks of histones in $CD4^+$ T cells through ChIP and high resolution sequencing (ChIP-seq), it was demonstrated that 4339 patterns existed in this cell out of which 1174 were associated with multiple genes[87]. Among these a combination of 17 modifications and a histone variant was associated with 25% of the promoters, a majority with the highest expression in this cell line.

Most of the histone modifications are dynamic and therefore a set of histone modifying enzymes are required for the establishment and removal of these marks. These enzymes are classified depending on what type of modification they add or remove[85]. The classes are histone acetyltransferase (HAT), deacetylase (HDAC), methyltransferase (HMT), demethylase (HDM), arginine methyltransferases (PRMT), serine/threonine kinase, ubiquitilase, and proline isomerase. Many of the members of these classes act as cofactors in transcriptional regulation, changing the chromatin environment and thereby the accessibility of DNA[84,85,89].

At the time of transcriptional activation of a gene, not only the histones are modified, but also the position of the nucleosomes is changed, in particular those at and in the vicinity of the TSS [90,91]. This usually results in eviction of the nucleosome at the TSS, leaving a so-called nucleosome free region (NFR). This nucleosome displacement in the chromatin is performed by ATP-dependent nucleosome remodeling complexes, such as the SWI/SNF family[92], NURF, and MLL1-3[93].

The total outcome of the interplay between the different histone modifications, histone variants and the nucleosome position is two states of chromatin, the euchromatin and the heterochromatin (Figure 5)[94]. The euchromatin is the transcription permissive state of chromatin with highly transcribed genes and mostly active histone modifications[95]. On the other hand, the heterochromatin is the transcription non-permissive state with poorly transcribed genes, mostly inactive histone modifications, usually DNA methylation, and tighter packaging of DNA. There are two types of heterochromatic DNA in cells, constitutive and facultative. The former designates regions of chromatin that are heterochromatic in every generation of the cell, *e.g.* centromeric regions of the chromosome[96], while the latter denotes regions that are heterochromatic in some cells, but euchromatic in others. Examples of such regions involve processes of X chromosome inactivation, imprinting, and allelic exclusion[97]. The process of heterochromatin formation involves RNA interference, DNA methylation and repressive marks on the histones, as well as repositioning of the chromatin in the nucleus[97-99].

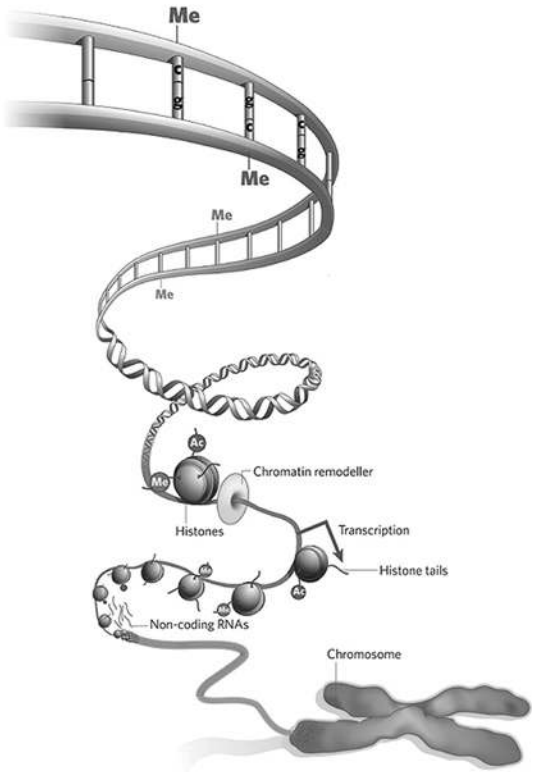

Figure 5. The epigenome. Two anti-parallel strands base pair with each other and together form the double helix. The double helix is wrapped around the histone core proteins twice, which constitute the nucleosome. The nucleosomes build up an array, which is further compacted to 10 and 30 nm fibers for higher compaction and formation of the chromosome. The chromatin and the nucleosomes are constantly modified to permit or repress transcription from the DNA. Adopted from Qiu, J. 2006. Nature[100].

DNA methylation

The second epigenetic mechanism widely spread in organisms is the methylation of DNA. This occurs by addition of a methyl group to the cytosine in a CpG dinucleotide by DNA methyltransferases (DNMTs). CpGs occur in a lower frequency than expected in the human genome[2,3]; however in some regions of the DNA they are highly abundant and form the so-called CpG-islands. As mentioned earlier, these epigenetic marks are involved in silencing of gene expression and formation of heterochromatic regions. The mechanism of silencing is either through direct hindrance of protein binding, *e.g.* inhibition of CTCF binding in the imprinting control region (ICR) of IGF2 and H19 genes[101], or through recruitment of factors with a methyl-binding domain (MBD), *e.g.* MeCP2, which in turn engage HDACs[102,103].

Several methods have been developed for studies of DNA methylation[104], some not suitable for genome-wide studies and some too laborious. The first genome-wide study of DNA methylation in humans employed a high-throughput method with somewhat high resolution based on immunoprecipitation of methylated DNA (MeDIP) and detection on BAC-clone and CpG island promoter arrays[105,106]. Here, it was concluded that in human primary lung fibroblasts gene-rich regions are highly methylated, while gene-poor regions are less methylated and that CpG-islands are hypomethylated. Further, it was shown that the inactive chromosome X is less methylated than expected.

Due to lack of a unifying definition of CpG-islands Saxonov *et al*. suggested a separation of regions based on CpG content[107]. They found that the promoters can easily be divided into two classes, the high CpG content (HCP) constituting 72% of the promoters, and the low CpG content (LCP) with 28%. In their follow-up study, Weber *et al*., took this classification further and suggested a third class, the intermediate CpG content (ICP)[108]. Then, using a promoter microarray with higher resolution, it was demonstrated that 42% and 33% of the hypermethylated promoters are in LCPs and ICPs, respectively, although they only form 35% of all the promoters. By analyzing the RNA polII and H3K4Me pattern in the same cell using ChIP, the authors concluded that methylation in the LCP does not always exclude transcription, and that inactive promoters in the HCP class were mostly hypomethylated. It was also suggested that H3K4Me2 may be a protective modification for keeping silent genes with HCP unmethylated. A genome-wide study in mouse embryonic stem cells applying the reduced representation bisulphate sequencing method confirmed these findings and inferred that DNA methylation is correlated with H3K4Me2, H3K4Me3, H3K9Me3, and H3K27Me3 pattern rather than the genomic sequence[109].

Spatial organization of chromatin

There is a growing body of evidence that DNA and chromatin is in constant movement and that the DNA is relocalized depending on the external signals and the operations under way. Newly developed methods such as chromosome conformation capture (3C)[110] and the modifications thereof, *i.e.* 4C[111,112], 5C[113], and 6C[114], and RNA FISH TRAP[115] (tagging and recovery of associated proteins) have revealed an extensive network of intra- and inter-chromosomal interactions[116,117]. Moreover, there is accumulating data signifying that there are "transcription factories" in the cell nucleus to which the DNA is translocated to be transcribed[118-120]. This contradicts the general belief that the RNA polII is recruited to the active gene. Together with these new models of transcription and long-range interactions, comes the theory of DNA-looping which is thought to play an important role in transcriptional regulation[120,121]. It is further implied that genes, depending on their level of

activity, are localized in different areas of the nucleus with the more active genes toward the inner nuclear space and the less active genes toward the nuclear periphery[122]. One example is the heterochromatic DNA, which has been shown to be tethered to the lamin protein in the nuclear periphery. It should be mentioned that there is also evidence of some active genes positioned in the vicinity of the nuclear envelope and the nuclear-pore complexes, although the reason for this is not yet completely clear.

Nuclear organization and the 3D structure of chromatin and nucleus is currently a hot topic in the genome biology. Nevertheless, the current research, although at its infancy, has an important take-home message: DNA and chromatin is not static in the nucleus, rather it is in constant motion and movement, whether it is for a few hundred bases or several megabases (Mb).

Non-B DNA structures

Repeated sequences are prone to form other structures than B-DNA. B-DNA, the common structure in cells, is the double-helical right-handed conformation with anti-parallel chains and Watson-Crick base-pairing[1]. Alternative secondary structures involve left- or right-handed conformation that is not necessarily double-helical nor anti-parallel and usually contain base-pairings other than Watson-Crick[123]. Composition of nucleotides in the repeat decides the type of conformation, *e.g.* left-handed Z-DNA, cruciform, slipped strand, triplex, *etc.* Formation of Z-DNA, a left-handed double helix, is favored by CG-repeats[124]. CpG islands, one form of CG-repeats, are associated with approximately 50% of human gene promoters[125]. Methylation of these islands is coupled with gene-silencing; however, the methylation rather than the conformation is considered to cause this inhibition. Methylation of CpG-islands is demonstrated to stabilize Z-DNA *in vitro*[124]. Although the non-B DNA structures are common in regulatory sequences of genes and predicted to bind nuclear factors, it is not clear what purpose they fulfill. Some of these secondary structures are suggested to relieve the negative supercoiling that arises in DNA during transcription[124].

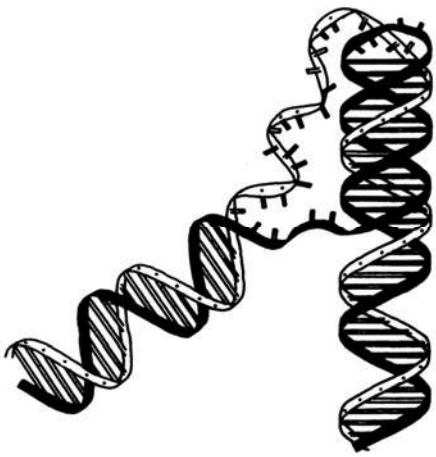

Figure 6. Representation of a triple helix. The triple helix or H-DNA is formed by a bend in the DNA double helix. One strand of the unwounded DNA interacts with the double helix in the major groove, while one strand becomes naked. Adopted from Wells, R.D. *et al.* 1988. Faseb J[126].

Pyrimidine/purine rich sequences or polypyrimidines/polypurines (PPy/PPu) are predicted to occur at 6-8 kb intervals[127] and are proposed to form a tetra-stranded structure[128]. This structure, commonly called H-DNA[129], constitutes of a triplex with a PPy·PPu forming a regular double-helix and a third strand, either a PPy or PPu, filling the major groove of the double-helix, leaving the fourth strand unpaired (Figure 6)[130]. The third and fourth strand of this tetraplex conformation can be either the 5' or 3' half of the PPy/PPu-sequence. Three different combinations of nucleotide base-pairings are common, each involving a Watson-Crick base pair, C·G·C^+, T·A·T, and G·G·C (Figure 7). It is not clear which base-pairing is normally preferred, nonetheless the length of the PPy/PPu tract may be a determinant[131]. Moreover, triplex DNA is shown to influence the conformation of adjacent sequences independent of the base composition of neighboring sequence[132,133].

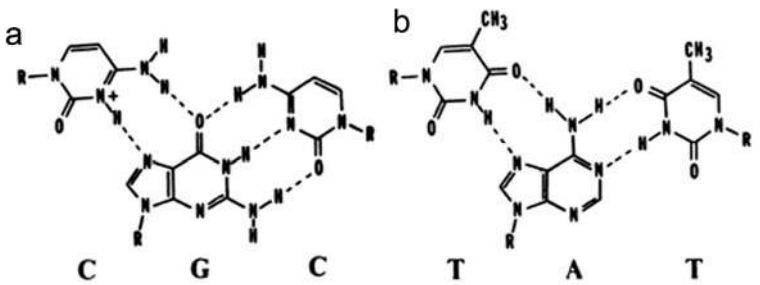

Figure 7. Base pairings in the triple helix. Hydrogen-bonding scheme of C·G·C^+ (a) and T·A·T (b) in the triple helix DNA. Adopted from Wells, R.D. *et al.* 1988. Faseb J[126].

S1-nuclease is regularly used for probing sites of triplex-formation, since it has an affinity for single stranded DNA. Through this, many PPy/PPu sequences have been mapped, most of them occurring in promoters of genes and thus suggested to be part of regulation of transcription[134]. They are also recognized as sites of translocation breakpoints involved in cancers and other disorders[135,136]. One of these disorders is Friedrich's ataxia, where two triplex structures are formed that can interact and form a bitriplex, called sticky DNA[137]. H-DNA or triplex-DNA is a dynamic conformation, so that for this structure to occur there is need of a stabilizing mechanism. *In vitro* this is achieved by low pH and different salt-concentrations. Although physiological pH can lead to formation of triplex[131], there might be other ways of stabilization *in vivo*, *e.g.* RNA-binding[138], methylation[139], and protein-binding.

DNA sequence polymorphisms

There are diverse types of variations in the human genome, ranging from changes in one base pair up to several Mb[140]. Variations in a certain position of a sequence, which is present on >1% of the alleles in the population, are called polymorphisms. Whether such variations lead to recognizable phenotypes or have no effect depends on the sequence involved. The phenotypic outcome of these changes is either a disease or a change in quantitative traits. Polymorphisms can be two or more variants at a single nucleotide level, single nucleotide polymorphism (SNP), or a short or long tandem repeat of one to several bases, called micro-, mini-, and major satellite, depending on the type of repeat[141]. SNPs and microsatellites are the two polymorphisms that have caught much attention, due to their high abundance in the human genome. At first, they were great tools in the search for genes causing diseases by linkage and association studies, but subsequently themselves became a subject of investigations for possible phenotypic effects.

Microsatellites

Microsatellites are tandem repeats of nucleotides with less than 10 bp iterated unit, which can have different base compositions and be hypervariable[141,142]. It is established that long forms of some microsatellites in the coding sequences may lead to genetic instability and thus cause disorders, *e.g.* Huntington's disease (OMIM+143100) and Fragile X-syndrome (OMIM+309550). Despite accumulating evidence of the role of microsatellites in gene regulation, this subject is still a controversial issue[143-147]. Different ways of participation are found depending on the location, sequence and size of the repeat[148]. Mechanisms such as changing the relative position of surrounding TFBS[149,150], transcription arrest[151,152], nucleosome exclusion[153],

and epigenetic silencing[154] has been suggested. One of the most exciting findings is that microsatellites are transcribed and form double stranded RNA (dsRNA) that act as templates for small interfering RNAs (siRNA). This was shown the first time for a CTG-repeat in the 3'-UTR of the *DMPK* gene[155]. Expansion of this repeat is a cause of myotonic dystrophy (DM1). However, transcription of repeats and their involvement in heterochromatin formation is not unique to microsatellites, but has also been found in pericentromeric repeats of chromosomes[99]. As mentioned above, repeats can act as regulatory elements through formation of alternative structures than the classical double helix.

Other means by which microsatellites are proposed to participate in transcriptional regulation is TF binding[148]. If the repeat contains a TFBS, different repeat lengths might allow binding of one or more of the actual factor or of different factors depending on the composition of bases, which would result in differential response depending on allele. Some of the described microsatellite binding proteins are: CGCBP[156] and hnRNPs (described above).

The regulatory sequence of inducible nitric oxide synthase (*iNOS* or *NOS2*) contains two microsatellites, a $\{TAAA\}_{3/4}$ (rs12720460) at -752 bp[157] upstream of transcriptional start site, and a $\{CCTTT\}_{8\text{-}17}$ (rs3833912) at -2.6 kb[158]. Additionally, a 92 bp PPy/PPu sequence (here called the CT-repeat) with a similar base composition as the CCTTT-microsatellite is located in intron 1 of this gene. Alleles of the {TAAA} bind protein and lead to different activities for the promoter, but only one study have found an association between this repeat and complications of type 2 diabetes[159].

The CCTTT-microsatellite is neither found in the mouse nor rat *Nos*-promoters, but has been genotyped in chimpanzee, gorilla, orangutan, and macaque[158]. Orangutans and macaques had a monomorphic repeat, while gorillas and chimpanzees were polymorphic, with gorillas having equal numbers of alleles as humans. As expected, the frequency of alleles in humans differs between ethnically diverse populations[160]. In Europe, distribution of alleles is unimodal with 11, 12, and 13 as the more common repeats; while in Africa, the distribution is bimodal with 9, 10, and 13 being the more common alleles.

Several association studies have been performed in order to detect possible connections of the CCTTT-microsatellite to different disorders, some of them with positive associations as result[161-164]. The microsatellite has not been tested for its protein binding capabilities, but has been investigated for its role in inducibility of *NOS2*. In two separate studies, a few alleles of the microsatellite were tested in constructs containing the *NOS2* promoter transfected into DLD-1 (colonic carcinoma cell line) and fibroblasts, and induced with interleukin-1β (IL-1β)[165,166]. In both experiments, the $\{CCTTT\}_{14}$-allele induced the reporter gene significantly more than the other alleles. However,

regulatory sequence constructs of *NOS2* with a large portion deleted, including the microsatellite, have retained the same activity as constructs without a deletion[167,168]. Differences in the aforementioned studies that have to be considered before conclusions can be drawn are the usage of different cell lines and inductions in these experiments. Furthermore, a 2.6 kb deletion in the regulatory sequences will probably affect the composition of regulatory elements and thus may not reflect the true nature of regulation for this gene.

SNPs

Single nucleotide polymorphisms is the most common type of variation in the human genome, with an average incidence of one in 200 bp in total and once in 1000 bp in an individual (NCBI dbSNP, Build 129). SNPs can be divided in two major groups, coding and non-coding, depending on their position in the gene. Coding SNPs usually influence the encoded protein[169], while the non-coding can affect splicing[170] or transcriptional regulation of the gene, or may be silent. As for microsatellites, there are an increasing number of studies indicating that SNPs can influence transcriptional regulation. Among these the total abolishment of binding has been demonstrated for RUNX-1 in an intron of *PDCD-1* in the SLE-disease[171]. Furthermore, a one base change in the binding motif of NF-κB has been found to effect the cofactor interactions[172].

Depending on the population, different sets of SNPs can be detected in the human genome. Some of these are private for each particular population and some are common. The proximal regulatory sequence of *NOS2* has been studied in relation to cardiovascular disorders[173], hepatitis C virus infection[174], and most extensively malaria[175-177]. No single SNP has yet been associated with the etiology of a disease; however, haplotypes of the *NOS2* promoter seem to be related to some aspects of malaria[175] and hepatitis C virus infection[174]. In the former study, a C/T SNP at position -1659 was tested for protein binding by electrophoretic mobility shift assay (EMSA), where the C-allele bound more proteins than the T-allele. The significance of this finding is unclear.

Nitric oxide synthase

Nitric oxide (NO) is an invaluable molecule in the human body. Due to its small size, it can easily and rapidly diffuse through cell membranes and thus function as a signaling molecule, which is utilized in the central and peripheral nervous system[178], by the vascular endothelium in regulation of blood pressure[179], in muscles throughout the body and specifically in cardiac muscles[180], regulation of apoptosis[181], and modification of transcriptional regulators[182,183]. Furthermore, NO participates in the immunological defense, *e.g.*

against microbial attack and some cancers[184]. Due to the reactivity of NO, it can also function as oxidizing agent or form toxic compounds. This has lead NO to be suspected or proven for participation in the pathology of several diseases[180] such as Alzheimer's[185,186], Parkinson[187], multiple sclerosis[188], systemic lupus erythematosus[189], malaria[190], type-1 diabetes[191], diabetic retinopathy[165], and glaucoma[192-194]. The variety in function and the potential to participate in or cause disorders gives an understanding of the significance of expressional regulation of this molecule.

Nitric oxide is produced by three isoforms of nitric oxide synthase (*NOS*). Two of these isoforms are constitutive and dependent on Ca^{2+} and Calmodulin, while the third is inducible and not Ca^{2+}-dependent[195]. Upon activation, the constitutive NOS immediately start producing a small amount of nitric oxide and do so under a longer period, whereas the inducible NOS starts later and produces a higher amount of NO under a shorter time. Each of these three is encoded by a specific gene. The constitutive forms are coded by neuronal *NOS* (*NOS1 or nNOS*) and endothelial *NOS* (*NOS3* or *eNOS*).

Inducible/immunological *NOS* (*iNOS*) or *NOS2*, the third gene in the trio, has two non-functional pseudogenes *NOS2P1* (*NOS2C*) and *NOS2P2* (*NOS2B*)[196]. All members of *NOS2*-family are on chromosome 17, with the *NOS2* positioned at 17q11.2[197,198]. It is a 43.76 kb long gene with 27 exons and 26 introns, a non-coding first exon, a near-consensus TATA-box 24 bp upstream of TSS, and translation initiation site in the middle of exon 2. There are reports of alternative transcription initiation sites upstream of the TATA-box, and alternative mRNA splicing[199,200]. Studies so far has characterized a 16 kb long region containing regulatory sequences[168]. Regulation of expression is mainly at the transcriptional level for this gene, with both the promoter and the 3'-UTR involved[201]. A number of TFBS have been identified in the regulatory sequence, most of them responding to cytokine signaling. These cytokines are mainly involved in inflammatory- and cellular stress responses, each inducing a specific response depending on the cell type[202-205]. The three most common cytokines for induction of *NOS2* are IL-1β, IFN-γ (interferon-γ), and TNF-α (tumor necrosis factor- α). One of the important findings is the characterization of a putative enhancer element between -5.2 and -6.5 kb upstream of TSS with five NF-κB sites recognized[167,206]. Furthermore, a negative regulatory element (NRE) has been identified in vicinity of the enhancer, located at -6.7 kb, that binds NF-κB-repressing factor (NRF)[207]. This is the first time an NRE is located at such far distance from the core promoter, once again indicating a long regulatory sequence for the *NOS2*. NRF exerts its effect through direct interaction with NF-κB. NF-κB-binding in the enhancer region is suggested to induce a three dimensional fold of the DNA, which may both cause an effective transcription and influence the interactions between NRF and the NF-κBs. This is though far from a complete picture of the elements involved in regulation of this gene, since it is surrounded by ca 175 kb of non-coding sequences.

NOS2 is also activated by many different stimuli, so it is therefore likely that this gene has an unusually complex regulation.

Nitric oxide is a common signaling and defense molecule in eukaryotes, bacteria, and plants[208], produced by the same type of enzyme(s) as in human. In 1997, there was a report of 120 laboratories working with different mouse models of *Nos*-knock-outs[209,210]. The mouse and rat homolog of human *NOS2* have 80% nucleotide- and amino acid sequence identity, which is down to 50% at the promoter level[197,211]. The murine promoter is constituted of two conserved clusters, region I (-48 to -209) and II (-913 to -1029), out of which region I seems to be conserved in the human promoter[212,213].

Present investigations

Aims of the present studies

Papers I-II

To investigate how polymorphisms can affect transcriptional regulation of genes and how they are involved in etiology of diseases. As a model, variations in the promoter of *NOS2* were studied in relation to primary open angle glaucoma. Focus of the experiments was the CCTTT-microsatellite in the regulatory sequences of *NOS2* and how it could affect the regulation of this gene.

Papers III-V

To identify all the putative binding sites for liver specific TFs and other interacting factors in the human genome in order to establish the transcriptional regulatory circuitry in the liver. As a feasibility study, some of the most well known TFs, *i.e.* SREBP1, HNF4α, GABPα, FOXA1, FOXA2, and FOXA3 were examined in HepG2 cells.

Materials and methods

Papers I-II

Subjects

The studied individuals were recruited from the glaucoma clinic at the Department of Ophthalmology, University Hospital, Uppsala, and the Department of Ophthalmology, Tierps Hospital, Tierp in Sweden. Inclusion criteria for the 200 patients were increased intraocular pressure (IOP) and glaucomatous damages to the optic nerve head and/or glaucomatous damage to the visual field. In the patient group, 99 were male with an average age at first diagnosis of 65.1 [range 27 – 90] and 101 female with an average of 66.5 [range 39 – 91]. In total, 50 had at least one case of glaucoma in the family. The mean pressure for the right eye of patients was 30.9 [range 18 – 57, median 29] and for the left eye 31.3 [range 19 – 59, median 30]. The threshold for elevated IOP is 24. For the control group 200 individuals were matched with the patients for age, sex, and geographic and ethnic origin. Glaucoma was excluded in this group by IOP-measurements and ophthalmoscopy of the optic disc. None of the studied individuals were related. Random population samples were DNA from 204 blood donors in Uppsala. There is no information regarding age or geographic and ethnic origin of the donors or about any eventual hereditary diseases for these individuals. Informed consent was obtained from all participants. This study was approved by the local Research Ethics Committee at Uppsala University and performed according to the Declaration of Helsinki.

Genotyping length polymorphisms

The used genotyping method is based on a separation on a denaturing polyacrylamide gel of differentially sized PCR-products, representing different alleles of a particular polymorphic marker. A sequence containing the studied length polymorphism was amplified by PCR, with fluorescently labeled primers. Amplicons were analyzed on a 4% polyacrylamide gel on ABI Prism™ 377 DNA Sequencer and the lengths were compared to a size marker.

SNaPshot™

This is a multiplex mini-sequencing method, applied for genotyping of SNPs. Oligonucleotides are annealed adjacent to the target SNP and then extended with a fluorescently labeled di-deoxy nucleotide (ddNTP) by a polymerase. The nucleotide incorporated, identifies the two alleles of the individual. Hereafter the excessive nucleotides are removed and the sample is electrophoresed and detected *e.g.* in a capillary sequencer.

Sequencing

Sanger sequencing is the widely used method for "reading" the order of nucleotides. Here primers annealed to a target sequence is extended with dNTPs until a labeled ddNTP is incorporated, resulting in a ladder of fragments. These fragments can then be separated on a gel or in a capillary sequencer and the order is detected. In this study, the DYEnamic™ ET dye terminator cycle sequencing kit was used together with MegaBACE™ 1000.

Haplotype reconstruction

A haplotype is a specific composition of alleles on a single chromosome. Haplotypes can be constructed for one gene or for a larger genomic area and is based on the polymorphic markers found within this gene or area in an individual. For reconstruction of haplotypes, statistical programs are applied to the genotyping data. Haplotype reconstruction program PHASE was utilized for our data.

Electrophoretic mobility shift assay, EMSA

EMSA is an *in vitro* approach to examine whether a specific sequence interacts with proteins or a certain nuclear factor. For this, double stranded oligonucleotides with the studied sequence are radioactively labeled and incubated with nuclear extract of a cell type or recombinant purified proteins and then analyzed on a native polyacrylamide gel. Protein-binding to the oligonucleotide results in a large complex, which has a lower mobility in the gel and thus gives rise to a band shifted in position compared to the naked oligonucleotide. The specificity of this interaction can then be tested by various methods, *e.g.* competition with unlabeled oligonucleotide, competition with unrelated oligonucleotide, mutation in sequence of oligonucleotide, and use of an antibody against the suspected nuclear factor. In the latter case, if the protein is recognized by the added antibody, an even larger complex will arise, which has an even lower mobility than the oligonucleotide-protein complex and will therefore be supershifted.

This method was also applied in paper III.

S1-nuclease assay

S1-nuclease recognizes single strands that may arise in alternative structures of DNA and makes one cut in that strand. As a result the nicked molecule is relaxed and the complementary strand exposed and can therefore be cut by S1. For circular plasmids, this treatment often results in linearization of the plasmid.

Papers III-V

Material

For the purpose of these studies, the human hepatocellular carcinoma cell line HepG2 was chosen, which originates from a Caucasian male. HepG2 cells have been shown to be good models for genome-wide localization studies, since in one study 66% of the binding sites for HNF6 in primary liver were also found in HepG2 cells[214].

Cells were grown in RPMI1640 complemented with non heat-inactivated 10% fetal bovine serum and 2% L-glutamine in 37°C at 5% CO_2 for all studies, unless otherwise stated.

ChIP, ChIP-chip, and ChIP-seq

Chromatin immunoprecipitation is one of the most widely used methods in studies of transcriptional regulation and epigenetics, since it allows the researcher to get a snapshot of the ongoing protein-DNA interactions *in vivo*. Generally, cells or tissues are treated with a cross-linking agent, *e.g.* formaldehyde, to stabilize the protein-DNA interactions in the cell (Figure 8). The final concentration of the agent and the time length of treatment is an important step; otherwise there is a risk of higher background due to extensive cross-linking. In studies of histone modifications, cross-linking is not required due to the tight compaction of DNA around histones and non cross-linked chromatin might even result in lower background.

The chromatin is then isolated and sheared to smaller fragments usually by sonication or treatment with micrococcal nuclease I (MNase I). MNase I cuts double stranded DNA at nucleosome free regions and if chromatin is fully treated, mononucleosome sized DNA (147 bp) can be obtained. The shearing is another central step in this protocol, since the size of DNA fragments directly affects the final resolution irrespective of final detection technique.

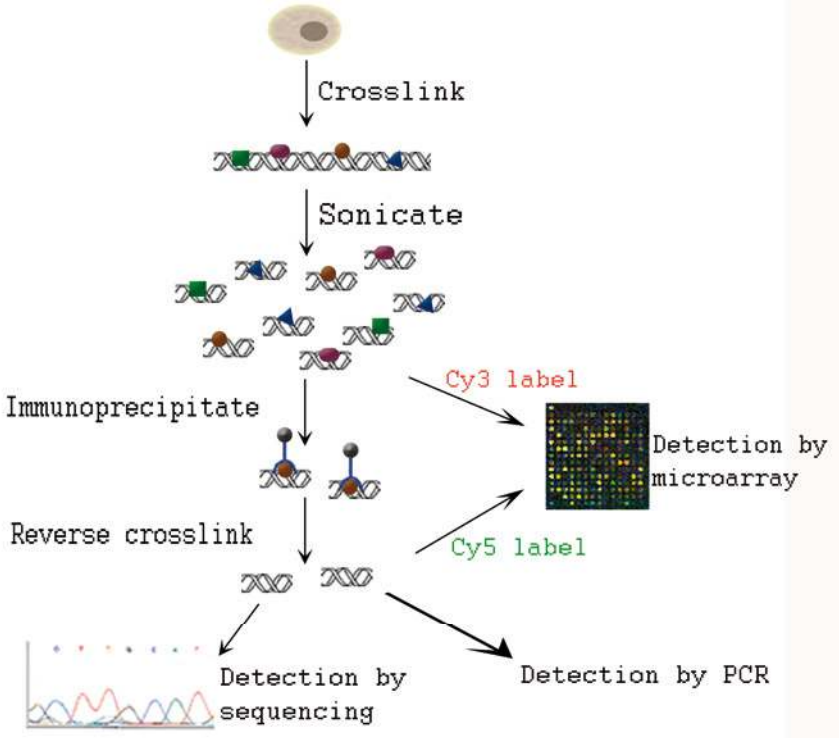

Figure 8. Chromatin immunoprecipitation (ChIP). Cells or tissues are treated to cross-link the interacting proteins with the DNA. The chromatin is sheared and the studied protein is immunoprecipitated by an antibody-solid phase complex. After reversal of cross-linking, the DNA is purified and can be detected by PCR (ChIP-PCR), microarray (ChIP-chip), or by sequencing (ChIP-seq).

Thereafter, the DNA-protein complex is immunoprecipitated with the antibody of choice coupled to a solid phase. The antibody used for the study should be evaluated prior to use by Western blot, siRNA, or other means to ensure the specificity of antibody, although, this does not guarantee the quality of the ChIP. Furthermore, in most of the studies polyclonal antibodies are preferred. However, this is not always suitable if there are multiple isoforms of the target protein with high homology at the antigenic site.

The DNA-protein-antibody-solid phase complex is washed extensively to remove all unbound antibody and DNA-protein complexes, as well as other debris. Subsequently, the DNA-protein complex is precipitated from the beads and the cross-linking is reversed, before removal of all proteins by proteinase-K treatment. Finally, the DNA is phenol-chloroform extracted or purified through columns. This DNA is usually referred to as *IP*. A fraction of the chromatin is saved before the immunoprecipitation and later treated as the IP from the reversal of cross-linking step. This sample is usually referred to as *input*.

There are alternative ways of how the resulting DNA can be studied. While detection by semiquantitative or quantitative PCR (ChIP-PCR) is the method of choice for single gene- and small-scale studies or verification of results from large-scale studies, it is not suitable and can be cumbersome for comprehensive studies. Here, the choice is between detection with microarrays (ChIP-chip) or with the new generation massively parallel sequencers (ChIP-seq). While there are a large number of studies with ChIP-chip, microarrays are losing ground to the new MPSs due to the cost-effectiveness and simplicity of genome-wide studies using the sequencers.

The variety of microarrays available for ChIP-chip is large. These can be PCR or oligo based tiling promoters, CpG-islands, or the whole genome with very low or very high resolution. The arrays used in paper III were PCR based covering 1% of the human genome as defined by the ENCODE consortium[4]. In these arrays, the repeat regions were included. The hybridization of samples is performed with both the IP and the input, each labeled with a fluorescent dye, usually Cy5 and Cy3. A region is defined as enriched when the signal from the IP exceeds that of the input and is generally expressed as a ratio (Figure 9).

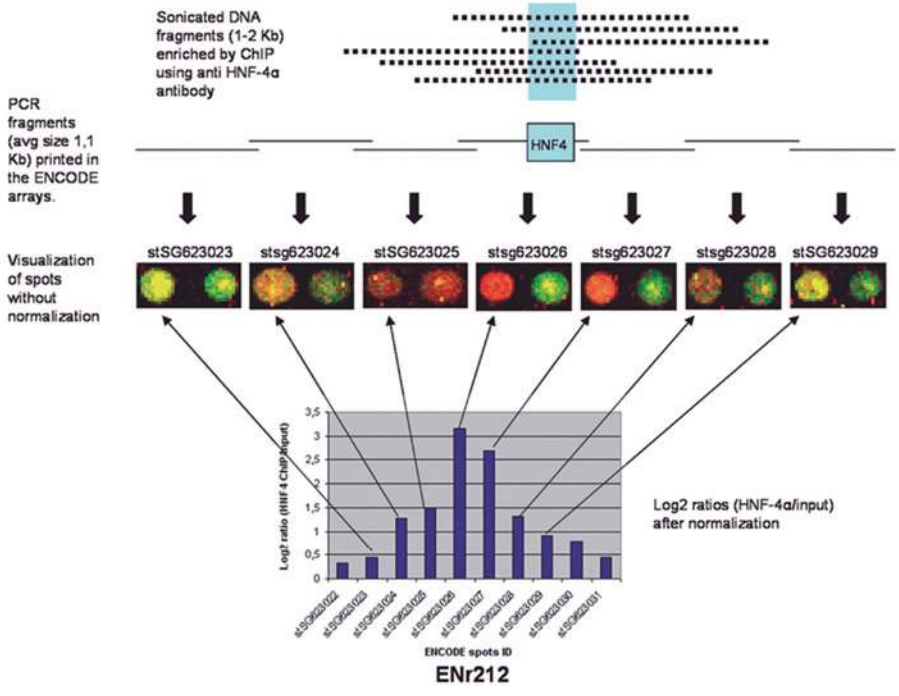

Figure 9. Definition of an enriched spot with a microarray. Enriched signals of IP are seen as red spots, while green is the signal from input. Each spot represents one fragment on the microarray. Enrichment is calculated as $\log_2$-ratio for IP signal over input. The spot with the highest signal in a distinct region of positive peaks is defined as the unique enriched spot (UES). Here, the UES is the fifth bar from left. Adopted from Rada-Iglesias, A. *et al*. 2005. Hum Mol Gen[75].

At the time of performance of these studies, there were three MPSs available: 454® from Roche, Solexa/Illumina from Illumina, and SOLiD™ from Applied Biosystems. All three sequencers are based on attachment of single stranded DNA on a solid phase, either beads or a surface, and clonal expansion of each DNA. These are then sequenced by a sequencing-by-synthesis (Illumina) or by ligation (SOLiD™) approach using fluorescent dyes. In ChIP-seq only the IP sample is needed for detection of enrichment. However, it is advisable to also sequence the input sample, which then can be used as a negative control. The generated data consists of short (25-50 bp for Illumina and SOLiD™) or longer (400 bp 454®) reads, which can be aligned to the genome using different bioinformatic tools. Regions in the genome with multiple reads correspond to enrichment and thus a putative TFBS (Figure 10). In paper IV the IPs were sequenced by Illumina, while in paper V the IPs were detected on a SOLiD™.

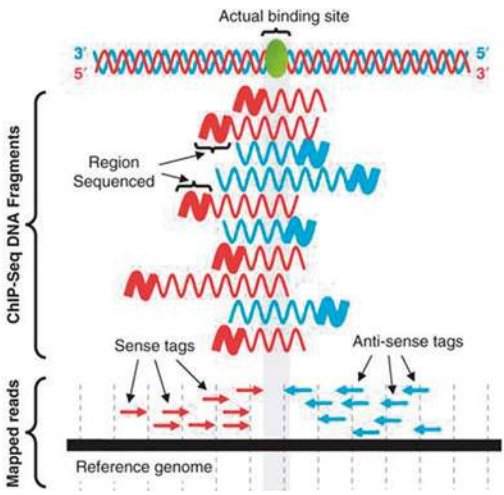

Figure 10. Defining a putative binding site by ChIP-seq. The sequenced end-tags of enriched fragments in the IP are aligned to the reference genome. Sites of TF binding are defined as regions with a distinct number of reads from the sense and anti-sense end-tags. Adopted from Jothi, R. *et al.* 2008. Nucl Acid Res[215].

ChIP-PCR was also applied in paper II.

Co-immunoprecipitation

Co-IP is a method for the study of protein-protein interactions *in vivo*. For this, cells are lysed and the solid phase coupled antibody targeting one of the proteins is added. The protein-antibody-solid phase is washed with different stringencies depending on the studied complex. Thereafter the sample is boiled and separated on a polyacrylamide gel for a Western blot. In the immunoblot, the antibody targeting the second protein in the complex can now be used to detect possible interactions.

Results and discussion

Paper I

In vitro and *in vivo* studies have indicated *NOS2* as a candidate for causing glaucoma, but no genetic studies have been performed[204,216,217]. Glaucoma is the common name for a heterogeneous group of neuropathological ophthalmic disorders involving retinal ganglion cell (RGC) death, optic nerve damage and visual field loss, starting from the periphery, and is one of the common causes of blindness in the world[218]. The majority of glaucoma patients are diagnosed with primary open angle glaucoma (POAG) which is a genetically complex disorder with late age at onset and with elevated intraocular pressure (IOP) in most cases[219].

To perform this association study, 200 POAG-patients and 200 age-matched controls were genotyped for the CCTTT-microsatellite. Distribution of alleles in both groups were found to be in concordance with the unimodal pattern found in the European population[160], but there was a difference in the overall distribution between patients and controls [$\chi^2 = 18.456$, DF = 7, P = 0.0101]. Furthermore, the 14-allele of the microsatellite, $\{CCTTT\}_{14}$, had a higher frequency, while $\{CCTTT\}_{13}$ had a lower frequency in POAG-patients, compared to age-matched controls. To verify this difference in allelic distribution, 204 randomly chosen individuals, here called random population, were genotyped for the same microsatellite. In agreement with these data, the frequency of the alleles differed between random population and POAG-patients. There was no difference between age-matched controls and random population.

Other polymorphisms were also genotyped in the POAG-patients and age-matched controls. The $TAAA_{3/4}$ and a GTGTGTT insertion/deletion had an equal distribution in both groups. Sequencing of 5.7 kb, which includes intron 1, exon 1 and ~4 kb of regulatory sequences upstream of TSS, lead us to find 8 SNPs, out of which 4 had not been reported at the time of this study. These SNPs were also genotyped and haplotypes were constructed, however no difference could be detected between POAG-patients and age-matched controls.

These results lead us to hypothesize that allele(-s) of the CCTTT-microsatellite are associated with primary open angle glaucoma. Since some of the alleles of this microsatellite have been reported to induce an activity for a reporter gene with differential strength depending on the number of

repeats[165,166], we decided to investigate how the CCTTT may exert its effect. For this an EMSA was performed with an oligonucleotide containing only the repeat. Using HeLa-nuclear extracts, we could demonstrate a binding of nuclear factors to the double stranded microsatellite. Based on the consensus sequence, possible candidates were predicted, yet none of these candidates led to a positive finding. We could though establish that the binding factor did not contain a Zn-finger motif.

The fact that the CCTTT-microsatellite bound nuclear factors, although *in vitro*, and that it had been demonstrated to induce an activity independent of its position, lead us to suggest that the microsatellite may be involved in regulation of inducible nitric oxide synthase by acting as an enhancer. Another way of exerting a regulatory effect, that we suggested, was facilitation of interaction between elements in the vicinity of the microsatellite.

Paper II

Based on what was found in paper I, *i.e.* a difference in distribution of the CCTTT-microsatellite alleles and protein binding for the repeat, we decided to further investigate what factor(-s) that interact with this microsatellite and thereby get an insight into how this microsatellite functions.

Polypyrimidine tracts have been shown to form triplex structures. Therefore, we investigated whether the microsatellite and its surrounding sequence was capable of assuming such conformation. A PCR was performed, to ascertain the double-strandedness of the region. To this PCR-product single stranded oligonucleotides of CCTTT or the complementary AAAGG were added in increasing gradients and analyzed on a polyacrylamide gel after an overnight incubation in a buffer with neutral pH and polyamines. Here it was found that the AAAGG, but not the CCTTT-oligonucleotide, could interact with the double stranded amplicon. Addition of these oligonucleotides to a PCR fragment with the CT-repeat did not cause any upshifted bands. This led us to suggest that the PCR form a triplex composed of G·G·C.

Since longer sequences of PPy have been demonstrated to form triplex even without being positioned in a supercoiled molecule[220], the PCR-products were incubated with different concentrations of S1-nuclease and analyzed on a gel. As expected, low mobility bands other than the amplicon were diminished upon treatment. PCR-products of the CT-repeat and the TAAA did not contain any extraneous bands and were not affected by the normal treatment with S1-nuclease. This indicates that the PCR-product of the CCTTT-microsatellite can form a triplex structure *in vitro* without any constraints.

The ability of the repeat to form a triplex was also tested in supercoiled conditions. Two different constructs of the *NOS2* promoter were prepared, pNOSP and pNOSPEI, where the former contained ~4 kb of the regulatory

sequences, while the latter in addition contained exon and intron 1. Homozygous individuals were used for cloning of the five most common alleles in the European population (alleles 10, 11, 12, 13, and 14 for the CCTTT-microsatellite) in the pNOSPEI-constructs. pNOSPEI was treated sequentially with S1-nuclease and a restriction enzyme with a single recognition site in the construct and the resulting fragments were analyzed on gel. For confirmation of results obtained with the first restriction enzyme (HindIII), a second enzyme (EcoRI) was used. These experiments all yielded the expected fragments, which indicate that the CCTTT-microsatellite forms a tetrastranded structure with a triplex and a single strand as core elements. These tests also showed extraneous bands for all five constructs, why we suspected that the CT-repeat in the intron 1 of *NOS2* forms a secondary structure and is thus cut by S1. To corroborate this, we compared digestion patterns of pNOSPEI and pNOSP, which resulted in less number of fragments for pNOSP. This further supported the hypothesis that PPy sequences in inducible nitric oxide synthase promoter adopt non-B DNA conformation. The S1-nuclease treatment approach for characterizing promoters of genes have been applied before[134]. In those studies, PPy-sequences in vicinity of TSSs proved to be susceptible to S1 and were therefore concluded to participate in transcriptional regulation of those genes.

There are several studies indicating members of the hnRNP family to bind single stranded DNA with a preference for polypyrimidine sequences[126]. The main function of these proteins in cells is to interact with newly transcribed mRNAs and participate in regulation of translation. In an EMSA, the single stranded CCTTT-oligonucleotide could bind nuclear factors of HeLa-extract and the factors were identified as PTBP1 and hnRNPK. To verify the findings, a ChIP was performed using HepG2 cells treated with a cytokine mixture of IFNγ, TNFα, and IL-1β. We found that the CCTTT-microsatellite did not bind PTBP1 or hnRNPK *in vivo*, while the CT-repeat was enriched for both factors in our ChIP. Furthermore, we found H3Ac, H4Ac and RNA polII in the CT-repeat region.

Based on our findings we concluded that both the CCTTT-microsatellite and the CT-repeat are capable of triplex-formation and that in the studied conditions the CT-repeat, but not the microsatellite, bound PTBP1 and hnRNPK *in vivo*. Therefore we indicated that the CT-repeat might play a role in the regulation of *NOS2*.

Paper III

Here we intended to map the binding sites of SREBP-1 in the ENCODE-regions as a feasibility study for genome-wide studies. We therefore performed a ChIP-chip in HepG2 cells. Due to the high turnover of SREBP-1 in cells, the HepG2 was treated with a proteasome inhibitor prior to the ChIP.

Furthermore, RNA polII has been demonstrated to be a good indicator of transcription in cells[221], why we decided to map this factor in HepG2 cells for correlation with SREBP-1 binding.

Three independent biological replicates were performed for each factor. Thereafter, a cut-off value was chosen and the unique enriched spots (UES) were defined. For SREBP-1 we found 45 UES and for RNA polII 143 UES in 1% of the human genome. For each of these, a set of putative binding sites were chosen for verification by ChIP-PCR. For SREBP-1 we could only verify 7 of 18 binding sites, while all 25 RNA polII sites were confirmed. We hypothesized that the high value of false positives were due to the high repeat content of SREBP-1 binding sites, which in combination with the repeat-containing microarrays used for this study may have created false positive spots.

The verified SREBP-1 binding sites were mostly positioned in vicinity of TSSs; however, three intragenic and intergenic sites could also be identified. Most importantly, two of these were located in the host cell factor C1 (*HCFC1*) and filamin A (*FLNA*) genes and coincided with RNA polII binding. We further demonstrated that SREBP-1 and RNA polII binding at both sites were sterol-regulated and we also identified the SREBP-1 binding site in *HCFC1*. HCFC1 interacts with E2F2 during G_1/S phase and recruits HMTs that target H3K4[222]. The other corroborated binding sites of SREBP-1 were positioned in TSSs of genes involved in lipid metabolism, *e.g. APOC3*. All genes reported in this study were novel binding sites.

From this study, we concluded that SREBP-1 participates in regulation of an expanded repertoire of genes, most interestingly genes involved in cell proliferation, and therefore is a suitable candidate for unbiased genome-wide studies.

Paper IV

In a recent paper it was shown that upstream stimulatory factor 2 (USF2), bound at distal elements and that this binding coincided with FOXA2 and HNF4α binding[20]. Moreover, the recognition sequence for GABP was frequently found in vicinity of USF binding. This led us to examine the whole genome binding sites for these factors. ChIPs were performed in HepG2 cells and the IPs were sequenced using the Illumina Genome Analyzer.

FOXA2 was found to bind 7253 sites in the human genome, HNF4α 18693 sites, and GABP 3060 sites. Out of these, 6.3%, 10%, and 85.2% were found within 500 bp of a TSS, respectively, indicating that a great majority of FOXA2 and HNF4α binding sites are located farther upstream of promoters. Additionally, GABP binding motifs were located closer to TSS than NRF1. Correlation of the TFBS with expression in HepG2 demonstrated that these factors are commonly associated with higher level of expres-

sion, however, while FOXA2 and HNF4α binding were correlated with higher expression in HepG2, GABP showed no such preference. Co-IP studies implicated interactions between FOXA2, HNF4α, and GABP. USF2 did not seem to interact, in spite of the finding that HNF4α, FOXA2, and USF2 were located in the same distal regulatory modules.

These findings suggest that FOXA2 and HNF4α mainly regulate uncharacterized transcripts or that they participate in long-distance regulation of genes by looping. None of these scenarios exclude the other.

Paper V

In this study, we aimed to understand the relation between the three members of the FOXA family, FOXA1, FOXA2, and FOXA3, as well as between the FOXA family and the active histone mark H3K4Me3. For this we performed genome-wide location analysis of FOXA1, FOXA3, and H3K4Me3 through ChIP-seq. For these experiments, the MPS SOLiD™ was employed.

We identified 8175 and 4598 putative binding sites for FOXA1 and FOXA3, respectively, and 41780 regions with H3K4Me3, corresponding to 160.000 nucleosomes. As expected, a majority of FOXA binding sites were located more than 1 kb away from a TSS, although FOXA3 had a higher percentage of binding within 1 kb compared to FOXA1. For H3K4Me3 42% of enriched regions were less than 1 kb from a TSS. We compared the H3K4Me3 regions with the available CAGE-tag data for HepG2; still we could not associate 28% of the regions with any transcript. Additionally, 13 binding sites for FOXA1, 8 for FOXA3, and 27 for H3K4Me3 showed a preferential binding to one of the two alleles in the genome, revealed by superior ChIP enrichment of one allele at heterozygous SNPs.

To understand the interrelationship in the FOXA family we compared the datasets for FOXA1, -2, and -3 binding and found 2304 regions in common. However, when we examined the protein-protein interactions by Co-IP, we found that FOXA2 interacts with both FOXA1 and FOXA3, but FOXA1 and FOXA3 do not interact. Therefore, we classified the TFBS for the three factors in pairs and found that 51% of FOXA2-FOXA3 bindings occur within 5 kb of TSS, while only 10% of FOXA2-FOXA1 complexes bind within the same distance.

By k-means clustering of H3K4Me3 signals around TSS, we made seven clusters. All seven clusters showed diverse distribution of signals and gene expression. The cluster with the highest expression in HepG2 was more enriched for H3K4Me3 than other clusters and 11% of the genes in the cluster had a FOXA3 binding site within 1 kb of the TSS. Furthermore, more than 30% of the genes in clusters I-III contained bidirectional promoters. We also found that at sites of FOXA1-2-3 binding, a double peak of H3K4Me3 surround the FOXA site. k-means clustering of these signals revealed

K4Me3 patterns reminiscent of those at the TSS. Looking at the TSSs within 5 kb of the triple binding sites, we found that 15% of these were associated with a CAGE-tag cluster.

In summary, we found that FOXA3 and FOXA1 have differential binding patterns in HepG2 cells and that they do not seem to interact *in vivo*. Furthermore, a majority of the FOXA1-2-3 bindings coincided with H3K4Me3 and we found that there might be a directionality of this histone modification at these sites, although no TSS is found in close vicinity. Finally, we suggest SNPs in combination with ChIP-seq as a tool for identification of monoallelic binding of TFs and enrichment of histone modifications.

Concluding remarks and future perspectives

In this thesis, I have presented genetic and genomic studies of transcriptional regulation in human cells. In the genetic studies, we specifically examined if and how repeat sequences participate in regulation of gene expression. Like other studies, we found that repeats may contribute by forming non-B DNA structures and by binding factors that may be involved in transcriptional regulation. In the genomic studies, we mapped the binding sites of six TFs important in liver metabolism, in addition to an epigenetic histone mark, and concluded that these TFs play a bigger role in cells than known before and that structure of regulatory sequences may be reflected in the epigenetic pattern at these sites.

The genetic view

We demonstrated that the CCTTT-microsatellite in the regulatory sequence of inducible nitric oxide synthase is potentially functional. We mainly base this conclusion on the *in vitro* ability of this microsatellite to form triplex and bind nuclear factors. This is further strengthened by the finding that another PPy/PPu tract in the regulatory sequence of *NOS2*, with a similar sequence as the microsatellite, formed triplex *in vitro* and bound regulatory factors *in vivo*. Our proposed model is that both the CCTTT-microsatellite and the CT-repeat can assume a triplex-DNA formation with part of the repeat in a triple-helix and a fourth strand unpaired. This single strand can then be recognized and bound by a complex of PTBP1 and hnRNPK.

In our system, we could not detect binding of the hnRNP family members at the microsatellite *in vivo*. This could be due to several reasons: 1) we have only examined one cell line and the microsatellite might not be active or not bind these factors in this tissue. 2) The microsatellite might not be responsive to the cytokine treatment and may require other signaling pathways to become active. 3) The microsatellite might be inactive at this stage of cell life and instead be more important at the earlier stages of the tissue development, for example before the cell has differentiated.

As reported, the CT-repeat in intron 1 of *NOS2* interacts with both PTBP1 and hnRNPK, but also RNA polII *in vivo*. For now, we do not have a clear picture of how this repeat might be involved in transcriptional regulation of the gene. However, a possible hypothesis is that cytokine signaling activates TFs, which bind the regulatory sequences of *NOS2* and engage histone modifying enzymes. These enzymes then modify the chromatin to make it ac-

cessible, which may permit for formation of a secondary structure in the CT-repeat opening the chromatin further. The single stranded DNA in the formed triple helix is then more accessible than the surrounding regions and makes an entry point for the GTFs and the RNA polII, allowing the CT-repeat to act as an alternative TSS. As mentioned in the introduction, there is evidence of alternative TSSs and alternative splicing of *NOS2* mRNA. Furthermore, hnRNP proteins are suggested to play a role in alternative splicing of pre-mRNAs, which strengthens this hypothesis further. Another possibility is that the CT-repeat acts as a TSS for antisense transcription.

An observation made in the genetic studies is the following: In our association study we found that alleles of the microsatellite were implicated in the etiology of POAG. Due to the high prevalence of SNPs, they appear to be the best markers for mapping disease-causing elements in the human genome. SNPs throughout the whole human genome have been genotyped in several populations to construct haplotypes[223]. The information obtained is then utilized for mapping the loci that contribute to all common diseases[224,225] and a range of other phenotypes[226,227]. The results of our study indicate that SNPs alone may not be powerful enough in all instances; rather a combination of SNPs and microsatellites could make a larger contribution to such studies. Microsatellites are usually more polymorphic than SNPs, which most often have two alleles, why one SNP-haplotype can be linked to different alleles of a microsatellite. In addition, microsatellites have a higher mutation frequency that further can break down the SNP-microsatellite linkage disequilibrium.

Polypyrimidine/polypurine tracts and other repetitive sequences are very common and spread in the human genome. If they indeed are able to adopt secondary structures and interact with proteins, they may have much larger impact on transcriptional regulation of genes in the human genome than is assumed today. In the future this could modify the perception of transcriptional regulation, since another level of complexity is added to what already is an intricate network of interactions between transcription factors and DNA, at the same time as it facilitates our understanding of how regulation of transcription functions in the human cell.

The genomic view

Looking at the genomic data for the TFs presented in papers III-V, we can divide the factors in two groups: distal element binding TFs (DEBTF) with HNF4α, FOXA1, FOXA2, and FOXA3 and proximal element binding TFs (PEBTF) with SREBP-1 and GABP. The former are factors that mostly bind at longer distances from TSSs, while the latter are factors that bind at or in vicinity of TSSs. This is obviously not a clear cut line, since we for example found that a considerable part of FOXA3 binding is within 500 bp of a PCG. As it appears in our data, the PEBTFs are restricted to always bind at promoters, since even sites that at first appear intragenic, are in fact novel TSSs.

DEBTFs on other hand are more promiscuous binders that seem to mostly participate in long-range interactions for regulation of target genes. The reason for that is unclear, although one could imagine different scenarios. One would be that these factors are crucial for expression of certain genes and therefore bind at proximal sites, while for other genes they function merely for raising the expression to higher levels than normal. If this is true, a knock-down of expression of DEBTFs may lead to complete silencing of genes with the factor at their promoter, while the other genes should only suffer loss of a part of the expression.

Another scenario could be that DEBTFs are binding at uncharacterized TSSs. Our current definitions of what a gene is and the techniques for transcriptome analysis used during the previous years, have restricted us in defining the transcriptional activity in our genome. However, with the current development of more sensitive molecular tools we might find that the transcriptome of the human genome is much larger than we can imagine. As mentioned in the introduction, recent findings in fact demonstrate that a significant part of the human genome is transcribed. With a better definition of the transcriptome, we might find that the DEBTFs are in fact PEBTFs of other crucial less-abundant transcripts. This is in part supported by our finding that FOXA-binding sites at distal elements have H3K4Me3 enrichment in close proximity. As data suggests today, H3K4Me3 is mostly enriched at TSSs.

An additional interesting finding was binding of SREBP in 3'-end of genes together with RNA polII, suggestive of presence of other promoters. As it was presented, these genes also had RNA polII binding at their known TSSs. Whether this 3'-end promoter results in shorter transcripts of the gene, in antisense transcription, or in competition with the main TSS is not yet known. Furthermore, a genome-wide location analysis of RNA polII in HepG2 cells is needed to get a better view of the on-going transcription in these cells and to understand if some of the intragenic bindings for the FOXAs and HNF4α might also be sites of new promoters. However, this is yet another strong implication that there is more transcription in the cells than meets the eyes now and that each gene in the human genome is certainly a scene for more than one player.

A technically important aspect of the work presented in papers IV and V is that with the current available methods and tools, we can refine the genetic information obtained by these high resolution studies. For TFs, it is now possible to study their interaction with DNA at resolutions less than a 100 bp, even down to tens of base pairs depending on the analysis method. This is shown in paper IV, where we can locate the TF binding motif in a majority of the putative binding sites. For histone modifications, the resolution is at the highest possible level, *i.e.* at nucleosome level, as presented in paper V.

The remaining question is: what is the use of genome-wide localization of TFBSs? Understanding the biology of transcription factors starts with understanding which processes these factors are involved in and with which other factors they interact. Certainly, this cannot be performed without combining this type of information with data from other genetic and epigenetic studies as well as transcriptome and proteome data. With this information in hand, we then can comprehend how our genes are regulated and how they function with the ultimate goal of understanding the biology of our genome. This not only benefits the scientific progression, but will undoubtedly help us to understand the etiology of human disorders and to design drugs that can target the point of "failure" while having mild side-effects.

And the future…

We live in an exciting time, when we are witnessing the development of genome biology in a faster rate than ever before. We have come far during these 55 years, since the determination of DNA structure, but there is still a long way to go.

What lies ahead? We need to understand how the ~1850 transcription factors function in each and every human tissue and how the regulatory circuitry in each tissue operates. We should certainly not forget that each of these transcription factors can carry one or a combination of different post-translational modifications, each affecting the function of the transcription factor. At the same time we need to understand the combinatorial pattern of histone modifications. Another component of transcriptional regulation, which has not been described deeply in this thesis, is the regulatory RNAs. Finally, we need to merge the information from these three wide areas, the transcription factors, the histone modifications, and the regulatory RNAs, to comprehend how the combinations of these constitute the *transcriptional regulatory code*.

What lies ahead is an exhilarating time, when we will witness astonishing findings beyond our imagination.

Acknowledgements

The road to and through the period of life as a PhD-student is long, rough and not always straight, with many ups and downs. Paving this road would not have been possible for me without the following people:

Professor **Ulf Gyllensten**, head of the department of Genetics and Pathology.

Professor **Claes Wadelius**, my head supervisor, who gave me the opportunity to stay in his group for such a long time. Thank you for the things you have taught me and all the scientific and non-scientific discussions. Someday we will visit Iran and ski there.

Associate professor **Gunnar Westin**, my co-supervisor, who I unfortunately had a few chances to discuss science with. Thank you for sharing your wisdom at those few moments.

Associate professor **Lena Åslund**, who created the opportunity for me to start in this project through her course.

I would also like to express my deepest gratitude towards the patients and other participants of the glaucoma project, who made those studies possible by blood donation. Parts of the projects in this thesis were funded by: Rönnows Stipendiestiftelse, Bergmarks Stipendiestiftelse, Rektors Wallenbergsfond, and Kungliga Vetenskapssamhällets forskningsfond.

All the members of the Wadelius group, in order of disappearance: **Benjamin Bakall**, **Mattias Jansson**, **Fredrik Aspgren**, and **Alvaro Rada-Iglesias** (thank you for everything you taught me, for all the scientific and non-scientific discussions, for sharing my occasional frustration, and all the fun we have had outside of Rudbeck), as well as current members **Ola Wallerman** (the only thing missing is a Cell paper), and **Madhu Sudhan Reddy Bysani** (for always being standby and helping me when I urgently needed it). **Kalicharan Patra**, a former member/student (for your enthusiasm). All the short/long time students (for putting up with me and my crazy ideas).

Jan Komorowski and the Komorowski boys at the Linnaeus center for Bioinformatics **Adam Ameur**, **Robin Andersson**, and **Stefan Enroth** for taking care of the data, interpreting them and making them understandable for me. Without you guys, I would have been lost in that world. I will try to stop being so picky with minor details☺.

Chandrasekhar Kanduri and his hard-working boys, **Faizaan Mohammad**, **Tanmoy Mondal**, **Radhamaran Pandey**, and **Gaurav Kumar Pandey** for introducing me to the world of RNA. It has been really nice to know you.

Johan Ericsson, head of the SREBP-fan club at Ludwig institute for cancer research, together with **Tanel Punga** and **Maria T Bengoechea-Alonso**. It has been a privilege to share this interest with you.

All the friendly folks at Uppsala Genome Center for your help and advice on all kinds and sorts of things, especially **Inger Jonasson** (I started my time at Rudbeck with your apparatus, I will finish my time here with your apparatus) and **Katarina Davidsson**. People at Applied Biosystems in Foster City and Beverly, USA, for running our samples. Also, **Jussi Vanhatalo** at the Nordic Applied, and **Jonathan Mangion** at Applied in UK. Other collaborators: **Ian Dunham** and **Christoph Koch** at Welcome Trust Sanger Institute; **Fredrik Pontén**, **Kenneth Wester**, **Adila Elobeid**, and **Ann-Catrin Andersson** at HPR, for providing us with all sorts of cells and some antibodies; **Agneta Lukinius**, for teaching me a few things about EM.

People in the slalom-room (in order of appearance): **Anh-Nhi Tran**, **Maja Nyström** (you'll do just fine), **Christina Edeby** (the doctor to be…), **Hanna Göransson**, **Patricia Respuela-Alonso** (my department of complaints ;-) thank you for always listening, helping me with ideas for the experiments, sharing chemicals and plastics, always being happy and optimistic, and for all the laughs), **Sara Ekvall** (the other coach in the room), and the all short-timers.

Professor **Niklas Dahl** and his crew **Jitendra Badhai** (you filled up the empty room after Hasse in the lab at nights), **Johanna Dahlqvist** (welcome to the club), **Ed Davey** (for all the discussions), **Miriam Entesarian** (I guess this is the first time after 10 years we'll go separate ways), **Ann-Sophie Fröjmark** (the best neighbor in the world), **Joakim Klar** (Jasså?!), **Hans Matsson** (see Jitendra's comment), **Malin Melin**, **Jens Schuster** (now I'm finished), **Maria Thuveson**, and **Hanna Örlén**.

Doktorandrådet: **Anna Wetterbom**, **Malin Engelmark**, **Hanna Edlund**, **Emma Ivansson**, and **Malin Jarvius**.

Other friendly faces at Rudbeck, among them **Anna-Karin Abelsson**, **Patrick Buckley** and **Sarah Walsh**, **Lucia Cavelier** (thank you for always being so supportive), **Tim Conze**, **Spyros Darmanis** (you wanna share the hair?), **Lioudmila Elfineh** (our SYBR source), **Fredrik Granberg** (time for you to test BMC), **Yin Hong**, **Mathias Howell**, **Max Ingman**, **Henrik Johansson**, **Sergey Kozyrev**, **Mårten Lindén**, **Fiona Murray**, **Martina Nilsson** (I haven't been in the gelroom for over 3 months now ☺), **Katerina Pardali**, **Ludmila Prokunina**, **Johanna Sandgren**, **Lena Spångberg**, **Mia Thorselius**, **Angelica Delgado Vega**, **Jenny von Salomé**, **Ammar Zaghlool**, **…** (if you don't see your name, magnify the dots with a microscope and you will find it☺). All the **administrators** for the great job you do.

Anna Hammarberg, **Sofie Mellberg**, and **Dan Edholm**, my good teaching colleagues at the CBB1-course.

People whom I would not survive Rudbeck without: **Mahmoud "Larry" Mansouri** (I am immensely grateful for your friendship), **Marina Tshuikina Wiklander** (without you Rudbeck would be so silent – and I mean it in a positive way! Thank you for always looking after me!), **Masood Kamali-Moghaddam** (Thank you for all the laughs and for not letting me making hesitated decisions), **Cecilia De Bustos** (Good luck to you and Todd in the South American adventures), **Antonia Kalushkova** (I will come back for yours, I promise), and **Gholamreza Tavoosidana** (you know what it takes to solve it…).

Birger Scholz for sharing your knowledge.

The **Green group** in Pedagogiska kursen. All the hardworking and enthusiastic people in **MDR** and **Doktorandnämnden**, as well as **NVF** and **SACO studentråd**. All of you in **ISF** (especially **Saeed Kamali-Moghaddam** and **Farideh Tarbiat**), **SIU** and **Yari**, who work non-stop for other people's well-being. Everyone at **KFUM Uppsala Basket** for keeping my head away from the lab from time to time. In particular **all the parents** for being extremely supportive and **all the players** in **Tigers** and **P93Röd** for being so understanding when I had much to do. You guys have given me some of my best memories. **Fredrik Blomqvist** and **Anders Lemmeke**, many thanks for being such a good colleagues.

Barbro Schröder, **Håkan Abrahamsson** (*QIP*) and **Krister Åkervall** for arousing my interest for science and research and for your support when I needed it the most. I am in great debt to all of you!

Viktor Leander (it's so nice to have contact with you again), **Ylva Lindman** (we both got where we wanted), **Dieter Fuchs** (for the physical and

verbal punches – keep them coming, you won't win anyway), **Siavoush Mohammadi** and family (for your incredible support, hospitality, and all the shared interest), **Feren** (for the long distance friendship and for the short trip in Iran).

Yvonne Niemi, you taught me how a good teacher should be.

All my relatives in Iran, US, and Canada, for your never-ending love, support, hospitality, and your interest for my work. I miss you all and share this with each and every one of you!

The other two musketeers, my dearest brothers **Reza** and **Hossein**, for all the help and support, for always being there for me especially during the most difficult times, and for all the great travels, which I hope will be many more of. Millions of thanks for everything! **Afsaneh**, for always having so close to laughter.

Last but not least, my beloved parents, **Ahmad** and **Vida**, for sacrificing your life in Iran for our well-being. I would not have come so far without your support and encouragement. You have taught me the value of endurance, hard work, and honesty. I hope I can at least make up for some of the million things you have done for me. Thank you with all my heart!

Uppsala, November 2008

References

1. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
2. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
3. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
4. Birney, E. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
5. Sandelin, A. et al. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* **8**, 424-36 (2007).
6. Sultan, M. et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956-60 (2008).
7. Kapranov, P., Willingham, A.T. & Gingeras, T.R. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet* **8**, 413-23 (2007).
8. Gerstein, M.B. et al. What is a gene, post-ENCODE? History and updated definition. *Genome Res* **17**, 669-81 (2007).
9. Smale, S.T. Core promoters: active contributors to combinatorial gene regulation. *Genes Dev* **15**, 2503-8 (2001).
10. Butler, J.E. & Kadonaga, J.T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* **16**, 2583-92 (2002).
11. Cooper, S.J., Trinklein, N.D., Anton, E.D., Nguyen, L. & Myers, R.M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res* **16**, 1-10 (2006).
12. Kadonaga, J.T. Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* **116**, 247-57 (2004).
13. Peterlin, B.M. & Trono, D. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nat Rev Immunol* **3**, 97-107 (2003).
14. Schug, J. et al. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol* **6**, R33 (2005).
15. Carninci, P. et al. The transcriptional landscape of the mammalian genome. *Science* **309**, 1559-63 (2005).
16. Carninci, P. et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-35 (2006).
17. Brodsky, A.S. et al. Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol* **6**, R64 (2005).
18. Singer, G.A. et al. Genome-wide analysis of alternative promoters of human genes using a custom promoter tiling array. *BMC Genomics* **9**, 349 (2008).
19. Trinklein, N.D. et al. An abundance of bidirectional promoters in the human genome. *Genome Res* **14**, 62-6 (2004).
20. Rada-Iglesias, A. et al. Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. *Genome Res* **18**, 380-92 (2008).

21. Lin, J.M. et al. Transcription factor binding and modified histones in human bidirectional promoters. *Genome Res* **17**, 818-27 (2007).
22. Maston, G.A., Evans, S.K. & Green, M.R. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**, 29-59 (2006).
23. Bell, A.C., West, A.G. & Felsenfeld, G. Insulators and boundaries: versatile regulatory elements in the eukaryotic. *Science* **291**, 447-50 (2001).
24. Kim, T.H. et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231-45 (2007).
25. Li, Q., Peterson, K.R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077-86 (2002).
26. Nobrega, M.A., Ovcharenko, I., Afzal, V. & Rubin, E.M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
27. Pabo, C.O. & Sauer, R.T. Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61**, 1053-95 (1992).
28. Wilson, M.D. et al. Species-specific transcription in mice carrying human chromosome 21. *Science* **322**, 434-8 (2008).
29. Spiegelman, B.M. & Heinrich, R. Biological control through regulated transcriptional coactivators. *Cell* **119**, 157-67 (2004).
30. Giandomenico, V., Simonsson, M., Gronroos, E. & Ericsson, J. Coactivator-dependent acetylation stabilizes members of the SREBP family of transcription factors. *Mol Cell Biol* **23**, 2587-99 (2003).
31. Lonard, D.M. & O'Malley B, W. Nuclear receptor coregulators: judges, juries, and executioners of cellular regulation. *Mol Cell* **27**, 691-700 (2007).
32. Rachez, C. & Freedman, L.P. Mediator complexes and transcription. *Curr Opin Cell Biol* **13**, 274-80 (2001).
33. Lewis, B.A. & Reinberg, D. The mediator coactivator complex: functional and physical roles in transcriptional regulation. *J Cell Sci* **116**, 3667-75 (2003).
34. Thomas, M.C. & Chiang, C.M. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41**, 105-78 (2006).
35. Smale, S.T. & Kadonaga, J.T. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**, 449-79 (2003).
36. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S. & Burd, C.G. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321 (1993).
37. Krecic, A.M. & Swanson, M.S. hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**, 363-71 (1999).
38. Brunel, F., Zakin, M.M., Buc, H. & Buckle, M. The polypyrimidine tract binding (PTB) protein interacts with single-stranded DNA in a sequence-specific manner. *Nucleic Acids Res* **24**, 1608-15 (1996).
39. Rustighi, A. et al. A polypyrimidine/polypurine tract within the Hmga2 minimal promoter: a common feature of many growth-related genes. *Biochemistry* **41**, 1229-40 (2002).
40. Ferguson, M., Henry, P.A. & Currie, R.A. Histone deacetylase inhibition is associated with transcriptional repression of the Hmga2 gene. *Nucleic Acids Res* **31**, 3123-33 (2003).
41. Swinburne, I.A., Meyer, C.A., Liu, X.S., Silver, P.A. & Brodsky, A.S. Genomic localization of RNA binding proteins reveals links between pre-mRNA processing and transcription. *Genome Res* **16**, 912-21 (2006).
42. Soderberg, M., Raffalli-Mathieu, F. & Lang, M.A. Regulation of the murine inducible nitric oxide synthase gene by dexamethasone involves a heterogeneous nuclear ribonucleoprotein I (hnRNPI) dependent pathway. *Mol Immunol* **44**, 3204-10 (2007).

43. Pautz, A. et al. The polypyrimidine tract-binding protein (PTB) is involved in the post-transcriptional regulation of human inducible nitric oxide synthase expression. *J Biol Chem* **281**, 32294-302 (2006).
44. Michelotti, E.F., Michelotti, G.A., Aronsohn, A.I. & Levens, D. Heterogeneous nuclear ribonucleoprotein K is a transcription factor. *Mol Cell Biol* **16**, 2350-60 (1996).
45. Postel, E.H., Mango, S.E. & Flint, S.J. A nuclease-hypersensitive element of the human c-myc promoter interacts with a transcription initiation factor. *Mol Cell Biol* **9**, 5123-33 (1989).
46. Choi, J. et al. Purification and characterization of cellular proteins associated with histone H4 tails. *J Biol Chem* **282**, 21024-21031 (2007).
47. Bomsztyk, K., Denisenko, O. & Ostrowski, J. hnRNP K: one protein multiple processes. *Bioessays* **26**, 629-38 (2004).
48. Oberstrass, F.C. et al. Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* **309**, 2054-7 (2005).
49. Tomonaga, T. et al. Unrestraining genetic processes with a protein-DNA hinge. *Mol Cell* **1**, 759-64 (1998).
50. Bengoechea-Alonso, M.T. & Ericsson, J. SREBP in signal transduction: cholesterol metabolism and beyond. *Curr Opin Cell Biol* **19**, 215-22 (2007).
51. Horton, J.D. & Shimomura, I. Sterol regulatory element-binding proteins: activators of cholesterol and fatty acid biosynthesis. *Curr Opin Lipidol* **10**, 143-50 (1999).
52. Brown, M.S. & Goldstein, J.L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331-40 (1997).
53. Farmer, S.R. Transcriptional control of adipocyte formation. *Cell Metab* **4**, 263-73 (2006).
54. Harada, N. et al. Alternative splicing produces a constitutively active form of human SREBP-1. *Biochem Biophys Res Commun* **368**, 820-6 (2008).
55. Felder, T.K., Klein, K., Patsch, W. & Oberkofler, H. A novel SREBP-1 splice variant: tissue abundance and transactivation potency. *Biochim Biophys Acta* **1731**, 41-7 (2005).
56. Toth, J.I., Datta, S., Athanikar, J.N., Freedman, L.P. & Osborne, T.F. Selective coactivator interactions in gene activation by SREBP-1a and -1c. *Mol Cell Biol* **24**, 8288-300 (2004).
57. Shimomura, I., Shimano, H., Horton, J.D., Goldstein, J.L. & Brown, M.S. Differential expression of exons 1a and 1c in mRNAs for sterol regulatory element binding protein-1 in human and mouse organs and cultured cells. *J Clin Invest* **99**, 838-45 (1997).
58. Raghow, R., Yellaturu, C., Deng, X., Park, E.A. & Elam, M.B. SREBPs: the crossroads of physiological and pathological lipid homeostasis. *Trends Endocrinol Metab* **19**, 65-73 (2008).
59. Yang, F. et al. An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature* **442**, 700-4 (2006).
60. Naar, A.M. et al. Chromatin, TAFs, and a novel multiprotein coactivator are required for synergistic activation by Sp1 and SREBP-1a in vitro. *Genes Dev* **12**, 3020-31 (1998).
61. Lee, Y.S. et al. Chromatin remodeling complex interacts with ADD1/SREBP1c to mediate insulin-dependent regulation of gene expression. *Mol Cell Biol* **27**, 438-52 (2007).
62. Reed, B.D., Charos, A.E., Szekely, A.M., Weissman, S.M. & Snyder, M. Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals

novel functional roles and combinatorial regulation of distinct classes of genes. *PLoS Genet* **4**, e1000133 (2008).
63. Schrem, H., Klempnauer, J. & Borlak, J. Liver-enriched transcription factors in liver function and development. Part I: the hepatocyte nuclear factor network and liver-specific gene expression. *Pharmacol Rev* **54**, 129-58 (2002).
64. Petrescu, A.D., Hertz, R., Bar-Tana, J., Schroeder, F. & Kier, A.B. Ligand specificity and conformational dependence of the hepatic nuclear factor-4alpha (HNF-4alpha). *J Biol Chem* **277**, 23988-99 (2002).
65. Hayhurst, G.P., Lee, Y.H., Lambert, G., Ward, J.M. & Gonzalez, F.J. Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. *Mol Cell Biol* **21**, 1393-403 (2001).
66. Ryffel, G.U. Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. *J Mol Endocrinol* **27**, 11-29 (2001).
67. Rosmarin, A.G., Resendes, K.K., Yang, Z., McMillan, J.N. & Fleming, S.L. GA-binding protein transcription factor: a review of GABP as an integrator of intracellular signaling and protein-protein interactions. *Blood Cells Mol Dis* **32**, 143-54 (2004).
68. Scarpulla, R.C. Nuclear control of respiratory gene expression in mammalian cells. *J Cell Biochem* **97**, 673-83 (2006).
69. Friedman, J.R. & Kaestner, K.H. The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci* **63**, 2317-28 (2006).
70. Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412-20 (1993).
71. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-70 (2008).
72. Odom, D.T. et al. Core transcriptional regulatory circuitry in human hepatocytes. *Mol Syst Biol* **2**, 2006 0017 (2006).
73. Duncan, S.A., Navas, M.A., Dufort, D., Rossant, J. & Stoffel, M. Regulation of a transcription factor network required for differentiation and metabolism. *Science* **281**, 692-5 (1998).
74. Wederell, E.D. et al. Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Res* **36**, 4549-64 (2008).
75. Rada-Iglesias, A. et al. Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. *Hum Mol Genet* **14**, 3435-47 (2005).
76. Cirillo, L.A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279-89 (2002).
77. Shim, E.Y., Woodcock, C. & Zaret, K.S. Nucleosome positioning by the winged helix transcription factor HNF3. *Genes Dev* **12**, 5-10 (1998).
78. Lee, C.S., Friedman, J.R., Fulmer, J.T. & Kaestner, K.H. The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944-7 (2005).
79. Goldberg, A.D., Allis, C.D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128**, 635-8 (2007).
80. Fraga, M.F. et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* **102**, 10604-9 (2005).
81. Fan, Y. et al. Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **123**, 1199-212 (2005).

82. Tremethick, D.J. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* **128**, 651-4 (2007).
83. Sarma, K. & Reinberg, D. Histone variants meet their match. *Nat Rev Mol Cell Biol* **6**, 139-49 (2005).
84. Berger, S.L. The complex language of chromatin regulation during transcription. *Nature* **447**, 407-12 (2007).
85. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
86. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-37 (2007).
87. Wang, Z. et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903 (2008).
88. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41-5 (2000).
89. Li, B., Carey, M. & Workman, J.L. The role of chromatin during transcription. *Cell* **128**, 707-19 (2007).
90. Schones, D.E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-98 (2008).
91. Shivaswamy, S. et al. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol* **6**, e65 (2008).
92. Sudarsanam, P. & Winston, F. The Swi/Snf family nucleosome-remodeling complexes and transcriptional control. *Trends Genet* **16**, 345-51 (2000).
93. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735-45 (2007).
94. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **293**, 1074-80 (2001).
95. Gilbert, N. et al. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555-66 (2004).
96. Morris, C.A. & Moazed, D. Centromere assembly and propagation. *Cell* **128**, 647-50 (2007).
97. Yang, P.K. & Kuroda, M.I. Noncoding RNAs and intranuclear positioning in monoallelic gene expression. *Cell* **128**, 777-86 (2007).
98. Grewal, S.I. & Elgin, S.C. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399-406 (2007).
99. Zaratiegui, M., Irvine, D.V. & Martienssen, R.A. Noncoding RNAs and gene silencing. *Cell* **128**, 763-76 (2007).
100. Qiu, J. Epigenetics: unfinished symphony. *Nature* **441**, 143-5 (2006).
101. Szabo, P.E., Tang, S.H., Silva, F.J., Tsark, W.M. & Mann, J.R. Role of CTCF binding sites in the Igf2/H19 imprinting control region. *Mol Cell Biol* **24**, 4791-800 (2004).
102. Jones, P.L. et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-91 (1998).
103. Nan, X. et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-9 (1998).
104. Suzuki, M.M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-76 (2008).
105. Weber, M. et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**, 853-62 (2005).

106. Wilson, I.M. et al. Epigenomics: mapping the methylome. *Cell Cycle* **5**, 155-8 (2006).
107. Saxonov, S., Berg, P. & Brutlag, D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412-7 (2006).
108. Weber, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* **39**, 457-66 (2007).
109. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-70 (2008).
110. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
111. Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-54 (2006).
112. Zhao, Z. et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38**, 1341-7 (2006).
113. Dostie, J. et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* **16**, 1299-309 (2006).
114. Tiwari, V.K., Cope, L., McGarvey, K.M., Ohm, J.E. & Baylin, S.B. A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res* **18**, 1171-9 (2008).
115. Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. & Fraser, P. Long-range chromatin regulatory interactions in vivo. *Nat Genet* **32**, 623-6 (2002).
116. Horike, S., Cai, S., Miyano, M., Cheng, J.F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* **37**, 31-40 (2005).
117. Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R. & Flavell, R.A. Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637-45 (2005).
118. Cook, P.R. The organization of replication and transcription. *Science* **284**, 1790-5 (1999).
119. Osborne, C.S. et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* **36**, 1065-71 (2004).
120. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**, 413-7 (2007).
121. Misteli, T. Beyond the sequence: cellular organization of genome function. *Cell* **128**, 787-800 (2007).
122. Akhtar, A. & Gasser, S.M. The nuclear envelope and transcriptional control. *Nat Rev Genet* **8**, 507-17 (2007).
123. Wells, R.D. Unusual DNA structures. *J Biol Chem* **263**, 1095-8 (1988).
124. Rich, A., Nordheim, A. & Wang, A.H. The chemistry and biology of left-handed Z-DNA. *Annu Rev Biochem* **53**, 791-846 (1984).
125. Ioshikhes, I.P. & Zhang, M.Q. Large-scale human promoter mapping using CpG islands. *Nat Genet* **26**, 61-3 (2000).
126. Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M. & Wohlrab, F. The chemistry and biology of unusual DNA structures adopted by oligopurine.oligopyrimidine sequences. *Faseb J* **2**, 2939-49 (1988).

127. Birnboim, H.C., Sederoff, R.R. & Paterson, M.C. Distribution of polypyrimidine . polypurine segments in DNA from diverse organisms. *Eur J Biochem* **98**, 301-7 (1979).
128. Johnson, D. & Morgan, A.R. Unique structures formed by pyrimidine-purine DNAs which may be four-stranded. *Proc Natl Acad Sci U S A* **75**, 1637-41 (1978).
129. Mirkin, S.M. et al. DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature* **330**, 495-7 (1987).
130. Kohwi, Y. & Kohwi-Shigematsu, T. Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proc Natl Acad Sci U S A* **85**, 3781-5 (1988).
131. Kohwi, Y. & Kohwi-Shigematsu, T. Structural polymorphism of homopurine-homopyrimidine sequences at neutral pH. *J Mol Biol* **231**, 1090-101 (1993).
132. Kohwi-Shigematsu, T. & Kohwi, Y. Poly(dG)-poly(dC) sequences, under torsional stress, induce an altered DNA conformation upon neighboring DNA sequences. *Cell* **43**, 199-206 (1985).
133. McCarthy, J.G. & Heywood, S.M. A long polypyrimidine/polypurine tract induces an altered DNA conformation on the 3' coding region of the adjacent myosin heavy chain gene. *Nucleic Acids Res* **15**, 8069-85 (1987).
134. Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. S1-hypersensitive sites in eukaryotic promoter regions. *Nucleic Acids Res* **12**, 8043-58 (1984).
135. Abeysinghe, S.S., Chuzhanova, N., Krawczak, M., Ball, E.V. & Cooper, D.N. Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs. *Hum Mutat* **22**, 229-44 (2003).
136. Bacolla, A. et al. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A* **101**, 14162-7 (2004).
137. Sakamoto, N. et al. Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. *Mol Cell* **3**, 465-75 (1999).
138. Boles, T.C. & Hogan, M.E. DNA structure equilibria in the human c-myc gene. *Biochemistry* **26**, 367-76 (1987).
139. Lee, J.S., Woodsworth, M.L., Latimer, L.J. & Morgan, A.R. Poly(pyrimidine) . poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res* **12**, 6603-14 (1984).
140. Sharp, A.J., Cheng, Z. & Eichler, E.E. Structural variation of the human genome. *Annu Rev Genomics Hum Genet* **7**, 407-42 (2006).
141. Chambers, G.K. & MacAvoy, E.S. Microsatellites: consensus and controversy. *Comp Biochem Physiol B Biochem Mol Biol* **126**, 455-76 (2000).
142. Ellegren, H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* **5**, 435-45 (2004).
143. Albanese, V. et al. Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite. *Hum Mol Genet* **10**, 1785-92 (2001).
144. Gharani, N. et al. Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. *Hum Mol Genet* **6**, 397-402 (1997).
145. Lapoumeroulie, C. et al. Genetic variations in human fetal globin gene microsatellites and their functional relevance. *Hum Genet* **104**, 307-14 (1999).
146. Yamada, N. et al. Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema. *Am J Hum Genet* **66**, 187-95 (2000).

147. Chiba-Falek, O., Kowalak, J.A., Smulson, M.E. & Nussbaum, R.L. Regulation of alpha -Synuclein Expression by Poly (ADP Ribose) Polymerase-1 (PARP-1) Binding to the NACP-Rep1 Polymorphic Site Upstream of the SNCA Gene. *Am J Hum Genet* **76**(2005).
148. Usdin, K. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res* **18**, 1011-9 (2008).
149. Lalioti, M.D., Scott, H.S. & Antonarakis, S.E. Altered spacing of promoter elements due to the dodecamer repeat expansion contributes to reduced expression of the cystatin B gene in EPM1. *Hum Mol Genet* **8**, 1791-8 (1999).
150. Alakurtti, K., Virtaneva, K., Joensuu, T., Palvimo, J.J. & Lehesjoki, A.E. Characterization of the cystatin B gene promoter harboring the dodecamer repeat expanded in progressive myoclonus epilepsy, EPM1. *Gene* **242**, 65-73 (2000).
151. Ditlevson, J.V. et al. Inhibitory effect of a short Z-DNA forming sequence on transcription elongation by T7 RNA polymerase. *Nucleic Acids Res* **36**, 3163-70 (2008).
152. Belotserkovskii, B.P. et al. A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J Biol Chem* **282**, 32433-41 (2007).
153. Wang, Y.H., Gellibolian, R., Shimizu, M., Wells, R.D. & Griffith, J. Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *J Mol Biol* **263**, 511-6 (1996).
154. Biacsi, R., Kumari, D. & Usdin, K. SIRT1 inhibition alleviates gene silencing in Fragile X mental retardation syndrome. *PLoS Genet* **4**, e1000017 (2008).
155. Cho, D.H. et al. Antisense transcription and heterochromatin at the DM1 CTG repeats are constrained by CTCF. *Mol Cell* **20**, 483-9 (2005).
156. Deissler, H. et al. Rapid protein sequencing by tandem mass spectrometry and cDNA cloning of p20-CGGBP. A novel protein that binds to the unstable triplet repeat 5'-d(CGG)n-3' in the human FMR1 gene. *J Biol Chem* **272**, 16761-8 (1997).
157. Bellamy, R. & Hill, A.V. A bi-allelic tetranucleotide repeat in the promoter of the human inducible nitric oxide synthase gene. *Clin Genet* **52**, 192-3 (1997).
158. Xu, W., Liu, L., Emson, P.C., Harrington, C.R. & Charles, I.G. Evolution of a homopurine-homopyrimidine pentanucleotide repeat sequence upstream of the human inducible nitric oxide synthase gene. *Gene* **204**, 165-70 (1997).
159. Morris, B.J. et al. Association of a functional inducible nitric oxide synthase promoter variant with complications in type 2 diabetes. *J Mol Med* **80**, 96-104 (2002).
160. Xu, W. et al. Survey of the allelic frequency of a NOS2A promoter microsatellite in human populations: assessment of the NOS2A gene and predisposition to infectious disease. *Nitric Oxide* **4**, 379-83 (2000).
161. Burgner, D. et al. Inducible nitric oxide synthase polymorphism and fatal cerebral malaria. *Lancet* **352**, 1193-4 (1998).
162. Kun, J.F. et al. Polymorphism in promoter region of inducible nitric oxide synthase gene and protection against malaria. *Lancet* **351**, 265-6 (1998).
163. Glenn, C.L., Wang, W.Y. & Morris, B.J. Different frequencies of inducible nitric oxide synthase genotypes in older hypertensives. *Hypertension* **33**, 927-32 (1999).
164. Tatemichi, M. et al. Increased risk of intestinal type of gastric adenocarcinoma in Japanese women associated with long forms of CCTTT pentanucleotide repeat in the inducible nitric oxide synthase promoter. *Cancer Lett* **217**, 197-202 (2005).
165. Warpeha, K.M. et al. Genotyping and functional analysis of a polymorphic (CCTTT)(n) repeat of NOS2A in diabetic retinopathy. *Faseb J* **13**, 1825-32 (1999).

166. Kawaguchi, Y. et al. NOS2 polymorphisms associated with the susceptibility to pulmonary arterial hypertension with systemic sclerosis: contribution to the transcriptional activity. *Arthritis Res Ther* **8**, R104 (2006).
167. Taylor, B.S. et al. Multiple NF-kappaB enhancer elements regulate cytokine induction of the human inducible nitric oxide synthase gene. *J Biol Chem* **273**, 15148-56 (1998).
168. de Vera, M.E. et al. Transcriptional regulation of human inducible nitric oxide synthase (NOS2) gene by cytokines: initial analysis of the human NOS2 promoter. *Proc Natl Acad Sci U S A* **93**, 1054-9 (1996).
169. Thorleifsson, G. et al. Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma. *Science* **317**, 1397-400 (2007).
170. Graham, R.R. et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet* **38**, 550-5 (2006).
171. Prokunina, L. et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat Genet* **32**, 666-9 (2002).
172. Leung, T.H., Hoffmann, A. & Baltimore, D. One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. *Cell* **118**, 453-64 (2004).
173. Cambien, F. et al. Sequence diversity in 36 candidate genes for cardiovascular disorders. *Am J Hum Genet* **65**, 183-91 (1999).
174. Yee, L.J. et al. Inducible nitric oxide synthase gene (NOS2A) haplotypes and the outcome of hepatitis C virus infection. *Genes Immun* **5**, 183-7 (2004).
175. Burgner, D. et al. Haplotypic relationship between SNP and microsatellite markers at the NOS2A locus in two populations. *Genes Immun* **4**, 506-14 (2003).
176. Levesque, M.C. et al. A Review of Polymorphisms in the Human Gene for Inducible Nitric Oxide Synthase (NOS2) in Patients with Malaria. *Sepsis* **4**, 213-31 (2000).
177. Cramer, J.P. et al. iNOS promoter variants and severe malaria in Ghanaian children. *Trop Med Int Health* **9**, 1074-80 (2004).
178. Bredt, D.S., Hwang, P.M. & Snyder, S.H. Localization of nitric oxide synthase indicating a neural role for nitric oxide. *Nature* **347**, 768-70 (1990).
179. Moncada, S. & Higgs, E.A. Endogenous nitric oxide: physiology, pathology and clinical relevance. *Eur J Clin Invest* **21**, 361-74 (1991).
180. Christopherson, K.S. & Bredt, D.S. Nitric oxide in excitable tissues: physiological roles and disease. *J Clin Invest* **100**, 2424-9 (1997).
181. Kim, Y.M., Bombeck, C.A. & Billiar, T.R. Nitric oxide as a bifunctional regulator of apoptosis. *Circ Res* **84**, 253-6 (1999).
182. Chung, K.K. et al. S-nitrosylation of parkin regulates ubiquitination and compromises parkin's protective function. *Science* **304**, 1328-31 (2004).
183. Reynaert, N.L. et al. Nitric oxide represses inhibitory kappaB kinase through S-nitrosylation. *Proc Natl Acad Sci U S A* **101**, 8945-50 (2004).
184. MacMicking, J., Xie, Q.W. & Nathan, C. Nitric oxide and macrophage function. *Annu Rev Immunol* **15**, 323-50 (1997).
185. Wallace, M.N., Geddes, J.G., Farquhar, D.A. & Masson, M.R. Nitric oxide synthase in reactive astrocytes adjacent to beta-amyloid plaques. *Exp Neurol* **144**, 266-72 (1997).
186. Xu, W. et al. The CCTTT polymorphism in the NOS2A gene is associated with dementia with Lewy bodies. *Neuroreport* **11**, 297-9 (2000).

187. Levecque, C. et al. Association between Parkinson's disease and polymorphisms in the nNOS and iNOS genes in a community-based case-control study. *Hum Mol Genet* **12**, 79-86 (2003).
188. Barcellos, L.F. et al. Linkage and association with the NOS2A locus on chromosome 17q11 in multiple sclerosis. *Ann Neurol* **55**, 793-800 (2004).
189. Ho, C.Y., Wong, C.K., Li, E.K., Tam, L.S. & Lam, C.W. Elevated plasma concentrations of nitric oxide, soluble thrombomodulin and soluble vascular cell adhesion molecule-1 in patients with systemic lupus erythematosus. *Rheumatology (Oxford)* **42**, 117-22 (2003).
190. Kremsner, P.G. et al. High plasma levels of nitrogen oxides are associated with severe disease and correlate with rapid parasitological and clinical cure in Plasmodium falciparum malaria. *Trans R Soc Trop Med Hyg* **90**, 44-7 (1996).
191. Johannesen, J., Pociot, F., Kristiansen, O.P., Karlsen, A.E. & Nerup, J. No evidence for linkage in the promoter region of the inducible nitric oxide synthase gene (NOS2) in a Danish type 1 diabetes population. *Genes Immun* **1**, 362-6 (2000).
192. Kotikoski, H., Moilanen, E., Vapaatalo, H. & Aine, E. Biochemical markers of the L-arginine-nitric oxide pathway in the aqueous humour in glaucoma patients. *Acta Ophthalmol Scand* **80**, 191-5 (2002).
193. Neufeld, A.H. Nitric oxide: a potential mediator of retinal ganglion cell damage in glaucoma. *Surv Ophthalmol* **43 Suppl 1**, S129-35 (1999).
194. Neufeld, A.H., Hernandez, M.R. & Gonzalez, M. Nitric oxide synthase in the human glaucomatous optic nerve head. *Arch Ophthalmol* **115**, 497-503 (1997).
195. Moncada, S. & Higgs, A. The L-arginine-nitric oxide pathway. *N Engl J Med* **329**, 2002-12 (1993).
196. Bloch, K.D. et al. Three members of the nitric oxide synthase II gene family (NOS2A, NOS2B, and NOS2C) colocalize to human chromosome 17. *Genomics* **27**, 526-30 (1995).
197. Chartrain, N.A. et al. Molecular cloning, structure, and chromosomal localization of the human inducible nitric oxide synthase gene. *J Biol Chem* **269**, 6765-72 (1994).
198. Marsden, P.A. et al. Localization of the human gene for inducible nitric oxide synthase (NOS2) to chromosome 17q11.2-q12. *Genomics* **19**, 183-5 (1994).
199. Eissa, N.T. et al. Alternative splicing of human inducible nitric-oxide synthase mRNA. tissue-specific regulation and induction by cytokines. *J Biol Chem* **271**, 27184-7 (1996).
200. Chu, S.C., Wu, H.P., Banks, T.C., Eissa, N.T. & Moss, J. Structural diversity in the 5'-untranslated region of cytokine-stimulated human inducible nitric oxide synthase mRNA. *J Biol Chem* **270**, 10625-30 (1995).
201. Nunokawa, Y., Oikawa, S. & Tanaka, S. Expression of human inducible nitric oxide synthase is regulated by both promoter and 3'-regions. *Biochem Biophys Res Commun* **233**, 523-6 (1997).
202. Rao, K.M. Molecular mechanisms regulating iNOS expression in various cell types. *J Toxicol Environ Health B Crit Rev* **3**, 27-58 (2000).
203. Liu, B. & Neufeld, A.H. Nitric oxide synthase-2 in human optic nerve head astrocytes induced by elevated pressure in vitro. *Arch Ophthalmol* **119**, 240-5 (2001).
204. Liu, B. & Neufeld, A.H. Activation of epidermal growth factor receptor signals induction of nitric oxide synthase-2 in human optic nerve head astrocytes in glaucomatous optic neuropathy. *Neurobiol Dis* **13**, 109-23 (2003).
205. Neufeld, A.H. & Liu, B. Comparison of the signal transduction pathways for the induction of gene expression of nitric oxide synthase-2 in response to two different stimuli. *Nitric Oxide* **8**, 95-102 (2003).

206. Taylor, B.S. & Geller, D.A. Molecular regulation of the human inducible nitric oxide synthase (iNOS) gene. *Shock* **13**, 413-24 (2000).
207. Feng, X. et al. Identification of a negative response element in the human inducible nitric-oxide synthase (hiNOS) promoter: The role of NF-kappa B-repressing factor (NRF) in basal repression of the hiNOS gene. *Proc Natl Acad Sci U S A* **99**, 14212-7 (2002).
208. Guo, F.Q., Okamoto, M. & Crawford, N.M. Identification of a plant nitric oxide synthase gene involved in hormonal signaling. *Science* **302**, 100-3 (2003).
209. Nathan, C. Inducible nitric oxide synthase: what difference does it make? *J Clin Invest* **100**, 2417-23 (1997).
210. Mungrue, I.N., Bredt, D.S., Stewart, D.J. & Husain, M. From molecules to mammals: what's NOS got to do with it? *Acta Physiol Scand* **179**, 123-35 (2003).
211. Geller, D.A. et al. Molecular cloning and expression of inducible nitric oxide synthase from human hepatocytes. *Proc Natl Acad Sci U S A* **90**, 3491-5 (1993).
212. Laubach, V.E., Zhang, C.X., Russell, S.W., Murphy, W.J. & Sherman, P.A. Analysis of expression and promoter function of the human inducible nitric oxide synthase gene in DLD-1 cells and monkey hepatocytes. *Biochim Biophys Acta* **1351**, 287-95 (1997).
213. Ables, G.P., Hamashima, N. & Watanabe, T. Analysis of genetic factors associated with nitric oxide production in mice. *Biochem Genet* **39**, 379-94 (2001).
214. Odom, D.T. et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nat Genet* **39**, 730-2 (2007).
215. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res* **36**, 5221-31 (2008).
216. Liu, B. & Neufeld, A.H. Expression of nitric oxide synthase-2 (NOS-2) in reactive astrocytes of the human glaucomatous optic nerve head. *Glia* **30**, 178-86 (2000).
217. Neufeld, A.H., Sawada, A. & Becker, B. Inhibition of nitric-oxide synthase 2 by aminoguanidine provides neuroprotection of retinal ganglion cells in a rat model of chronic glaucoma. *Proc Natl Acad Sci U S A* **96**, 9944-8 (1999).
218. Quigley, H.A. Number of people with glaucoma worldwide. *Br J Ophthalmol* **80**, 389-93 (1996).
219. Ray, K., Mukhopadhyay, A. & Acharya, M. Recent advances in molecular genetics of glaucoma. *Mol Cell Biochem* **253**, 223-31 (2003).
220. Evans, T. & Efstratiadis, A. Sequence-dependent S1 nuclease hypersensitivity of a heteronomous DNA duplex. *J Biol Chem* **261**, 14771-80 (1986).
221. Kim, T.H. et al. A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-80 (2005).
222. Tyagi, S., Chabes, A.L., Wysocka, J. & Herr, W. E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol Cell* **27**, 107-19 (2007).
223. Frazer, K.A. et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).
224. Kiemeney, L.A. et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet* **40**, 1307-12 (2008).
225. Saxena, R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).
226. Hillmer, A.M. et al. Susceptibility variants for male-pattern baldness on chromosome 20p11. *Nat Genet* **40**, 1279-81 (2008).
227. Richards, J.B. et al. Male-pattern baldness susceptibility locus at 20p11. *Nat Genet* **40**, 1282-4 (2008).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 403

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-9407

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2008