Medizinische Fakultät

der
Universität Duisburg-Essen

Aus dem Institut für Physiologie

Expressionsregulation von Androglobin

I n a u g u r a l – D i s e r t a t i o n
zur
Erlangung des Doktorgrades der Medizin
durch die Medizinische Fakultät
der Universität Duisburg-Essen

Vorgelegt von
Robert Caspar Markworth

aus
Bergisch Gladbach
2020

DuEPublico
Duisburg-Essen Publications online

UNIVERSITÄT DUISBURG ESSEN
Offen im Denken

ub | universitäts bibliothek

Diese Dissertation wird via DuEPublico, dem Dokumenten- und Publikationsserver der Universität Duisburg-Essen, zur Verfügung gestellt und liegt auch als Print-Version vor.

DOI: 10.17185/duepublico/74568
URN: urn:nbn:de:hbz:464-20210727-125312-2

Alle Rechte vorbehalten.

Dekan:	Herr Univ.-Prof. Dr. med. J. Buer
1. Gutachter:	Herr Prof. Dr. rer. nat. David Hoogewijs
2. Gutachter:	Herr Univ.-Prof. Dr. med. H. Ch. Reinhardt
3. Gutachter:	Herr Prof. Dr. B. Wielockx

Tag der mündlichen Prüfung: 19. April 2021

CONTENTS

1 INTRODUCTION

1.1 The formation of multicellular organisms – an evolutionary approach

Several explanations about the circumstances for the first shift from an anoxic to an oxygenatic atmosphere on earth, also known as the great oxidation event, geologically dated between 2.5 and 2 billion years ago, have been discussed (Kump, 2008, Kasting, 2001). Unarguably, the presence of oxygen changed the outer appearance and forms of life on this planet drastically.

The Proterozoic-Cambrian transition around 540 to 500 million years ago chronicles the first appearance of essentially all high-level metazoan body plans (phyla) in from a geologist point of view brief time interval (Knoll and Carroll, 1999, Sperling et al., 2013). Several possible clarifications have been investigated to determine the correlation between the vast increase in diversity and disparity and the specific timing of the event. Overall acknowledged is that the interplay between escalatory predator-prey arms races and ubiquitously stable ocean/atmospheric oxygenation are responsible for the evolution of greater body size and species diversification (Payne et al., 2011, Sperling et al., 2013).

The mentioned predator-prey arm race discussion - the evolution from the last common eukaryotes ancestor (LECA) - from unicellular to multicellular eukaryotes was also influenced by the evolutionary advantages of meiotic sexual reproduction, opposed to mitotic asexual cell division (Goodenough and Heitman, 2014).
In this context it can be assumed that the earliest eukaryotes species divided mitotically to reproduce their single set of chromosomes, symbolizing a less energy-consuming, thus more efficient way to replicate (Wilkins and Holliday, 2009). As organisms developed to be more complex, mitosis was commonly replaced by sexual reproduction, nevertheless consisting of the three key events: meiosis, sexual differentiation and fertilization. The sexual reproduction has been recognized to happen in just about all eukaryote species and corresponding meiotic cell division is evolutionary highly conserved (Zimmer, 2009, Bai, 2015).

In 1862 Charles Darwin described “the final cause of sexuality” as a “subject (…) yet hidden in darkness” (Darwin, 1862) and even current state of phylogenetic research lacks to shed light into the specific events for the final shift towards the more inefficient meiosis. The necessity for constant adaption against environmental stresses, competitors or parasites, as well as the hunt for prey are nowadays considered as driving forces for the evolutionary implementation of sexual reproduction (Colegrave, 2012, Zimmer, 2009). The constant selective pressure of maintenance of the own species demanded for the transmission of favorable genomic characteristics and increasing genetic variability and yielded finally to the replacement of mitotic division.

Compared to anaerobic metabolism, aerobic metabolism represents a more efficient source of energy, producing approximately 16 to 18 times more adenosine triphosphate (ATP), the universal source of energy for cells (Jiang et al., 2010). The increasing presence of atmospheric oxygen can thus be considered as a starting point for the evolution of complex life forms (Burmester and Hankeln, 2014).
Alongside to specialized respiratory organs and a corresponding circulatory system, proteins evolved to ensure proper transport and storage of oxygen and adapt to the new environment (Burmester and Hankeln, 2014).

1.2 The globin family

Among the respiratory proteins in protostome lineages, globins are the most widespread (Burmester and Hankeln, 2014). The expression of globins has been detected in vertebrates, including humans and rodents but also in invertebrates, prokaryotes, fungi, protozoa, as well as in plants (Vinogradov et al., 2005, Vinogradov et al., 2006, Vinogradov et al., 2013b, Vinogradov et al., 2013a, Hoogewijs et al., 2012a, Vinogradov et al., 2011). This broad field of expression was proven by extensive research, representing globins as one of the best-investigated proteins in medical and biological sciences.

Even though the primary sequence varies among the protein family, they commonly comprise about 150 amino acids and the tertiary structure is evolutionary conserved. The

structure typically consists of eight α-helical segments, named A to H, which form the characteristic "globin fold", a 3-over-3 α-helical sandwich structure and incorporate a central heme prosthetic group (Fe^{2+}-protoporphyrin IX) (Figure 1A) (Holm and Sander, 1993). The heme prosthetic group is designated to reversibly bind gaseous ligands such as oxygen (O_2), carbon monoxide (CO) and nitric oxide (NO) (Figure 1B) (Perutz, 1979). Among the individual globin proteins, the molecular structure of the N- and C-termini varies.

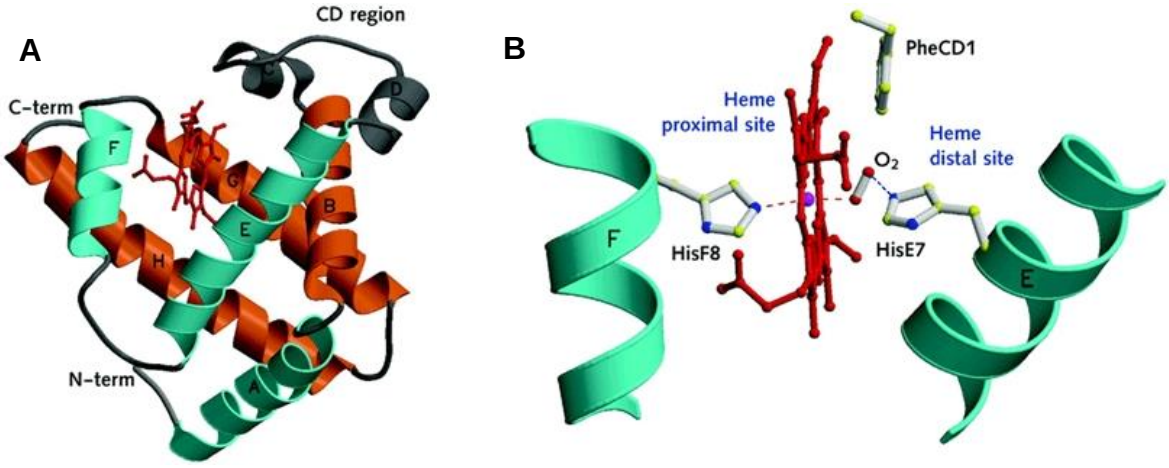

Fig. 1. The "classical globin" fold (Pesce et al., 2002). **(A):** The 3-over-3 α-helical globin sandwich structure is highlighted in orange and green. The C- and the N-terminus of the protein and the helices A - H are labeled. Furthermore, the CD region (linking helices C & D) is mentioned. **(B):** A closer look on the heme binding of an oxygen molecule. The central heme group is presented in red; the helices E & F are presented in green, with the corresponding amino acids that are involved in the binding.

Thus far, eight different types of globins have been identified in vertebrates (Burmester and Hankeln, 2014). Next to the commonly known and intensively studied hemoglobin (Hb) and myoglobin (Mb), an unprecedented diversity of globins has been revealed more recently.

Since the beginning of this century, neuroglobin (Ngb), cytoglobin (Cygb), eye-globin or globin E (GbE), globin X (GbX), globin Y (GbY) and androglobin (Adgb) have been added to the protein family, representing globins as a highly innovative part of current scientific research. A detailed look on specific globin expression unravels new astonishing globin features, such as sensing, signaling, and the detoxification of reactive oxygen species, however there is still much more to be explored (Vinogradov and Moens, 2008).

Taking into account the specific heme binding to the central iron in a deoxygenated state, the protein family can be divided into penta- or hexa-coordination (Kakar et al., 2010).

Hb, Mb and GbE are penta-coordinated, displaying an empty sixth Fe^{2+} binding, Ngb, Cygb, GbX and Adgb are hexa-coordinated, demonstrating a distal histidine or glutamine binding to the Fe^{2+}. The exact coordination of GbY still remains to be elucidated (Burmester and Hankeln, 2014). The functional background of the differing coordination remains unclear, however an improved gaseous sensing of hexa-coordinated proteins is anticipated (de Sanctis et al., 2004).

Hb represents the most established member of the globin family, mainly being responsible for the mentioned circulatory transportation of O_2. Mb is commonly known for its ability to extract oxygen from the blood, store it and supply the mitochondria of myocytes with it. In addition, Mb expression has been reported to contain a potential tumor suppressing effect in breast cancer (Kristiansen et al., 2011). Both globins have also been associated with the metabolism of NO, a signaling molecule involved in many physiological and pathological processes.

As suggested by the prefix, Ngb is mainly expressed in the central and peripheral nervous system and a role in the oxidative metabolism is anticipated. Despite extensive research no direct function has been linked to an expression so far (Burmester et al., 2000, Burmester and Hankeln, 2009). Cygb is primarily expressed in fibroblasts and an involvement in the oxidative metabolism is predicted as well (Burmester et al., 2002). The antioxidative effect has been shown in several cellular models, including renal podocyte-derived cell lines (Randi et al., 2020). The expression of GbE, GbX and GbY is restricted to certain vertebrate species only. Whereas GbE is thought to be involved in the supply of oxygen in the retina of birds, GbX is speculated to participate in oxygen sensing in the central nervous system of lampreys, sharks, ray-finned fishes, coelacanth, amphibians and reptiles (Roesner et al., 2005, Kugelstadt et al., 2004). Less is known about the role of GbY, which is mainly expressed in clawed frogs (Fuchs et al., 2006).

Androglobin was discovered in 2012 and represents the newest member of the globin family (Hoogewijs et al., 2012b).

1.3 Androglobin – a male fertility related globin

As suggested by its name, androglobin combines the two outlined evolutionary driving forces for complex organisms. On one hand the molecular structure suggests its globin identity. On the other hand, male fertility is inevitably dependent on androglobin expression.

Phylogenetic distribution analysis of androglobin reveals orthologues in almost all metazoan lineages and in choanoflagelates, which display the closest unicellular organism to metazoa, before the aforesaid increase in diversity of species (Hoogewijs et al., 2012b). It can be considered that the rate of evolutionary conservation of genomic regions is correlated to the increased likelihood of indispensable functions (Hirsh and Fraser, 2001). The findings about androglobin suggest an evolutionary very conserved function and therefore hint for an essential function.

Adgb mRNA expression was quantified in mouse tissue and revealed a most abundant expression in testis tissue, with a 10-fold lower expression in lung and a ~100-fold lower expression in the remaining tissues (Figure 2) (Hoogewijs et al., 2012b).

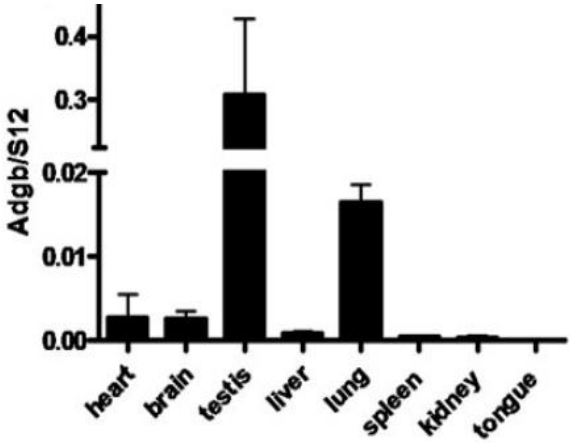

Fig. 2. Analysis of relative androglobin expression levels in different mouse tissues (Hoogewijs et al., 2012b).

Pronounced androglobin expression in mouse testis tissue is first detected after postnatal day 25, corresponding to late stages of spermatogenesis and elongation of postmeiotic spermatids (Figures 3A , B) (Hoogewijs et al., 2012b).

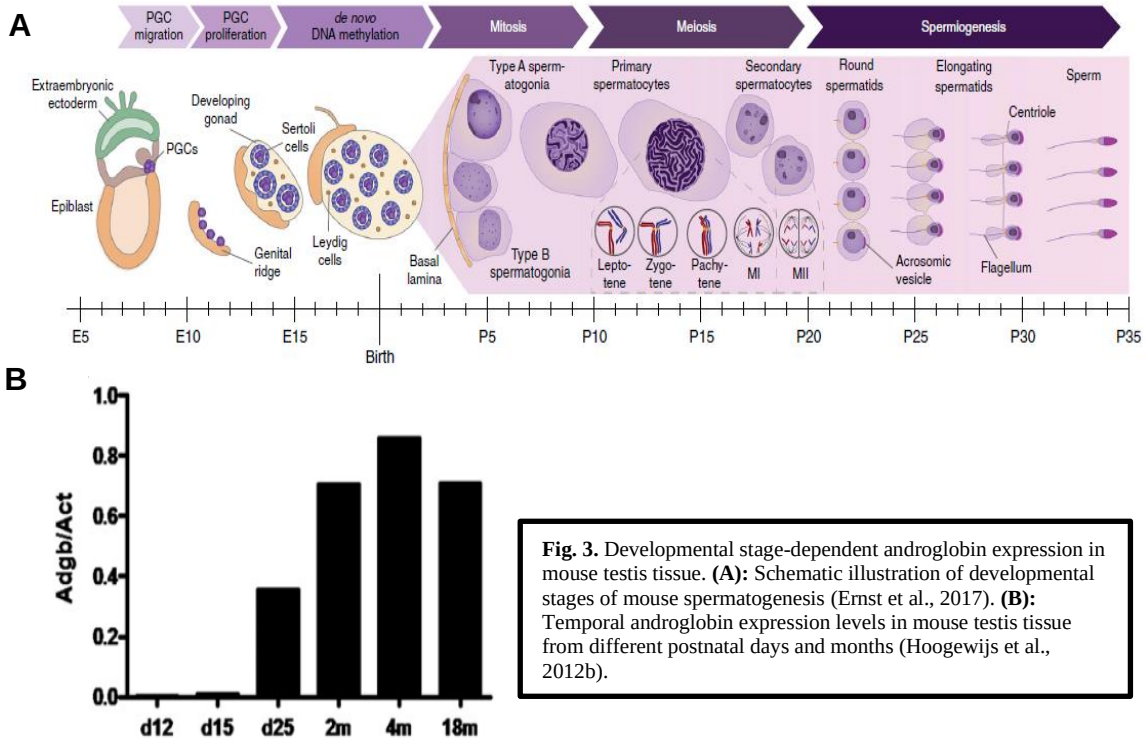

Fig. 3. Developmental stage-dependent androglobin expression in mouse testis tissue. **(A):** Schematic illustration of developmental stages of mouse spermatogenesis (Ernst et al., 2017). **(B):** Temporal androglobin expression levels in mouse testis tissue from different postnatal days and months (Hoogewijs et al., 2012b).

In situ hybridization of *Adgb* antisense RNA to mouse testis cryosections reveals an increasing expression towards the lumen of the seminiferous tubules (Hoogewijs et al., 2012b).

The tremendous impact on fertility is demonstrated by androglobin-deficient mice (*Adgb* $^{-/-}$), displaying infertility and corresponding accumulation of immature spermatids in the lumen of the seminiferous tubules and the epididymis (Figures 4A – B) (Keppner et al., unpublished). *Adgb*-lacking mice suffer additionally from pathologies, such as hydrocephalus or kidney cysts (Keppner et al, unpublished).

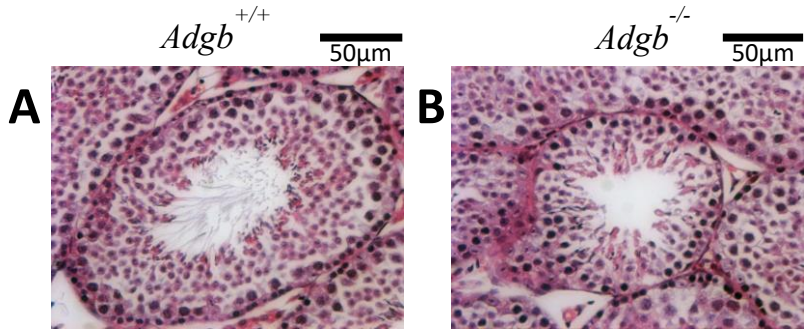

Fig. 4. Hematoxylin and eosin staining of mouse seminiferous tubules in testis tissue. **(A):** Wild type sample, *Adgb* $^{+/+}$ **(B):** Genetic knockout, *Adgb* $^{-/-}$ (60 times zoom) (Keppner et al., unpublished).

In line with its role in murine male fertility, spermatozoa of infertile men display lower androglobin expression levels (Hoogewijs et al., 2012b).

1.4 Androglobin – a chimeric protein

An initial overview of the protein structure reveals that androglobin is a chimeric protein and can be divided into four major domains (Figure 5).

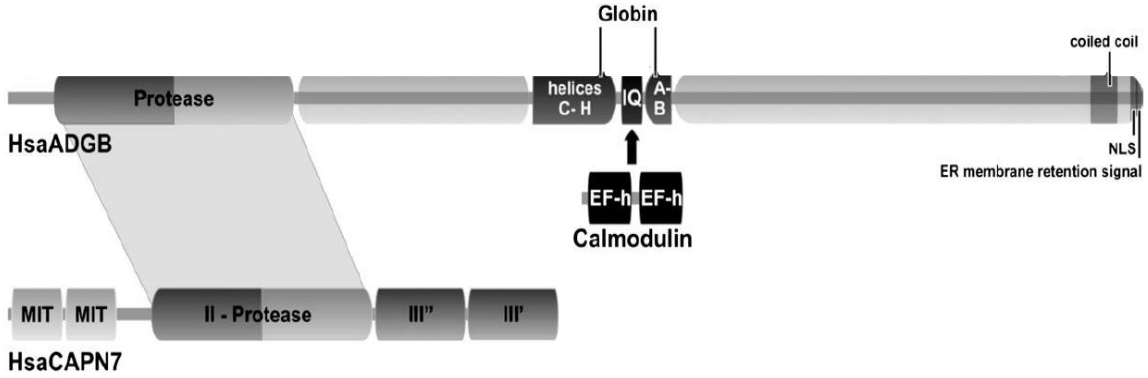

Fig. 5. An overview of the chimeric protein structure of androglobin. Starting at the N-terminus: a calpain-like protease, as comparison the structure of human calpain-7 (HsaCAPN7); the rearranged globin domain (helices A – H); the central IQ motif and the potentially binding calmodulin; a C-terminal coiled-coil region, candidate nuclear localisation (NLS) and ER membrane retention signals (Hoogewijs et al., 2012b).

The N-terminal part of the protein structure is homologous to the catalytic domain II of calpain 7 (Figure 5) (Hoogewijs et al., 2012b).

Overall, 15 different isoforms of calpains are found in humans. They represent a group of calcium-dependent cysteine proteases and have been implicated to modulate the corresponding function of their substrates in a broad variety of basic, cellular processes (Perrin and Huttenlocher, 2002). The calpain domain is followed by a protein region of yet to be characterized function (Figure 5). This domain is on its turn followed by the characteristic, central globin domain (Figure 5). As opposed to the other members of the globin family, androglobin represents the first globin with a divided and circularly permutated globin domain. More specifically, it is split in two parts, separated by a central, dividing IQ motif (Figure 5) (Hoogewijs et al., 2012b). The helices C to H are shifted towards the N-terminus and the helices A and B towards the C-terminus of the protein (Figure 5). The observed absorption spectra of various recombinantly expressed partial and full-length Adgb proteins in *Escherichia coli* (*E. coli*) and insect cells confirmed heme hexacoordination (Hoogewijs et al., 2012b). The IQ motif is commonly distributed in a diverse variety of tissues and recognized as a calmodulin binding amino acid sequence, in which context IQ refers to the first two amino acids of the motif, isoleucine and glutamine.

Calmodulin is a calcium-dependent protein that modulates cell signaling via the alteration of calcium availability in a broad variety of cell functions. Among the regulation of molecular motor activity, a calmodulin binding IQ motif is also found in the sperm surface protein 17 (Sp17) to facilitate sperm penetration and sperm – egg fusion (Bahler and Rhoads, 2002).

The remaining domain is localized at the C-terminus of androglobin and hence requires further characterization (Figure 5). A coiled-coil region and potential nuclear localization signals (NLS) as well as endoplasmic reticulum (ER) membrane retention signals have been identified so far, representing cell function regulating gene regions (Figure 5) (Hoogewijs et al., 2012b).

1.5 A molecular analysis of androglobin

1.5.1 The historical process to modern gene sequencing

This research topic analyzes the expression regulation of androglobin and focuses on the impact of a potential promoter and enhancer gene region. For this purpose, computational analysis of epigenomic data derived from public databases was performed of the human androglobin (*ADGB)* gene locus, enabling the identification of regulatory gene regions and potential ADGB expression regulating transcription factors.

It has been a long journey since the first discovery and isolation of DNA in 1869 by Friedrich Miescher to nowadays high-throughput DNA sequencing and public accessible data of sequenced and analyzed genomes (Dahm, 2005). Considering the corresponding milestones, it becomes apparent, that especially the last two decades marked a period of great acceleration regarding the insight and understanding about specific interactions in the genome.

The model of a double-helix structure of DNA, introduced by James Watson and Francis Crick in 1953, established the four nucleotides as backbone of every form of life and the complementary relation of the nucleotides as fundamental DNA structure. The next breakthrough was reached in 1955 by the first publication of the exact amino acids sequence of a protein by Frederick Sanger, which in this case was insulin (Sanger et al.,

1955). The chain termination and the chemical degradation method, established in the mid-1970s, were the first cost-efficient and reliable methods to enable sequencing of DNA segments, that were also larger (Maxam and Gilbert, 1977, Sanger et al., 1977). The following, ambitious thrive to sequence the entire human genome, led to the foundation of the Human Genome Project (HGP) in the late 1980´s, an organized collaboration between geneticists of all parts of the world (Hood and Rowen, 2013). As a result of continuous improvement of sequencing methods and the extension of collaborations among scientists, the HGP published the first working draft in 2001, covering the exact sequence of about 90% of the human genome (Lander et al., 2001). By now, the technology is as advanced, that sequencing has become a standard tool in scientific research and medical genomics.

The encoded genome represents the common background for every single cell. However, the astonishing diversity in specification and functionality of single cells in a multicellular organism is assured by specific gene regulation. By enabling or disabling the transcription of specific proteins, gene regulation coordinates the network in between the individual cells, to exert the specific function for which the cell is required.
The facilitated access to genome sequences as a foundation stone, has pathed the way to recognize sequence patterns and DNA interactions, to establish a more systematical insight into the genetic biology (Ideker et al., 2001).
To display the regulation and organization of genes and the whole genome, the Encyclopedia of DNA Elements (ENCODE) effort was launched in 2003 (Collins et al., 2003). Large numbers of sequence-based studies not only revealed that more than 80% of the human genome has actual biochemical function, but it also showed that the actual genome fraction for gene regulation is significantly higher than that for the actual biochemical function (Consortium, 2012).

The University of California Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu/) is a public database which displays detailed annotations on genes and gene predictions, plus the corresponding regulation. The data of the browser is based on diverse sequencing methods and, among other databases, includes the data from

the ENCODE project (Kent et al., 2002). The UCSC Genome Browser was also used for this thesis project, in order to analyze the gene locus of androglobin.

Furthermore, data from the JASPAR database (http://jaspar.genereg.net/) was used to predict transcription factor binding sites in specific genomic sequences.
The JASPAR database is an open-access database which displays matrix-based transcription factor binding profiles in eukaryotes (Sandelin et al., 2004).

1.5.2 The genomic structure of androglobin

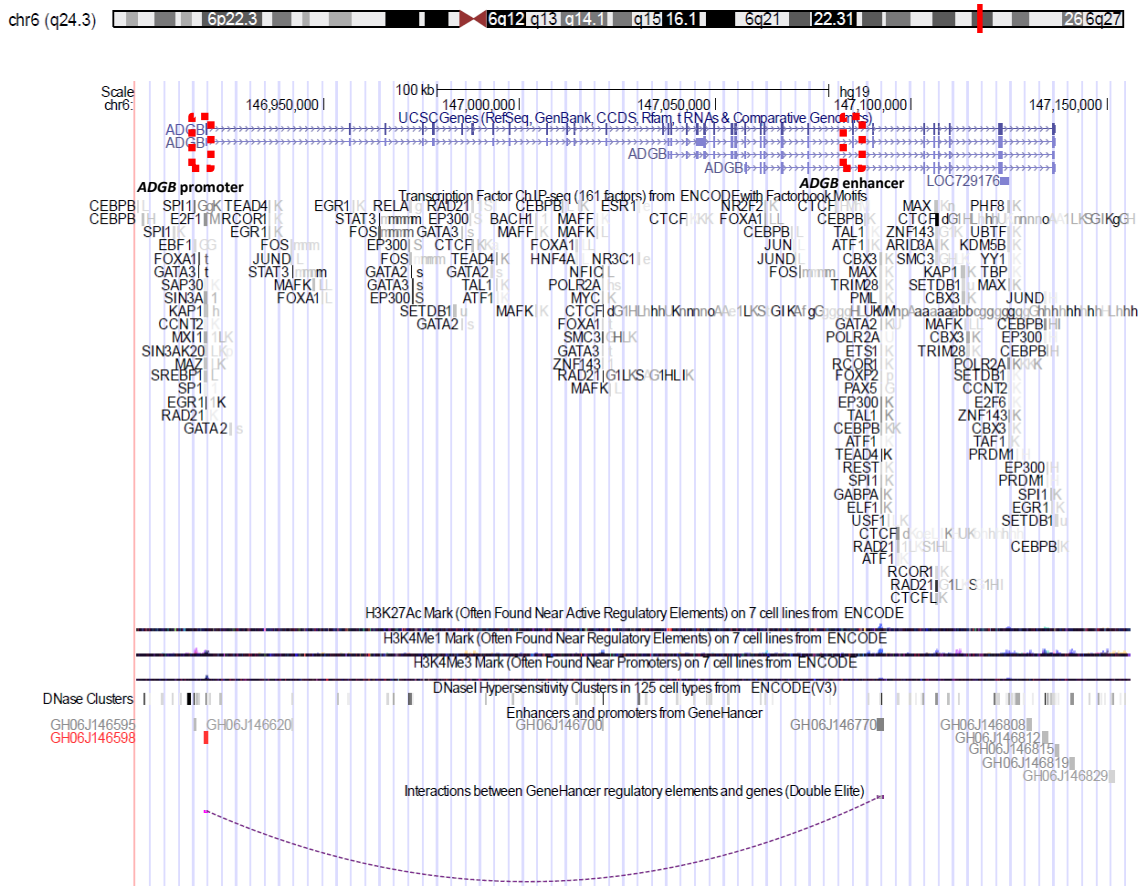

Fig. 6. The *ADGB* gene locus, as displayed by the UCSC Genome Browser and customized to show the information of interest (http://genome.ucsc.edu/).

The figure above shows the *ADGB* gene locus, displayed by the UCSC Genome Browser and customized to provide information about the transcriptional regulation.
The human *ADGB* gene locus lays on chromosome 6 and spans about 216 462bp. The longest transcription variant comprises 36 exons and 35 introns.

The first line of the figure is an overview of the 6th chromosome. The red vertical bar indicates the position of the *ADGB* locus on the long arm of the chromosome (Figure 6). Below the scale for base pairs, four transcription variants of androglobin are displayed in blue, the exons being marked as vertical lines and the introns as horizontal lines. The approximate genomic location of the investigated *ADGB* promoter and enhancer regions are circled in red dots (Figure 6).

The following track lists transcription factors that have been identified by chromatin immunoprecipitation and subsequent sequencing (ChIP-seq) to have a binding site at the genome region that they are annotated to (Figure 6). The experiments were performed in 91 cell types and scanned for a selection of 161 different factors. The grey vertical bar represents the exact region of the binding site. Next to the marker of the region, the name of the factor and the initials of each cell line, in which the binding was identified, are annotated (Figure 6).

The histone marks H3K27Ac, H3K4Me1, H3K4Me4, shown on the three black vertical bars below, are displayed as representatives for repressive or activating regulatory regions (Figure 6). Histone modifications alter the chromatin structure and consequently the corresponding DNA interaction. The nomenclature of the histone marks indicates the belonging to DNA packaging protein Histone H3. The following K is an abbreviation for the affected amino acid lysine, followed by the exact position of the amino acid residue. Ac correlates in this context to an acetylation, Me to a methylation. The following number displays the number of methyl groups that were added.

Whereas H3K27Ac is defined as an activating enhancer mark, provoking higher activation of transcription, H3K4Me1 and H3K4Me3 are considered as inactive histone marks that need to be further activated (Roadmap Epigenomics et al., 2015). H3K4Me1 is found in relation to enhancer activity and H3K4Me3 in proximity of promoter gene regions (Roadmap Epigenomics et al., 2015).

The following track shows DNaseI hypersensitivity areas that were plotted in 125 cell types within the framework of the ENCODE project. As suggested by its name, DNaseI hypersensitivity regions are chromatin regions that are sensitive to cleavage by the DNaseI enzyme, hence represent a decondensed structure accessible for transcriptional regulation (Thurman et al., 2012).

The last section of the figure indicates potential promoter gene regions as red and enhancer as grey vertical bars, according to GeneHancer, a database focused on mapping genomic enhancer regions (Fishilevich et al., 2017).
The predicted interaction between those regulatory gene regions is highlighted by the purple curve in the section beneath.

The genomic region of androglobin represents multiple potential layers of regulation. The scientific interest into the proposed *ADGB* promoter and enhancer regions and the corresponding interaction is underlined, taking the outlined measures of analysis and further predictions into consideration. The observations about the genomic regions match with many criteria for gene regulatory regions and propose to play an important role in the expression regulation of androglobin.
The analysis of the exact transcriptional regulation of androglobin can reveal further information about when, where and why the protein is expressed. Furthermore, it allows us to draw hypotheses and conclusions about yet to be discovered functions of androglobin.

2 MATERIAL AND METHODS

2.1 Expendable materials

0.2ml PCR-reaction tube:	Biozym, Oldendorf
6-well plates:	Sarstedt, Nürnbrecht
24-well plates:	Sarstedt, Nürnbrecht
96-well plates:	Sarstedt, Nürnbrecht
Adhesive film for 96-well plates:	4titude, Surrey, UK
Cell scraper:	Sarstedt, Nürnbrecht
Centrifuge tubes (15ml;50ml):	Sarstedt, Nürnbrecht
Filter tips:	Sarstedt, Nürnbrecht
Petri dish (92x16mm):	Sarstedt, Nürnbrecht
Petri dish (150x20mm):	Sarstedt, Nürnbrecht
Pipette tips:	Sarstedt, Nürnbrecht
Reaction tube (0.5ml, safe-lock lit):	Sarstedt, Nürnbrecht
Reaction tube (1.5ml, normal lit):	Sarstedt, Nürnbrecht
Reaction tube (1.5ml, safe-lock lit):	Sarstedt, Nürnbrecht
Reaction tube (2ml, safe-lock lit):	Sarstedt, Nürnbrecht
Serological pipettes:	Greiner bio-one, Essen

2.2 Equipment

-80°C freezer:	Panasonic MDF-U53865-PE, Osaka
-20°C freezer:	Liebherr Comfort, Kirchdorf
4°C fridge:	Liebherr Comfort, Kirchdorf
Autoclave:	Varioclav, Thermo Scientific, Waltham, MA, USA
Cell counter:	TC 20, Thermo Fisher, Oberhausen
Centrifuges:	Varifuge 3.0R, Heraeus, Hanau
	Centrifuge 5415R, Eppendorf, Wesseling Berzdorf

Imaging system for agarose gels:	BioDoc-IT Imaging System, UVP, Upland CA, USA
Heating block:	HTM 130, HLC, Haep-Labor, Bovenden
Incubation shaker:	SM 30 A, Edmund Bühler, Hechingen
Micro plate luminometer:	Centro LB 960 Microplate Luminoter, Berthold Technologies, Bad Wildbad
Micro plate spectrophotometer:	Epoch, Biotek, Vermont, USA
Mini Lab Roller:	Labnet, Edison, NJ, USA
pH-meter:	Eutech pH700, Thermo Scientific, Waltham, MA, USA
Pipettes:	Pipetman classic, Gilson, Middleton, WIS, USA
	Pipetboy comfort, IBS integra Biosciences, Biebertal
Power supply unit:	Powerpac 200, BioRad, Hercules, CA, USA
Sterile bench:	Hera Safe Typ HS12, Heraeus, Hanau
Sonicator:	Sonoplus, Bandelin, Berlin
Thermo cyclers:	Tpersonal, Biometra, Göttingen
	Tprofessional, Biometra, Göttingen
	CFX96 Real-Time System, Biorad, Hercules, CA, USA
Table centrifuges:	Biofuge pico, Heraeus, Hanau
	Model Sprout™, Biozym Scientific Hessisch Oldendorf
Vortex-mixer:	Select Vortexer, Select Bioproducts, Edison, NJ, USA

2.3 Chemicals

All chemicals are of the highest degree of purity.

The solutions were prepared with distilled and sterile-filtered water (filtered by a millipore-facility).

100bp DNA ladder:	AppliChem, Darmstadt
1Kb DNA ladder:	AppliChem, Darmstadt
Acetic acid (CH_3COOH):	Thermo Scientific, Waltham, MA, USA
Agarose basic:	AppliChem, Darmstadt
Ampicillin (100mg/ml):	Sigma Aldrich, St. Louis, MO, USA
Bovine serum albumin (BSA):	Serva, Hamburg
Calcium chloride (CaCl):	Sigma Aldrich, St. Louis, MO, USA
Deoxycholate:	Sigma Aldrich, St. Louis, MO, USA
Difco antibiotic:	Norwald, Homburg
Disodium phosphate (Na_2HPO_4):	Sigma Aldrich, St. Louis, MO, USA
DMEM:	Gibco, life technologies, Carlsbad, CA, USA
Ethanol:	Roth, Karlsruhe
Ethidiumbromid (1% solution):	AppliChem, Darmstadt
Ethylenediaminetetraacetate (EDTA):	AppliChem, Darmstadt
Fetal calf serum (FCS):	Biochrom, Berlin
Glycerol:	Thermo Scientific, Waltham, MA, USA
Isopropanol:	Thermo Scientific, Waltham, MA, USA
Lithium chloride (LiCl):	Sigma Aldrich, St. Louis, MO, USA
MOPS (3-(N-morpholino) propane sulfonic acid):	Sigma Aldrich, St. Louis, MO, USA
Nonoxynol 40 (Np40):	AppliChem, Darmstadt
Passive lysis buffer (PLB):	Promega, Madison, WI, USA
Penicillin/Streptomycin:	Gibco, life technologies, Carlsbad, CA, USA
Pierce Protein A/G Agarose:	Thermo Scientific, Waltham, MA, USA
Potassium acetate (KAc):	AppliChem, Darmstadt
Potassium chloride (KCl):	AppliChem, Darmstadt
Potassium dihydrogen phosphate (KH_2PO_4):	Merck, Darmstadt
Potassium hydrogen carbonate ($KHCO_3$):	Merck, Darmstadt
Rubidium chloride (RbCl):	AppliChem, Darmstadt

Salmon sperm DNA :	AppliChem, Darmstadt
Scacchite (MnCl):	Merck, Darmstadt
Select Agar powder:	Thermo Fisher, Oberhausen
Sodium acetate (NaAc):	Sigma Aldrich, St. Louis, MO, USA
Sodium bicarbonate ($NaHCO_3$):	Sigma Aldrich, St. Louis, MO, USA
Sodium chloride (NaCl):	Sigma Aldrich, St. Louis, MO, USA
Sodium hydrogen phosphate (Na_2HPO_4):	Merck, Darmstadt
Tris:	AppliChem, Darmstadt
Tris-HCl:	AppliChem, Darmstadt
Triton X-100:	Thermo Scientific, Waltham, MA, USA
Trypsin:	Gibco life technologies, Carlsbad, CA, USA
Trypton:	Sigma Aldrich, St. Louis, MO, USA
Yeast extract, BioChemica:	AppliChem, Darmstadt

2.4 Buffers

Tris-Acetate-EDTA buffer (TAE buffer):
45mM Tris
45mM Acetic acid
1mM EDTA, pH: 8

dNTP-Mix:
10mM dATP
10mM dCTP
10mM dGTP
10mM dTTP
in ddH_2O

Tris-EDTA buffer:
1M Tris-HCl
0.5M EDTA, pH: 8

Transformation Buffer 1 (TFB1) (preparation of competent cells):
30mM KAc
100mM RbCl
10mM $CaCl_2$
50mM $MnCl_2$
15% Glycerol
add ddH_2O up to one liter and adjust the pH to 5.8 with 0.2M acetic acid

Transformation Buffer 2 (TFB2) (preparation of competent cells):
10mM MOPS
75mM $CaCl_2$
10mM $RbCl_2$
15% Glycerol
add ddH_2O up to one liter and adjust the pH to 6.5 with KOH

Resuspension Buffer (Miniprep):

50mM Glucose

25mM Tris-HCl, pH: 8

10mM EDTA, pH: 8

Lysis Buffer (Miniprep):

0.2M NaOH

1% SDS

Neutralization Buffer (Miniprep):

3M NaAc, pH: 5.2

Phosphate-buffered saline (PBS):

150mM NaCl

8mM Na_2HPO_4

2mM KH_2PO_4

Preblocking Buffer (ChIP):

200µg BSA

200µg salmon sperm DNA

Lysis Buffer (ChIP):

1% SDS

10mM EDTA

50mM Tris-HCl (pH: 8.1)

Add fresh protease inhibitor (1:10)

IP Dilution Buffer (ChIP):

1% Triton X-100

2mM EDTA

150mM NaCl

20mM Tris-HCl (pH: 8.1)

150mM NaCl

TSEI (ChIP):

0.1% SDS

1% Triton X-100

2mM EDTA

20mM Tris-HCl (pH: 8.1)

TSEII (ChIP):

0.1% SDS

1% Triton X-100

2mM EDTA

20mM Tris-HCl (pH: 8.1)

500mM NaCl

Buffer III (ChIP):

0.25M LiCl

1% NP-40

1% Deoxycholate

1mM EDTA

10mM Tris-HCl (pH: 8.1)

Extraction Solution (ChIP):

1% SDS

0.1M $NaHCO_3$

2.5 Enzymes

BglII restriction enzyme + buffer:	New England BioLabs, Frankfurt
Go-Taq G2 DNA Polymerase + buffer:	Promega, Madison, WI, USA
HindIII restriction enzyme + buffer:	New England BioLabs, Frankfurt
HindIII-HF restriction enzyme + buffer:	New England BioLabs, Frankfurt
KpnI restriction enzyme + buffer:	New England BioLabs, Frankfurt
KpnI-HF restriction enzyme + buffer:	New England BioLabs, Frankfurt
M-MLV Reverse Transcriptase + buffer:	Promega, Madison, WI, USA
NcoI restriction enzyme + buffer:	New England BioLabs, Frankfurt
NheI restriction enzyme + buffer:	New England BioLabs, Frankfurt
Phusion High-Fidelity DNA Polymerase + buffer:	New England BioLabs, Frankfurt
Protease-inhibitor:	Roche, Basel, Switzerland
RNase A:	Thermo Scientific, Waltham, MA, USA
Rotifect:	Carl-Roth, Karlsruhe
Phusion High-Fidelity DNA Polymerase + buffer:	New England BioLabs, Frankfurt
SpeI-HF restriction enzyme + buffer:	New England BioLabs, Frankfurt
T4 DNA Ligase + buffer:	New England BioLabs, Frankfurt
XbaI restriction enzyme + buffer:	New England BioLabs, Frankfurt
XhoI restriction enzyme + buffer:	New England BioLabs, Frankfurt

2.6 Kits

Dual-Glo Luciferase Assay System:	Promega, Madison, WI, USA

Genomic DNA Purification Kit:	Thermo Scientific, Waltham, MA, USA
GeneJET Plasmid Miniprep Kit:	Thermo Scientific, Waltham, MA, USA
High Pure PCR Product Purification Kit:	Roche, Basel, Switzerland
QIAEX II Gel Extraction Kit:	Qiagen, Venlo, Netherlands
QuikChange Lightning Kit, Site-Directed Mutagenesis Kit:	Agilent Technologies, Santa Clara, CA, USA
RNeasy Mini Kit:	Qiagen, Venlo, Netherlands
Takyon qPCR Kit:	Eurogentec, Seraing, Belgium

2.7 Antibodies

Primary antibodies:

GATA Binding Protein 3 (GATA3), origin: rabbit IgG	Cell Signalling D13C9
Paired Box 5 (Pax5), origin: rat	Merk Millipore 05-1573

Secondary antibody:

β-actin:	origin: rabbit IgG	Sigma Aldrich, St. Louis, MO, USA

2.8 Primer sequences

2.8.1 Primer sequences for cloning

The following lists include the primer pairs that were used to amplify the *ADGB* promoter fragments and the *ADGB* enhancer. The utilized restriction enzymes are indicated by the primer name. The restriction sites for the cleavage of the enzymes are marked in red.

Amplicon:	**Primer name:**	**Primer sequence (5´ → 3´):**	**Product size:**
431bp *ADGB* promoter	*ADGB*-prom431-BglII-F	GGAAGATCTtccgaccccaggttttcaa Gcca	460bp
	ADGB-prom431-HindIII-R	CCCAAGCTTGGGtccgcgtctgcgtt gccagga	

1031bp *ADGB* promoter	*ADGB*-prom1031-BglII-F	GGAAGATCTTCCatctctgaagtg aggttata	1060bp
	ADGB-prom1031-HindIII-R	CCCAAGCTTGGGtccgcgtctgcgtt gccagga	
1981bp *ADGB* promoter	h*ADGB*1981promoter-XhoI-F	CCGCTCGAGCGGctcgtcacatcttttatgat	2007bp
	h*ADGB*1981promoter-NcoI-R	CATGCCATGGCATGtccgcgtctgcgt tgccagga	

Table 1. Primer sequences that were used to amplify the *ADGB* promoter fragments, to clone in the pGL3-Basic Vector.

Amplicon:	**Primer name:**	**Primer sequence (5´ → 3´):**	**Product size:**
ADGB enhancer	*ADGB*-enh-KpnI-F	GGGGTACCCCaaacacttgatggt	1707bp
	ADGB-enh-NheI-R	CTAGCTAGCTAGacggtatgagctaag	
ADGB enhancer	*ADGB*-enh-KpnI-F	GGGGTACCCCaaacacttgatggt	1707bp
	ADGB-enh-XbaI-R	CTAGCTAGCTAGacggtatgagctaag	

Table 2. Primer sequences that were used to amplify the *ADGB* enhancer, to clone in the pGL3B-xAP vector.

Amplicon:	**Primer name:**	**Primer sequence (5´ → 3´):**	**Product size:**
431bp *ADGB* promoter	*ADGB*-prom431-BglII-F	GGAAATCTTCCgaccccaggttttcaagcca	460bp
	ADGB-prom431-HindIII-R	CCCAAGCTTGGGtccgcgtctgcgttgccagga	
1031bp *ADGB* promoter	*ADGB*-prom1031-SpeI-F	TATATATAACTAGTTAatctctgaagt gaggttata	1060bp
	ADGB-prom1031-HindIII-R	CCCAAGCTTGGGtccgcgtctgcgttgccagga	
1981bp *ADGB* promoter	h*ADGB*1981promoter-XhoI-F	CCGCTCGAGCGGctcgtcacatcttttatgat	2007bp
	h*ADGB*1981promoter-NcoI-R	CATGCCATGGCATGtccgcgtctgcgt tgccagga	

Table 3. Primer sequences that were used to amplify the *ADGB* promoter fragments, to clone in the pGL3-Enhancer Vector.

Amplicon:	**Primer name:**	**Primer sequence (5´ → 3´):**	**Product size:**
ADGB enhancer	*ADGB*-enh-KpnI-F	CGGGGTACCCCGcaaacacttgatggt	1709bp
	ADGB-enh-XbaI-R	TGCTCTAGAGCAacggtatgagctaag	

Table 4. Primer sequences that were used to amplify the *ADGB* enhancer, to clone in the pGL3-Basic Vector.

Amplicon:	**Primer name:**	**Primer sequence (5´ → 3´):**	**Product size:**
ADGB enhancer	*ADGB*-enh-KpnI-F	CGGGGTACCCCGcaaacacttgatggt	1709bp
	ADGB-enh-XbaI-R	TGCTCTAGAGCAacggtatgagctaag	

Table 5. Primer sequences that were used to amplify the *ADGB* enhancer to clone in the pGL3-Promoter Vector.

2.8.2 Primer sequences used for real-time PCR

The following list includes the primer pairs that were used to analyze the endogenous mRNA expression of the transcription factors, in the examined cell lines HeLa, HEK293T and MCF7.

Amplicon:	Primer name:	Primer sequence (5´ → 3´):	Product size:
Pax5	Pax5-F	ggaggagtgaatcagcttgg	393bp
	Pax5-R	gttgggtggctgctgtactt	
Tripartite Motif Containing 28 (Trim28)	TRIM28-F	ctcgggatggtgaacgtact	213bp
	TRIM28-R	gcaatgttgcatgtttgtcc	
Upstream Transcription Factor 1 (USF1)	USF1-F	cctggcactggtcaattctt	200bp
	USF1-R	tgcacgatccagttgttgat	
GA Binding Protein Transcription Factor Subunit Alpha (GABPα)	GABPα-F	aagtgacaagatgggctgct	171bp
	GABPα-R	ccgaaatgttgagtgtggtg	
CCAAT Enhancer Binding Protein Beta (CEBPβ)	CEBPβ-F	gacaagcacagcgacgagta	158bp
	CEBPβ-R	agctgctccaccttcttctg	
RE1 Silencing Transcription Factor (Rest)	REST-F	gaggaggagggctgtttacc	160bp
	REST-R	tcacagcagctgccatttac	
Nuclear Factor, Erythroid 2 Like 3 (Nrf3)	Nrf3-F	tgtcgtaaacgcaaattgga	186bp
	Nrf3-R	tggcctaccttggtcatctc	
Chromobox 3 (HP1Gamma)	HP1Gamma -F	gctggcaaagaaaaagatgg	212bp
	HP1Gamma -R	agcaccaagtctgcctcatc	
Spi-1 Proto-Oncogene (PU.1)	PU.1-F	ccagctcagatgaggaggag	151bp
	PU.1-R	caggtccaacaggaactggt	
Nuclear Factor, Erythroid 2 Like 2 (Nrf2)	Nrf2-F	aaaccagtggatctgccaac	190bp
	Nrf2-R	gcaatgaagactgggctctc	
TEA Domain Transcription Factor 4 (Tead4)	TEAD4-F	gctccttctatggggtctcc	206bp
	TEAD4-R	gtgcttgagcttgtggatga	
Promyelocytic Leukemia 4 (Pml-IV)	PML-IV-F	ggagcaggatagtgcctttg	198bp
	PML-IV-R	ctggccatctcctcgtagtc	
DEAD-Box Helicase 6 (P54)	P54-F	acaagcctgtcattcctgct	187bp
	P54-R	ttcctctttccccatctcct	
Regulatory Factor X2 (Rfx2)	RFX2-F	cctctcttcctgccagtgac	198bp
	RFX2-R	agttacggatggcctgtgtc	
GATA Binding Protein 2 (GATA2)	hGATA2-F	gtcactgacggagagcatga	232bp
	hGATA2-R	gccttctgaacaggaacgag	
GATA3	hGATA3-F	ctcattaagcccaagcgaag	205bp
	hGATA3-R	tttttcggtttctggtctgg	

Table 6. Primer sequences that were used to examine the endogenous transcription factor expression in the examined cell lines.

2.8.3 Primer sequences for ChIP experiments

The following list includes the primer pairs that were used in ChIP experiments, to amplify the proposed GATA3 and Pax5 binding sites, in the *ADGB* promoter and enhancer.

Amplicon:	**Primer name:**	**Primer sequence (5´ → 3´):**	**Product size:**
Binding site 1 of GATA3 in the *ADGB* promoter	ChIP-Primer 1-F (GATA3) *ADGB*-Prom	cagcaacttttggcaagaca	217bp
	ChIP-Primer 1-R (GATA3) *ADGB*-Prom	ttggctgaaaacagcacaac	
Binding site 2 of GATA3 in the *ADGB* promoter	ChIP-Primer 2-F (GATA3) *ADGB*-Prom	agatgtttgcagggcagtct	181bp
	ChIP-Primer 2-R (GATA3) *ADGB*-Prom	ggctctccactcagcatagc	
Binding site 5 of GATA3 in the *ADGB* promoter	ChIP-Primer 5-F (GATA3) *ADGB*-Prom	gcgctttgcatcttatcctc	184bp
	ChIP-Primer 5-R (GATA3) *ADGB*-Prom	cttattagggcgcaggctta	
Binding site 6 of GATA3 in the *ADGB* promoter	ChIP-Primer 6-F (GATA3) *ADGB*-Prom	ctcgaaaccaggtggtctgt	230bp
	ChIP-Primer 6-R (GATA3) *ADGB*-Prom	cttattagggcgcaggctta	
Binding site 1 of GATA3 in the *ADGB* enhancer	ChIP-Primer 1-F (GATA3) *ADGB*-Enh	tgggctttgtcatcattcag	192bp
	ChIP-Primer 1-R (GATA3) *ADGB*-Enh	Ggataatgcaagggacgaaa	
Binding site 3 of GATA3 in the *ADGB* enhancer	ChIP-Primer 3-F (GATA3) *ADGB*-Enh	tgaacgagtggctgtgctac	190bp
	ChIP-Primer 3-R (GATA3) *ADGB*-Enh	atcgaagccttgtctgctgt	
Binding site 4 of GATA3 in the *ADGB* enhancer	ChIP-Primer 4-F (GATA3) *ADGB*-Enh	tgaacgagtggctgtgctac	160bp
	ChIP-Primer 4-R (GATA3) *ADGB*-Enh	ctcagtgtgctgggctgtta	
Binding site 6 of GATA3 in	ChIP-Primer 6-F (GATA3) *ADGB*-Enh	gcctggacctgatcaagaaa	210bp

the *ADGB* enhancer	ChIP-Primer 6-R (GATA3) *ADGB*-Enh	Gcagtgtgatgcgttgctat	
Binding site 2 of Pax5 in the *ADGB* promoter	ChIP-Primer 2-F (Pax5) *ADGB*-Prom	ctaggcctaagggctgtgtg	233bp
	ChIP-Primer 2-R (Pax5) *ADGB*-Prom	gcagcaaataggagccactc	
Binding site 3 of Pax5 in the *ADGB* promoter	ChIP-Primer 3-F (Pax5) *ADGB*-Prom	gcgctttgcatcttatcctc	184bp
	ChIP-Primer 3-R (Pax5) *ADGB*-Prom	cttattagggcgcaggctta	
Binding site 4 of Pax5 in the *ADGB* promoter	ChIP-Primer 4-F (Pax5) *ADGB*-Prom	agggaaaaactagcccaagg	151bp
	ChIP-Primer 4-R (Pax5) *ADGB*-Prom	caggtggtcctttgcagact	
Binding site 6 of Pax5 in the *ADGB* promoter	ChIP-Primer 6-F (Pax5) *ADGB*-Prom	ctcgaaaccaggtggtctgt	179bp
	ChIP-Primer 6-R (Pax5) *ADGB*-Prom	ttgccttgggctagtttttc	
Binding site 2 of GATA3 in the *ADGB* enhancer	ChIP-Primer 2-F (Pax5) *ADGB*-Enh	acagcagacaaggcttcgat	183bp
	ChIP-Primer 2-R (Pax5) *ADGB*-Enh	aagtgattgccaaagccaac	
Binding site 3 of GATA3 in the *ADGB* enhancer	ChIP-Primer 3-F (Pax5) *ADGB*-Enh	gcagacaaggcttcgatctc	180bp
	ChIP-Primer 3-R (Pax5) *ADGB*-Enh	aagtgattgccaaagccaac	
Binding site 4 of GATA3 in the *ADGB* enhancer	ChIP-Primer 4-F (Pax5) *ADGB*-Enh	tgtcatatgaaagccgcaaa	204bp
	ChIP-Primer 4-R (Pax5) *ADGB*-Enh	cattaggtgggtccaggaaa	
Binding site 5 of GATA3 in the *ADGB* enhancer	ChIP-Primer 5-F (Pax5) *ADGB*-Enh	gttggctttggcaatcactt	158bp
	ChIP-Primer 5-R (Pax5) *ADGB*-Enh	Agggctcaaattcacagcag	

Table 7. Primer sequences that were used in ChIP experiments to amplify the transcription factor binding sites in the *ADGB* promoter and enhancer.

2.8.4 Primer sequences for site-directed mutagenesis

The following list displays the primer pair that was used for the site-directed mutagenesis experiments, to mutate the GATA3 binding sites in the *ADGB* promoter.

Amplicon:	Primer name:	Primer sequence (5´ → 3´):
pGL3B-1981AP-Mut	pGL3B-1981AP-Mut-F	Gatgccaaaagaattagggaat<u>t</u>taagccttagaaatcaaaagatctacatc
	pGL3B-1981AP-Mut-R	Gatgtagatcttttgatttctaaggctta<u>a</u>attccctaattcttttggcatc

Table 8. Primer sequences that were used to mutate the GATA3 binding site in the 1981bp *ADGB* promoter. The changed bases for the mutation are underlined.

2.9 List of plasmids employed in this thesis

The following list includes the plasmids that were used to observe the alteration of gene expression, in the transfection experiments.

Plasmid:	Properties:	Origin:
pGL3-Basic Vector	Luciferase reporter vector, containing the luciferase gene; neither a promoter nor an enhancer.	Promega, Madison, WI, USA
pGL3B-431AP	Luciferase reporter vector, containing the luciferase gene and the 431bp *ADGB* promoter.	
pGL3B-1031AP	Luciferase reporter vector, containing the luciferase gene and the 1031bp *ADGB* promoter.	
pGL3B-1981AP	Luciferase reporter vector, containing the luciferase gene and the 1981bp *ADGB* promoter.	
pGL3B-431AP-AE	Luciferase reporter vector, containing the luciferase gene, the 431bp *ADGB* promoter and the *ADGB* enhancer.	
pGL3B-1031AP-AE	Luciferase reporter vector, containing the luciferase gene, the 1031bp *ADGB* promoter and the *ADGB* enhancer.	
pGL3B-1981AP-AE	Luciferase reporter vector, containing the luciferase gene, the 1981bp *ADGB* promoter and the *ADGB* enhancer.	
pGl3B-AE	Luciferase reporter vector, containing the luciferase gene and the *ADGB* enhancer.	
pGL3-Enhancer Vector	Luciferase reporter vector, containing the luciferase gene and a SV40 enhancer.	Promega, Madison, WI, USA

pGL3E-431AP	Luciferase reporter vector, containing the luciferase gene, a SV40 enhancer and the 431bp *ADGB* promoter.	
pGL3E-1031AP	Luciferase reporter vector, containing the luciferase gene, a SV40 enhancer and the 1031bp *ADGB* promoter.	
pGL3E-1981AP	Luciferase reporter vector, containing the luciferase gene, a SV40 enhancer and the 1981bp *ADGB* promoter.	
pGL3-Promoter Vector	Luciferase reporter vector, containing the luciferase gene and a SV40 promoter.	Promega, Madison, WI, USA
pGL3P-AE	Luciferase reporter vector, containing the luciferase gene, a SV40 promoter and the *ADGB* enhancer.	
pcDNA3.1	Empty vector, without any gene expression.	Invitrogen
pRL-TK-Renilla	Vector containing the wild-type Renilla gene.	Promega, Madison, WI, USA
pcDNA3.1-GATA3	Vector containing the GATA3 gene.	gift from Christof Dame (Berlin)
pRK5-Myc-TEAD4	Vector containing the TEAD4 gene.	Addgene, deposit by Kunliang Guan
pLPC-PML-IV	Vector containing the PML-IV gene.	Addgene, deposit by Gerardo Ferbeyre
pCMV-Myc-p54	Vector containing the P54 gene.	Addgene, deposit by Benjamin Blencowe
pLPC-REST	Vector containing the REST gene.	gift from Cormac Taylor (Dublin)
pcDNA3.1-Nrf2	Vector containing the Nrf2 gene.	gift from Cormac Taylor (Dublin)
pcDNA3.1-Nrf3	Vector containing the Nrf3 gene.	gift from Cormac Taylor (Dublin)
LZRSpBMN-linker-IRIS-EGFP-PU.1	Vector containing the PU.1 gene.	Addgene, deposit by Ellen Rothenberg
pCCL-cppt-PKG-WPRE-Pax5	Vector containing the Pax5 gene.	Addgene, deposit by Malin Parmar
pCDNA 3.1-HP1Gamma	Vector containing the HP1Gamma gene.	Addgene, deposit by Naoko Tanese

pKH3-Trim28	Vector containing the Trim28 gene.	Addgene, deposit by Fanxiu Zhu
pcDNA3.1-hygro-hUSF1	Vector containing the USF1 gene.	gift from Cheng-Jun Hu (Colorado)
pCMV6-XL5-GABPα	Vector containing the GABPα gene.	OriGene Technologies, Inc, Maryland, USA
pCMV6-XL5-CEBPβ	Vector containing the CEBPβ gene.	OriGene Technologies, Inc, Maryland, USA
pcDNA3.1-GATA2	Vector containing the GATA2 gene.	gift from Christof Dame (Berlin)
pcDNA 3.1-RFX2	Vector containing the RFX2 gene.	gift from Zijie Sun (Stanford)

Table 9. Plasmids that were used in the transfection experiments.

2.10 Preparation of LB broth

One liter of lysogeny broth medium (LB medium) was prepared by mixing 10g of Trypton, 5g of NaCl and 5g of yeast extract in a glass container. The container was filled up to one liter with ddH_2O and autoclaved.

LB-agar plates were prepared by mixing 17.5g of antibiotic medium with 15g of agar, which was filled up to one liter with ddH_2O and autoclaved.
After the temperature cooled down, 1ml of the specific antibiotic (in this case: ampicillin) was added to the solution. 10 – 15ml of the solution was distributed per petri dish (92 x 16mm) and left to dry.

2.11 Preparation of DH5α *E. coli* competent cells

DH5α competent *E. coli* cells were prepared following the Cohen principal. The Cohen principal implies a high concentration of cations that leads to a better permeability of the bacteria, which increases the likelihood of foreign DNA uptake (Cohen et al., 1972).
By inoculating one aliquot of incompetent DH5α *E. coli* cells in 5ml of LB medium, a pre-culture was grown for 12 hours, shaking (250rpm) at 37°C. The obtained pre-culture was then inoculated in 200ml of pre-warmed LB medium and incubated shaking

(250rpm) at 37°C, until it reached an optical density of 0.6 – 0.8. After the bacteria had replicated enough, the probe was put on ice for 10 minutes, followed by 15 minutes of centrifugation (3500rpm) at 4°C to separate the supernatant from the pelleted cells. The cells were resuspended in 15ml ice-cold TFB1 buffer and left to incubate for 15 minutes on ice. After another centrifugation under the same conditions as previously described, the cell sediment was resuspended in 15ml TFB2 buffer and put on ice to incubate for an additional 15 minutes. The mixture was divided into 100µl aliquots and shock-frozen in liquid nitrogen. The aliquots were stored at -80°C.

2.12 Cloning of plasmids containing *ADGB* promoter fragments & *ADGB* enhancer

Three *ADGB* promoter fragments and the *ADGB* enhancer gene region were cloned into three different vectors (pGL3-Basic Vector, pGL3-Enhancer Vector, pGL3-Promoter Vector). pGL3-Basic Vector contains a luciferase gene, pGL3-Enhancer Vector contains a luciferase gene and a viral, SV40 enhancer and pGL3-Promoter Vector contains a luciferase gene and a viral, SV40 promoter.
The cloning consisted of four major steps: The amplification of the DNA fragments via PCR, the DNA digestion, the ligation of the digested DNA sections and the transformation of the final DNA construct to competent cells.

2.12.1 PCR amplification of gene sections

The primers were ordered at Europrim, Invitrogen (https://www.thermofisher.com/de/de/home/brands/invitrogen.html).
The first step was to determine the most suitable annealing temperature (55°C – 70°C) for the designed primers by performing a gradient PCR. Once the optimal annealing temperature was selected, a conventional PCR was performed to amplify the desired amplicon. As a template 1µl of genomic DNA (gDNA) mixture (3500µg/µl) of HEK293T, MCF7 and REPC cells was used.

The PCR mixture consisted of the following components:

5µl Phusion HF buffer

0.5µl (10mM) dNTPs

1µl (10mM) forward primer

1µl (10mM) reverse primer

1µl gDNA template

0.25µl Phusion DNA Polymerase

16.25µl ddH_2O

The mixture was filled in reaction tubes, vortexed and placed in a thermocycler.

The PCR was programmed as following:

Initial phase:

95°C 5 min → Initialization: Break and separation of the nucleic acid base pairs

Cycles 1 – 35:

98°C 15 sec → DNA denaturation

72°C 135 sec → Annealing of the primers and elongation of the polymerase

72°C 5 min → Completion of unfinished reactions

The PCR amplification was confirmed on a 1.8% agarose gels and examined in a gel imaging system.

The 1.8% agarose gels were prepared by mixing 100ml TAE buffer with 1.8g agarose and 10µl ethidium bromide (1% solution) and heating it in a microwave, until the solution became precipitant free. The liquid agarose was filled in the suitable container and a comb was inserted to create pockets.

After the gel polymerized, 15µl of DNA ladder (100bp and 1Kb) and 20µl of each PCR probe was filled in the pockets to analyze the amplification of the primers and confirm the desired size and accuracy of the amplification.

2.12.2 DNA purification using QIAEX II Gel Extraction Kit

The amplified DNA was purified using the binding capacity of Silica beads. The purification was performed following the protocol of the QIAEX II Gel Extraction Kit.

The final concentrations of the obtained DNA samples were measured by a micro plate spectrophotometer. Tris-Cl was used as blank for the calibration of the machine.

2.12.3 Digestion of the PCR fragments

For the conventional cloning of DNA segments into plasmids, the desired genomic sequence and the corresponding plasmid were digested with restriction enzymes to generate fragments with compatible ends.

In the experimental setup of this project, 1µg of each DNA sample was digested with two different restriction endonucleases from New England BioLabs.

In case the utilized buffers of the two digestion enzymes were not compatible, the digestions were performed sequentially. In between the sequential digestions, the product of the first digestion was purified, using QIAEX II Gel Extraction Kit, before proceeding to the second digestion.

DNA-templates:	Restriction Enzymes:	Buffer:	Final construct:
pGL3-Basic Vector	BglII & HindIII	3.1/2.1	pGL3B-431AP
431bp *ADGB* promoter			
pGL3-Basic Vector	BglII & HindIII	3.1/2.1	pGL3B-1031AP
1031bp *ADGB* promoter			
pGL3-Basic Vector	NcoI & XhoI	CutSmart	pGL3B-1981AP
1981bp *ADGB* promoter			
pGL3B-431AP	KpnI-HF & NheI	CutSmart	pGL3B-431AP-AE
ADGB enhancer	KpnI-HF & XbaI		
pGL3B-1031AP	KpnI-HF & NheI	CutSmart	pGL3B-1031AP-AE
ADGB enhancer	KpnI-HF & XbaI		
pGL3B-1981AP	KpnI & NheI	1.1	pGL3B-1981AP-AE
ADGB enhancer			
pGL3-Basic Vector	KpnI-HF & NheI	CutSmart	pGL3B-AE
ADGB enhancer	KpnI-HF & XbaI		
pGL3-Enhancer Vector	BglII & HindIII	3.1/2.1	pGL3E-431AP
431bp *ADGB* promoter			
pGL3-Enhancer Vector	HindIII-HF & NheI-HF	CutSmart	pGL3E-1031AP
1031-bp *ADGB* promoter	HindIII-HF & SpeI-HF		
pGL3-Enhancer Vector	NcoI & XhoI	CutSmart	pGL3E-1981AP
1981bp *ADGB* promoter			
pGL3-Promoter Vector	KpnI-HF & NheI	CutSmart	pGL3P-AE
ADGB enhancer	KpnI-HF & XbaI		

Table 10. Overview of the DNA templates and vectors that were used to clone the desired plasmids. The table furthermore shows the implied restriction enzymes with the corresponding buffers and points out the name of the final plasmid constructs.

The digestion mixture consisted of the following:

1µl of restriction endonuclease 1

1µl of restriction endonuclease 2

5µl of restriction endonucleases buffer

1µg of DNA template

Filled up with ddH_2O to a total volume of 50µl.

The digestion mixture was transformed to a reaction tube, vortexed and incubated for 2.5 hours at 37°C.

After the incubation, the digested DNA fragments were purified using QIAEX II Gel Extraction Kit.

The final concentrations of the digested DNA samples were measured by a micro plate spectrophotometer. Tris-Cl was used as blank for the calibration of the machine.

2.12.4 Ligation of the digested PCR fragments

Different ratios of the digested plasmids and DNA fragments, plus T4 DNA Ligase from New England BioLabs as ligation enzyme, were used to obtain the desired ligated product.

The required ratios of the plasmid and the insert were calculated by New England BioLabs Calculator. For this calculation, the DNA base pair lengths of the constructs and the premise to use 50ng of vector DNA were taken into consideration, as suggested by the protocol. For the constructs of this thesis, all ligations were performed using insert-to-vector ratios of 3:1, 5:1 and 7:1.

The final ligation mixture consisted of the following:

50µg plasmid DNA mass

Varying µg of inserted DNA mass

1µl T4 DNA Ligase

2µl T4 DNA Ligase buffer

Filled up with ddH_2O to a total volume of 20µl.

The mixture was filled in a reaction tube, vortexed and incubated for 1.5 hours.

2.12.5 Transformation

5µl of the solution which contains the ligated plasmids was pipetted on top of an aliquot of DH5α *E. coli* competent cells and gently flicked a few times to assure an adequate distribution in the bacteria containing reaction tube.
After 15 minutes incubation on ice, the bacteria were heat-shocked at 42°C for one minute to ensure an opening of the pores and a proper uptake of the ligated plasmids. The reaction tubes were put back on ice afterwards, for an additional two minutes.
500µl of LB medium was added to the tubes and they were placed on the incubation shaker (250rpm) at 37°C. After one hour, 200µl of the bacteria containing solution was spread on a pre-warmed, ampicillin containing LB agar plate.
The ampicillin is important to ensure a selection advantage for the ampicillin resistant DH5α *E. coli* competent cells, opposed to potential precipitation. The agar plates were finally left to incubate for 12 hours at 37°C.

2.12.6 Replication of colonies

Single bacteria colonies were picked from the LB agar plate with a pipette tip and inoculated in a 15ml centrifuge tube with 2ml of LB medium and 2µl of ampicillin per tube. The tubes were incubated, rotating (250rpm) for 12 hours at 37°C.

2.12.7 DNA preparation of cloned plasmids, following a home-made protocol

The preparation of the plasmid DNA was performed following the rapid alkaline extraction procedure introduced by Birnboim and Doly (Birnboim and Doly, 1979).
To extract the plasmid DNA from the bacteria, the tubes were first centrifuged for 25 minutes (3500rpm) and the supernatant was discarded.
The pelleted cells were resuspended in the Resuspension Buffer (Miniprep) to prevent a degradation of the DNA and transferred to a 1.5ml reaction tube. In the following step, 200µl of Lysis Buffer (Miniprep) was added per tube. To ensure a proper lysis of the bacteria, the tubes were inverted a couple of times, followed by an incubation period of three minutes.
The lysis process was stopped by adding 150µl Neutralization Buffer (Miniprep) to the tube to precipitate the chromosomal DNA with the SDS from the Lysis Buffer (Miniprep).

The solutions were mixed by inversion and put on ice to incubate for five minutes. To separate the cell debris from the DNA containing solution, the reaction tubes were centrifuged at 13000rpm for 8 minutes.
The obtained supernatant was put into a fresh reaction tube and 270µl of isopropanol was added to further precipitate the DNA. After 8 minutes of centrifugation (1300rpm), the supernatant was discarded and 800µl of 70% ethanol was added to wash the pelleted DNA.
After three minutes of centrifugation (13000rpm), the supernatant was discarded and the pellet was dried at room temperature. The dried DNA pellet was finally resuspended in 50µl ddH_2O.

2.12.8 Digestion of the cloned plasmids

The prepped DNA of the cloned plasmids was digested with the same restriction enzymes that were used for the first, initial digestion. The second digestion was used to evaluate if the plasmids actually contain the desired insert.
The digestion mixture consisted of the following:

0.2µl of restriction endonuclease
1µl of prepped plasmid DNA
1µl of restriction endonuclease buffer
filled up to 10µl final volume with ddH_2O

The mixture was incubated for one hour at 37°C and the digested plasmids were analyzed on a 1.8% agarose gel, visualized by a gel imaging system.

2.12.9 DNA purification using a GenJET Plasmid Miniprep Kit

The DNA was transformed again into DH5α *E. coli* competent cells and purified, using the GenJET Plasmid Miniprep Kit.
In this case, the kit was used to ensure a higher purity of the final product. The included Thermo Scientific GeneJet Spin Columns contain a silica-membrane for higher DNA purity. The final plasmid DNA was prepared following the instructions of the manufacturer.
The final concentrations of the obtained DNA samples were measured by a micro plate spectrophotometer, the elution buffer of the GenJET Plasmid Miniprep Kit was used as

blank for the calibration of the machine. The final product was sequenced by Microsynth (https://www.microsynth.ch/), using the primers RVprimer3 and GLprimer2, provided by the company.

2.13 Dual luciferase reporter assays

The transcriptional effects of the cloned *ADGB* promoter fragments and the *ADGB* enhancer and their response to overexpression of specific transcription factors in selected cell lines were measured using a dual luciferase reporter assay.

The dual luciferase assay approach is based on two different luciferases. One enzyme is obtained from the firefly *Phontinus pyralis* and the other one originates from the sea pansy *Renilla reniformis*.

The firefly luciferase gene is expressed by the cloned vectors and allows a quantification of gene expression depending on the impact of the transcriptional activity of the *ADGB* gene fragments and/or the implied transcription factors. The sea pansy luciferase gene is independently added to the reactions and acts as an internal control of the transfections and overall cell growth.

Plasmids encoding the corresponding luciferases genes were transfected in HeLa, MCF7 and HEK293T cells, followed by cell lysis and measurement of the luciferase activities. A micro plate luminometer was used to measure the light emission which was linked to a computer program (Microwin 2010) for further analyzation.

The alteration of the transcriptional activity of the *ADGB* promoter and enhancer constructs is displayed as the ratio of firefly luciferase light signal and *Renilla* luciferase light signal (in relative luminescence units, RLU).

All results were plotted in GraphPad Prism.

2.13.1 Transfections

Transfections were performed via lipofection with a substance called Rotifect, which contains cationic lipids. The substance forms complexes with the desired nucleic acids and acts as a vehicle to transport it across the cell membrane using electrostatic interactions.

The cell medium used for transfections was antibiotic-free.

The experimental setup included 100ng of cloned construct, 300ng of transcription factor, 1ng of *Renilla* vector and 10ng/310ng of a vector that does not contain a gene (empty vector). The empty vector was added to the reaction to ensure an equal amount of DNA mass affecting the cells.

10ng of empty vector was added in this context to the reactions that analyzed the overexpression of 300ng transcription factor and 310ng of empty vector to the reference reactions without transcription factor overexpression.

The transfection procedure was performed under the sterile bench.

24 hours before the transfections, the cells were plated to a 30% confluence (approximately 0.4 – 2.0 x 10^3 cells/well) in 24-well plates to assure an approximate 50% – 60% confluence of adherent cells the day of transfections.

To start the transfections, the desired vectors were diluted in 50µl of serum and antibiotic-free medium in a 1.5ml tube. 2µl of Rotifect / transfected plasmid was diluted individually. After the dilution, the two components were pipetted together. The *Renilla* vector was added individually as the last substance of the mixture.

The final mixture was left to incubate for 15 – 20 minutes to form nucleic acid - lipid complexes. The plated cells were washed in the meantime with PBS and the surrounding medium was replaced by antibiotic-free medium.

After the incubation time, the obtained complexes were pipetted drop-by-drop on the cells and the cells returned to the incubation chamber for 24 hours.

The medium was exchanged to antibiotic and serum containing medium the following day. After an additional 24 hours of incubation, the cells were ready for further analysis.

2.13.2 Lysis of cells

The cells were first washed with PBS. Furthermore, 100µl of passive lysis buffer was added per well and left to incubate, rotating (25rpm) for 15 minutes to ensure total cell lysis.

2.13.3 Analysis of Dual-Glo luciferase assays

To analyze the final luciferase gene expression, 20µl of cell lysate was pipetted in duplicates on 96 well plates.

The firefly luciferase was measured by adding 20µl of *Luciferase Assay Reagent II* (*Lar II*), which contains the firefly substrate D-Luciferin (Figure 7A). After the measurement, 20µl of *1x Stop & Glo* reagent was added per well to measure the *Renilla*-luciferase (Figure 7B).

The *Stop & Glo* reagent contains on the one hand Dehydroluciferyl-AMP, which stops the firefly luciferase, and on the other hand Coelenterazin, which is a *Renilla*-luciferase substrate.

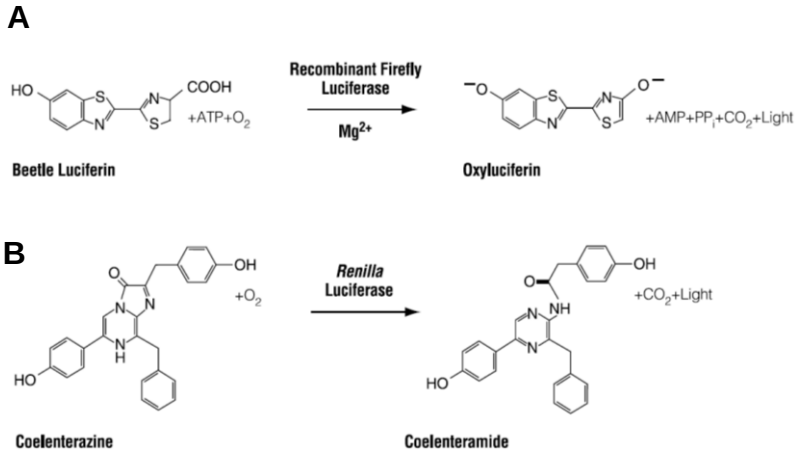

Fig. 7. Bioluminescent reactions catalyzed by firefly and *Renilla* luciferases.
(A): Schematic representation of the enzymatic firefly luciferase reaction: Beetle Luciferin + Recombinant Firefly Luciferase + magnesium → Oxyluciferin + light emission.
(B): Schematic representation of the enzymatic *Renilla* luciferase reaction: Coelenterazine + oxygen + *Renilla* Luciferase → Coelenteramide + carbon dioxide + light emission.
(Figures edited from the product manual)

2.14 Analysis of the endogenous transcription factor expression in the examined cell lines

The endogenous expression of the examined transcription factors in the cell lines (HeLa, HEK293T, MCF7), that were used in the Dual-Glo luciferase assay experiments, was examined via real time, quantitative PCR (qPCR). The corresponding RNA was extracted from the cells and converted into complementary DNA (cDNA) by a reverse transcriptase PCR (RT-PCR). The obtained cDNA was used in the qPCR.

2.14.1 RNA extraction

The RNA extraction of the Qiagen-Kit is based on the single step method of RNA isolation, first established by Chomczynski and Sacchi in 1987 (Chomczynski and Sacchi, 1987).

The lysis takes place in presence of RNases inactivating, highly denaturing guanidine thiocyanate containing buffer to ensure a high grade of intact RNA. The RNA was bound

to the provided spin columns that contain a purifying silica gel membrane. The final product was eluted in provided RNase-free ddH_2O.
The examined cells were first plated on petri dishes (92 x 16mm) and incubated until they reached confluency (approximately 1 x 10^7cells / petri dish).
On the day of cell harvest, the medium was extracted and the cells were washed with PBS, followed by a cell collection via cell scraper. The cell suspension was pipetted into a tube and centrifuged for five minutes at 2000rpm.
The obtained cell pellet was used to extract the RNA, following the suggested protocol of the manufacturer.
The final RNA concentration was measured by a micro plate spectrophotometer. RNase-free water was used as blank for the calibration of the machine.

2.14.2 Reverse transcriptase PCR (cDNA synthesis)

The following synthesis of the cDNA consists of two major steps. The first step is to anneal oligo-dTs to the polyA tail of the RNA to generate cDNA, which serves as a starting point for the second step, the actual RT-PCR.
To prepare the RNA, 1 – 3µg of RNA was mixed with 2µl of oligo-dTs and filled up to a final volume of 9,5µl with ddH_2O. The mixture was incubated for 10 minutes at 68°C and cooled down for 5 – 10 minutes on ice.
For the second step, 9.5µl of PCR mixture was added to the prepared RNA.
The PCR mixture consisted of the following:

4µl 5 x Moloney Murine Leukemia Virus (M-MLV) Reverse Transcriptase buffer
5µl dNTPs
1µl ddH_2O
0.5µl M-MLV Reverse Transcriptase

The RT-PCR program was divided in two steps. The first circle amplified the cDNA at 42°C for one hour, followed by the second circle of denaturation of the cDNA at 92°C for 10 minutes.

2.14.3 Quantitative PCR (qPCR)

Every cDNA probe of interest was analyzed by primer pairs which amplify the genes of the transcription factors, that were used in the Dual-Glo luciferase assays. Additionally, two primer pairs were used as internal control to amplify genes (L28 and HPRT) that demonstrate a constant expression pattern in every probe, as described before (Hoogewijs et al., 2008).

The qPCRs were performed in duplicates, using the Takyon No ROX SYBR 2X Master Mix.

A standardized PCR reaction consisted of the following:

10µl Takyon No ROX SYBR 2X Master Mix

8.8µl ddH_2O

0.4µl forward-Primer (10µM)

0.4µl reverse-Primer (10µM)

0.4µl cDNA

The mixtures were pipetted in a 96-well plate, sealed with a plastic foil and analyzed by the thermo cycler (CFX96).

The corresponding PCR program was divided in two parts. The first part consisted of the standard PCR steps, during the second part the temperature was increased 1°C / cycle, reaching from 65°C to 95°C.

The qPCR was programmed as following:

Initial phase:

95°C 5 min → Initialization: Break and separation of the nucleic acid base pairs

Cycles 1-45:

95°C 15 sec → DNA denaturation

60°C 1 min → Annealing of the primers and elongation of the polymerase

65°C 30 sec → Preheating for the second part

Cycles 46-75:

65°C 5 sec → Temperature dependent gene amplification

+1°C/Cycle

The expression level of the examined transcript is detected by the individual fluorescence signal of the probe and the relative expression ratio is calculated, compared to the internal reference gene transcripts.
The final data was visualized with Bio-Rad CFX Manager.

2.15 Confirmation of transcription factor binding via Chromatin immunoprecipitation (ChIP)

The ChIP experiments were performed to examine transcription factor binding to specific gene sequences of the *ADGB* promoter and enhancer.
The ChIP experiment is divided into four major steps: cell harvesting, fragmentation of the DNA via sonication, specific antibody binding and PCR amplification.

2.15.1 Chromatin immunoprecipitation (ChIP)

In preparation for the actual ChIP experiments, the utilized Pierce Protein A/G Agarose beads were pre-blocked and washed thoroughly to reduce the possibility of unspecific binding. In the meantime, the MCF7 and Hep3B-cells were grown to an approximate 95% confluence on 150mm plates in the incubator.

55µl of the bead solution was used per sample and distributed into reaction tubes. To separate the actual beads from the solution, the tubes were centrifuged at maximum speed (13,200 rpm) for one minute.
The pelleted beads were sequentially washed three times with 1ml ChIP IP Dilution Buffer, centrifuged and then resuspended in 1ml ChIP Preblocking Buffer. Finally, 2µl of salmon sperm DNA was added to the suspension and the beads were incubated rotating for 12 hours at 4°C.
As a next step, the beads were pelleted again, using a centrifuge at maximum speed (13,200 rpm) for one minute. The obtained beads were then washed twice with ChIP IP Dilution Buffer, centrifuged and filled up to the original volume with the same buffer.

Once a final cell count of approximately 3 – 5,000,000 cells / plate was reached, the cells were fixed with 1% formaldehyde and incubated for 10 minutes at 37°C.

After the incubation, glycine was added to a final concentration of 125mM to stabilize the reaction, which was followed by an additional incubation for five minutes at room temperature.
The cells were then washed with ice-cold PBS twice and collected with a cell scraper in a reaction tube. To separate the cells from the remaining supernatant, they were centrifuged (3500rpm) for five minutes at 4°C.
Next, the pelleted cells were sequentially washed with 1ml of ice-cold PBS + PMSF, ChIP Buffer 1 and ChIP Buffer 2. In between the washing steps, the cells were centrifuged (3500rpm) for five minutes at 4°C and the supernatant was discarded.
The pelleted cells were lysed adding 500µl of ChIP Lysis Buffer and 50µl of protease inhibitor, followed by a 10 minute incubation on ice. The cell suspension was distributed between two safe-lock tubes and stored at -80°C.

The sonication in the ChIP experiments aims at the shearing of the DNA in pieces around 500bp via ultrasonic frequencies in pieces. The cell suspension samples were sonicated with a maximum frequency of 30W in this case. The obtained size was analyzed running 2µl of sample on a 1.8% agarose gel, visualized by a gel imaging system.
One cycle for MCF7 cells consisted of 30 seconds sonication on ice, followed by a one-minute cool-down on ice.
The Hep3B cycles consisted of 15 seconds sonication on ice and 30 seconds cool-down.
A 50µl aliquot of each sonicated DNA sample was stored at -80°C to obtain an input control of the reactions. The rest of the material was distributed to reaction tubes and centrifuged at maximum speed (13,200 rpm) for 10 minutes at 4°C to separate the sheared DNA from the remaining cell debris. 200µl of the supernatant was collected in a new reaction tube and diluted with 800µl ChIP IP Dilution Buffer.
The final product was then incubated with the pre-blocked beads for 2.5 hours at 4°C, to prevent unspecific beads – DNA binding in later stages of the experiment.
The beads were separated from the DNA by centrifugation at maximum speed (13,200 rpm) for two minutes at 4°C. 250µl of each DNA sample was distributed to separate reaction tubes.

Either 5µl of one of the specific transcription factor antibodies (GATA3 or Pax5), 5µl of an unspecific antibody (actin) or 5µl of ddH_2O, as a control was added to the reaction tube to immunoprecipate the DNA samples. The reactions were subsequently incubated rotating (35rpm) for 12 hours at 4°C.

30µl of blocked beads was added next to each sample, followed by two hours of incubation at room temperature to bind the DNA – antibody complexes. To pellet the beads, the tubes were centrifuged at maximum speed (13,200 rpm) for 90 seconds and the supernatant was discarded.

The beads were then washed as explained in the following: First, 1ml of the solutions listed below was added, followed by 10 minutes of rotating (35rpm) incubation, 90 seconds of centrifugation and discard of the supernatant. The beads were first washed with ChIP-TSEI Buffer, followed by four times ChIP-TSEII Buffer, one-time ChIP Buffer III and finally three times Tris-EDTA.

The following elution of the DNA – antibody complexes from the beads was achieved by adding 75µl freshly prepared ChIP-Elution Buffer to the pelleted beads and incubation for 15 minutes at 65°C. To ensure proper distribution during the incubation time, the probes were vortexed every two minutes.

The reaction tubes were then centrifuged at maximum speed (13,200 rpm) for two minutes and the obtained supernatant was transferred to a new reaction tube. To raise the obtained yield of DNA concentration, this procedure was repeated twice.

After the elution, 15µl of 5M NaCl was added to the reaction tubes and the previously stored input control to avoid SDS precipitation in the final product. The samples were left to incubate for 12 hours at 65°C.

The final product was purified using the High Pure PCR Product Purification Kit.

2.15.2 DNA purification using High Pure PCR Product Purification Kit, Roche

The DNA purification is based on the DNA binding capacity of the provided filtered tubes that contain two layers of glass fiber fleece. The DNA purification steps were followed as suggested by the protocol of the manufacturer.

The concentration of the purified DNA – antibody complexes was measured by a micro plate spectrophotometer. The elution solution of the High Pure PCR Product Purification Kit was used as blank for the calibration of the machine.

2.15.3 PCR amplification of the transcription factor binding sites with GoTaq® G2 DNA-Polymerase

The sequences for the used primers were determined to amplify the section of the *ADGB* promoter and enhancer with highest probability of transcription factor binding. The JASPAR database was used in this context to scan the *ADGB* promoter and enhancer sequences. The primers were designed using Primer3Plus (https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi).

The primers were ordered at Europrim, Invitrogen.

A gradient PCR (55°C – 70°C) was performed first to determine the most suitable annealing temperature. Once the annealing temperature was selected, a conventional PCR was performed to amplify the desired amplicon. The purified, immunoprecipitated DNA was used as DNA template for the PCR.

The PCR mixture consisted of the following:

5µl 5 x GoTaq Green Reaction Buffer
0.5µl 10mM dNTPs
1.25µl (10mM) forward primer
1.25µl (10mM) reverse primer
1µg DNA template
0.125µl GoTaq G2 DNA Polymerase
16.375µl ddH_2O

The mixture was filled into a reaction tube, vortexed and amplified by the thermo cycler.

The PCR was programmed as following:

Initial phase:

95°C	15 sec	→ Initialization: Break and separation of the nucleic acid base pairs

Cycles 1 – 35:

95°C	15 sec	→ DNA denaturation
64°C	15 sec	→ Annealing of the primers
72°C	2 min	→ Elongation of the polymerase
72°C	5 min	→ Completion of unfinished reactions

The PCR amplification was confirmed on a 1.8% agarose gels, visualized in a gel imaging system.

2.16 Site-directed mutagenesis

For the site-directed mutagenesis experiments, GATA3 binding sites in the *ADGB* promoter were examined to mutate the corresponding nucleotide sequence and evaluate the specific impact of these binding sites on the transcriptional activity in transfection experiments using a Dual-Glo Luciferase Assay System.

The JASPAR database was employed to predict the highest probability of GATA3 binding sites in vertebrae in the *ADGB* promoter sequence. In case of the 1981bp androglobin promoter, the sequence 5´-GCTTATCT-3´ in between the nucleobase positions 264 and 271, represents the highest relative score (0.9738), thus the highest likelihood of being the actual transcription factor binding site for GATA3 (Figure 8).

Fig. 8. Matrix-based presentation of the GATA3 transcription factor binding site in the 1981bp *ADGB* promoter, in eukaryotes, as presented at the JASPAR website (http://jaspar.genereg.net/).

The performed ChIP experiments additionally display GATA3 antibody binding to this specific promoter sequence in MCF7 cell DNA, amplified with ChIP primer pair 1 *ADGB* Prom.

The primers for the site-directed mutagenesis were designed using the provided QuikChange Primer Design Program to mutate the sequence from 5´-GCTTATCT-3´ to 5´-GCTTATTT-3´ in the vector construct pGL3B-1981AP. The primers were ordered at Europrim, Invitrogen.

The PCR reaction was prepared following the suggested protocol of the kit.

The PCR was programmed as following:

Initial phase:

95°C	2 min	→ Initialization: Break and separation of the nucleic acid base pairs

Cycles 1 – 18:

95°C	20 sec	→ DNA denaturation
78°C	10 sec	→ Annealing of the primers
68°C	4 min	→ Elongation of the polymerase
68°C	5 min	→ Completion of unfinished reactions

The obtained DNA was digested with 2µl of DPN I restriction enzyme to select for the desired mutation-containing DNA and to sort out the undesired, parenteral DNA. The plasmid DNA was purified and transformed to DH5α *E. coli* competent cells, following the suggested protocol of the kit.

The bacteria were replicated in agar medium and plated on an ampicillin containing plate, as described above.

The DNA was purified using GenJET Plasmid Miniprep Kit.

To confirm the mutation, the final product was sequenced by Microsynth using the primers RVprimer3 and GLprimer2, provided by the company.

The mutated plasmid was used to transfect HeLa, MCF7 and HEK293T cells with and without GATA3 co-transfection, as described in the previous sections.

The alteration of the transcriptional activity was measured and examined using the Dual-Glo Luciferase Assay System.

All results were plotted in GraphPad Prism.

3 RESULTS

3.1 Cloning results

Cloning was performed using the supplied vector backbones of Promega, pGL3-Basic Vector (pGL3B), pGL3-Enhancer Vector (pGL3E) and pGL3-Promoter Vector (pGL3P). Three *ADGB* promoter fragments of 431, 1031 and 1981 base pairs in length were cloned in the multiple cloning regions of pGL3B and pGL3E (Figures 9A, B). The *ADGB* enhancer of 1685 base pairs was cloned in the multiple cloning regions of pGL3B and pGL3P (Figures 9A, C).

Initial cloning efforts aimed to clone the *ADGB* enhancer, 3' of the *luc+* gene, in the multiple cloning region SalI/BamHI of the pGL3B-xAP vector constructs to separate the gene locus of the promoter from the enhancer. After several unsuccessful attempts, the cloning strategy was adapted and the *ADGB* enhancer was cloned 5' of the *ADGB* promoter (Figure 9A).

The circular nature of the vector ensures separate transcription of the promoter and the enhancer nonetheless (Figure 9A).

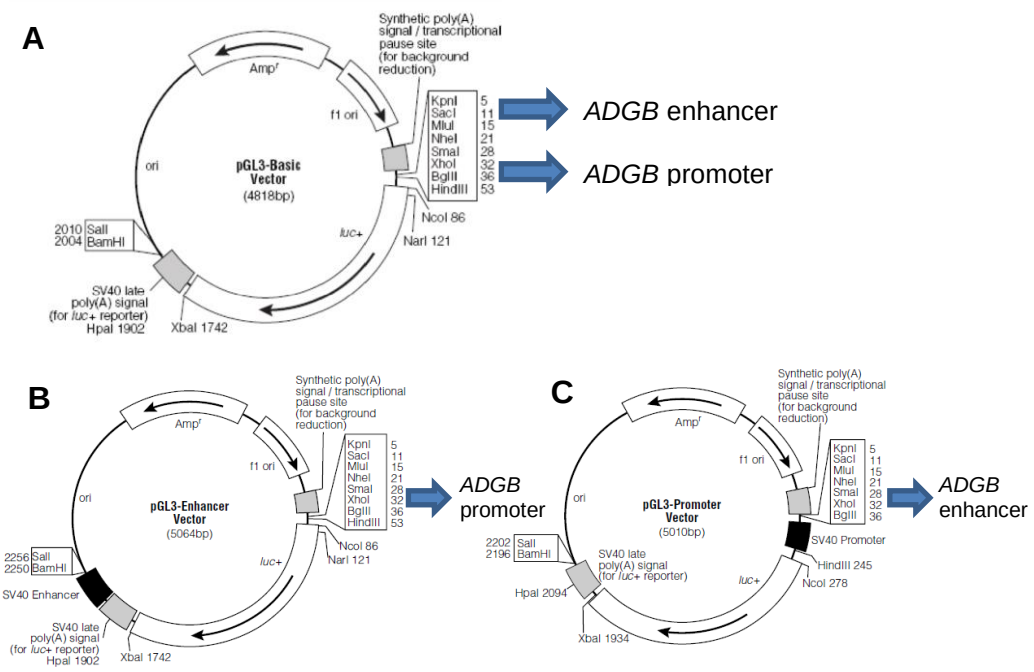

Fig. 9. The utilized vector backbones displaying the name and the bp size of the vector; the multiple cloning regions; the *luc+* gene; the potential restriction sites and other vector features. **(A):** pGL3-Basic Vector and the restriction sites where the *ADGB* enhancer and the promoter fragments were cloned in. **(B):** pGL3-Enhancer Vector and the restriction sites where the *ADGB* promoter fragments were cloned in. **(C):** pGL3-Promoter Vector and the restriction sites where the *ADGB* enhancer was cloned in. (Figures edited from the product manual)

As described in the methods presented in this thesis, the cloning was examined using a digestion of the cloned vector, followed by a 1.8% agarose gel separation, to analyze the size of the fragments. The results were verified by external sequencing.
Two exemplary pictures of the agarose gels are displayed below (Figures 10A, B).

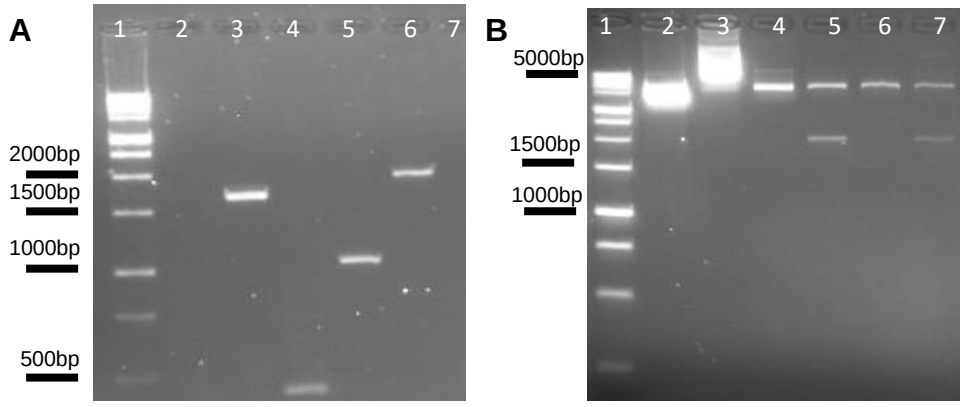

Fig. 10. 1.8% agarose gel electrophoresis displaying the amplified PCR products and the digested vector constructs. **(A):** (1) DNA ladder; (2) empty lane; (3) *ADGB* enhancer; (4) empty lane; (5) 431bp *ADGB* promoter; (5) 1031bp *ADGB* promoter; (6) 1981bp *ADGB* promoter; (7) empty lane. **(B):** (1) DNA ladder; (2) undigested pGL3Basic vector; (3) undigested pGL3B-1031AP; (4) unsuccessful cloning & digestion; (5) successful cloning & digestion; (6) unsuccessful cloning & digestion; (7) successful cloning & digestion.

3.2 Transfections and quantitative PCRs

After successful cloning, the vector constructs were transfected to HeLa, MCF7 and HEK293T cells to assess the *ADGB* promoter and enhancer dependent changes in luciferase gene expression.
Additionally, various transcription factors were co-transfected to examine their impact on the corresponding expression regulation. The transcription factors were chosen based on UCSC implemented ENCODE-based ChIP sequencing data and their proposed binding to the *ADGB* enhancer region.
Whereas some transcription factors display increasing luciferase gene transcription, others decrease the corresponding transcription.

It should be mentioned that some transcription factors display cell line-dependent effects in the examined cell lines. Unfortunately, this effect cannot be avoided and is one reason for the partly differing obtained results among the three cell lines. Due to the most stable

expression pattern in HeLa cells, the cell line was chosen as representative, in case of aberrant results.

Some of the transcription factors seem to have a transcriptional effect on the vector backbone itself. To rule out this bias, the obtained data of transcription factor co-transfection was transformed, according to the following protocol:

1) The means of the duplicates was calculated, e.g. (pGL3B + TF) probe 1 / (pGL3B + TF) probe 2
2) The means of the ratio (firefly expression /*Renilla* expression) was calculated, e.g. (pGL3B + TF) firefly gene expression / (pGL3B + TF) *Renilla* gene expression
3) The ratio of (construct + TF) / (construct without TF) was calculated, e.g. (pGL3B + GATA3) / (pGL3B)
4) The x-fold change in luciferase activity for the individual vector constructs was calculated dividing the obtained value of point 3 of the cloned vector constructs by the obtained value of point 3 for the vector backbone, e.g. (value of point 3 for pGL3B-431AP + GATA3) / (value of point 3 for pGL3B + GATA3) → plotted value for pGL3B-431AP + GATA3

The quantitative PCRs were performed to investigate the endogenous gene expression of the transcription factors in the examined cell lines. The results of the differing endogenous expression can explain some of the mentioned dissimilarities among the cell lines.
All experiments were at least performed in triplicates.

3.2.1 Transfection of the cloned *ADGB* promoter/enhancer luciferase constructs

In HeLa cells, the *ADGB* promoter enhances the luciferase gene expression. The strongest increase is observed in presence of the pGL3B-1031AP vector (Figure 11A).
The supplementary addition of the *ADGB* enhancer region slightly increases the *ADGB* promoter-driven luciferase activity of the shortest, 431bp promoter fragment but results in a reduced luciferase activity in presence of the 1031bp and 1981 bp *ADGB* promoter (Figure 11A).
The *ADGB* enhancer cloned in the pGL3-Basic Vector has no observable effect (Figure 11A).

Transfection experiments performed in MCF7 cells confirm the observed results of the pGL3-Basic Vector backbone constructs. As observed in HeLa cells, the *ADGB* promoter acts as a promoting DNA element on gene expression (Figure 11B). In combination with the *ADGB* enhancer however, a depleting transcription of all promoter lengths is observed (Figure 11B). The *ADGB* enhancer itself has a promoting effect on the gene expression level (Figure 11B).

The results observed in MCF7 cells are confirmed in HEK293T cells (Figure 11C).

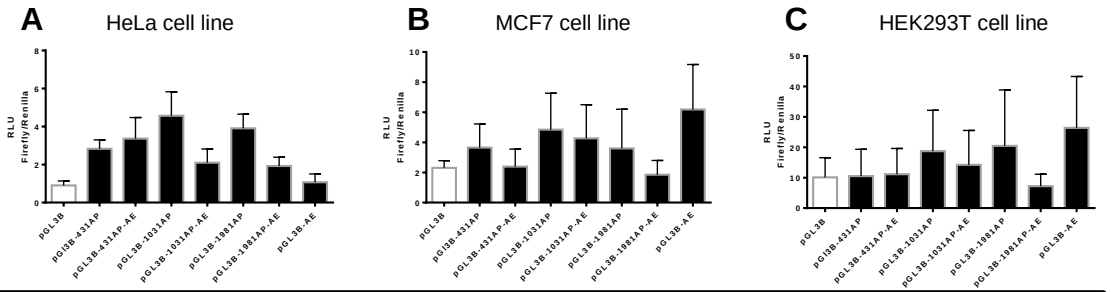

Fig. 11. Luciferase assays demonstrating the change of transcriptional activity through the *ADGB* promoter and/or enhancer in the pGL3Basic Vector. **(A):** Experiments performed in HeLa cells. **(B):** Experiments performed in MCF7 cells. **(C):** Experiments performed in HEK293T cells.

In comparison to the pGL3Basic Vector constructs, the introduction of the *ADGB* promoter to the viral, SV40 enhancer as present in the pGL3Enhancer Vector, results in profoundly increased luciferase gene transcription in the transfection experiments of all three cell lines (Figures 12A, B, C).

In presence of the 431bp and 1031bp *ADGB* promoter, substantially increased transcriptional activity is observed. In case of the largest 1981bp *ADGB* promoter however, a rather decreasing effect is seen (Figures 12A, B, C).

The observed effect is confirmed in MCF7 and HEK293T cell lines, but less pronounced (Figures 12B, C).

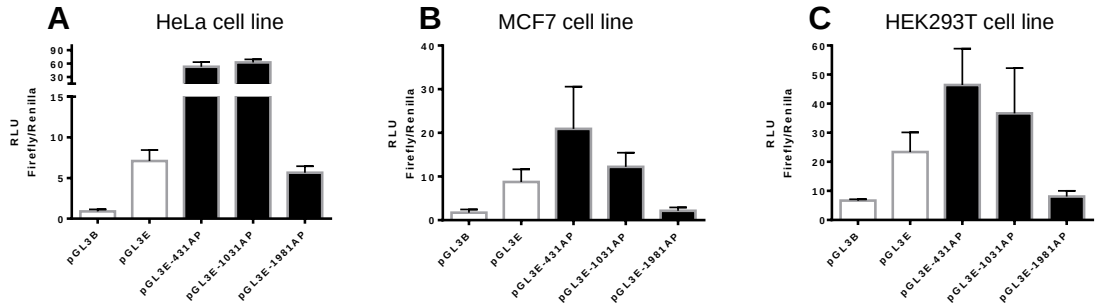

Fig. 12. Luciferase assays demonstrating the change of transcriptional activity through the *ADGB* promoter in the pGL3Enhancer Vector. **(A):** Experiments performed in HeLa cells. **(B):** Experiments performed in MCF7 cells. **(C):** Experiments performed in HEK293T cells.

The gene transcription of the viral, SV40 promoter as present in the pGL3Promoter Vector, is increased in presence of the *ADGB* enhancer, compared to without in HeLa cells (Figure 13A).

In MCF7 cells the effect is not as pronounced as in HeLa cells (Figure 13B). In HEK293T cells no significant change of transcription is distinguishable (Figures 13C).

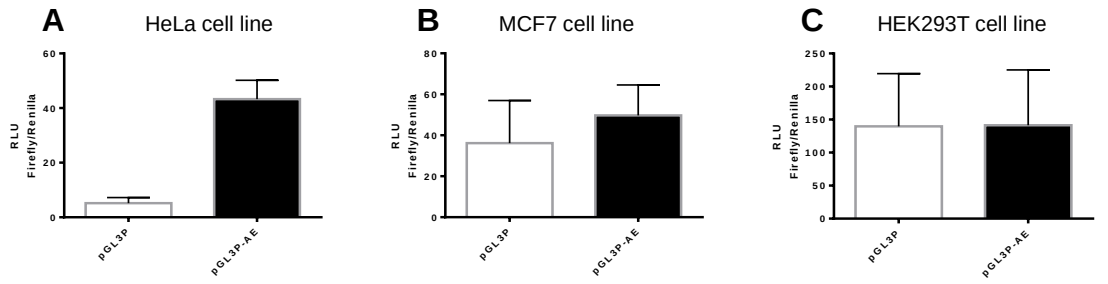

Fig. 13. Luciferase assays demonstrating the change of transcriptional activity through the *ADGB* enhancer in the pGL3Promoter Vector. **(A):** Experiments performed in HeLa cells. **(B):** Experiments performed in MCF7 cells. **(C):** Experiments performed in HEK293T cells.

3.2.2 Transcription factors with an enhancing transcriptional effect

The examined transcription factors Pax5, CEBPβ, REST, GATA2, GATA3 and TEAD4 display an enhancing effect on *ADGB* promoter-dependent luciferase gene transcription in HeLa cells.

The co-overexpression of Pax5 in presence of the *ADGB* promoter and enhancer leads to an approximate 1.8-fold increase of gene transcription. No additional impact is detectable upon addition of the *ADGB* enhancer (Figure 14A).

Pax5 demonstrates furthermore no significant transcriptional impact on the SV40 enhancer or promoter, in combination with the *ADGB* gene regulatory regions containing vector constructs (Figures 14B, C). A decreasing tendency is detectable in both cases (Figures 14B, C).

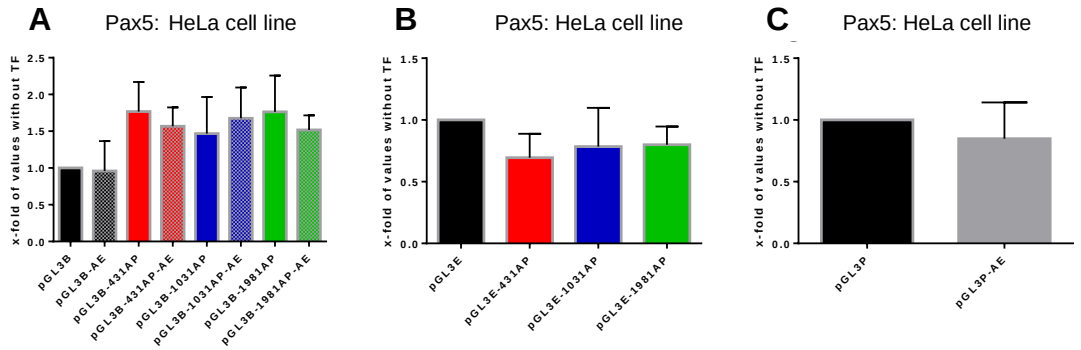

Fig. 14. Luciferase assays demonstrating the change of transcriptional activity through Pax5 overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

A dose-dependent transfection of Pax5 concerning the *ADGB* promoter-driven luciferase was evaluated. A dose-dependent increase of transcriptional activity up to 300ng TF / well is detectable. An increase to 400ng TF / well leads to a decreasing transcription (Figure 15).

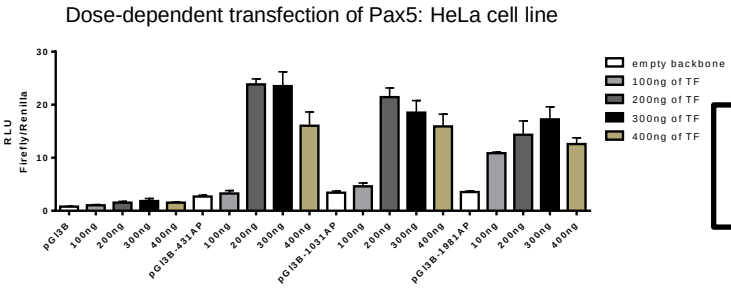

Fig. 15. Luciferase assays demonstrating the dose-dependent change of transcriptional activity through Pax5 overexpression in HeLa cells.

In regards of the pGL3Basic and pGL3Promoter Vector constructs, the experiments performed in MCF7 and HEK293T cells exhibit a similar effect as observed in HeLa (Figures 16A, C, 17A, C). In presence of the *ADGB* promoter and the SV40 enhancer however, a contrary effect compared to HeLa cells is observed. The transcription is enhanced in both cell lines (Figures 16B, 17B).

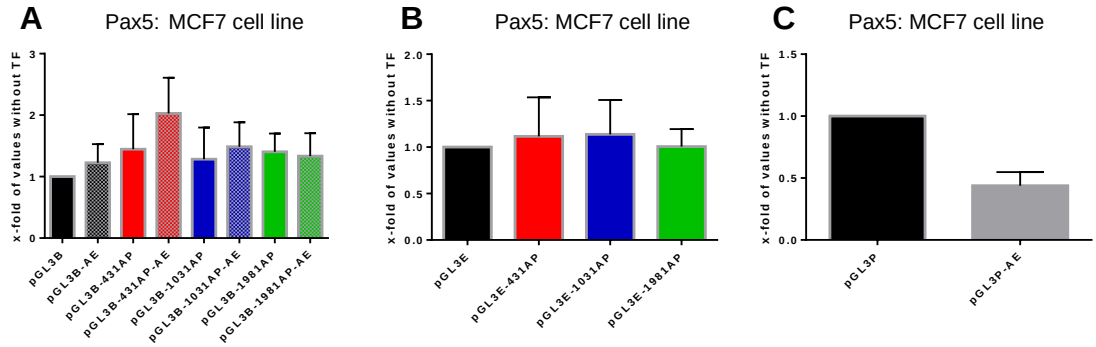

Fig. 16. Luciferase assays demonstrating the change of transcriptional activity through Pax5 overexpression in MCF7 cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

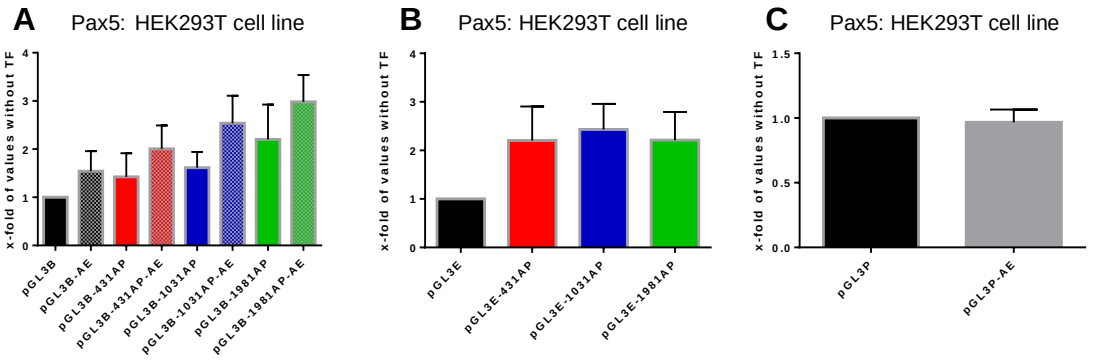

Fig. 17. Luciferase assays demonstrating the change of transcriptional activity through Pax5 overexpression in HEK293T cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The corresponding qPCR demonstrates similar Pax5 expression levels in HeLa and HEK293T cells and a more than 2-fold higher expression in MCF7 cells (Figure 18). This could explain the less pronounced results of the transfection experiments in MCF7 cells (Figure 18).

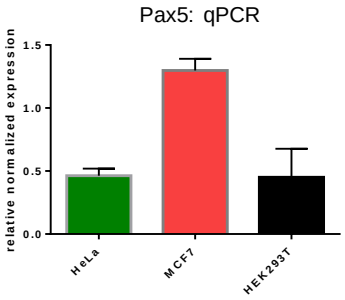

Fig. 18. qPCR demonstrating the endogenous Pax5 expression in HeLa, MCF7 and HEK293T cells.

CEBPβ increases the gene transcription more than 3-fold in presence of the *ADGB* promoter (Figure 19A). In presence of the *ADGB* enhancer a comparably decreasing effect is noticeable, also seen in the transfection of the pGL3B-AE vector construct (Figure 19A).

In presence of the SV40 enhancer and CEBPβ, the transfections containing the *ADGB* promoter demonstrate an increasing effect for the 431bp *ADGB* promoter and a decreasing effect for the longer promoter fragments (Figure 19B). No significant change of transcription can be observed in presence of the SV40 promoter and the *ADGB* enhancer (Figure 19C).

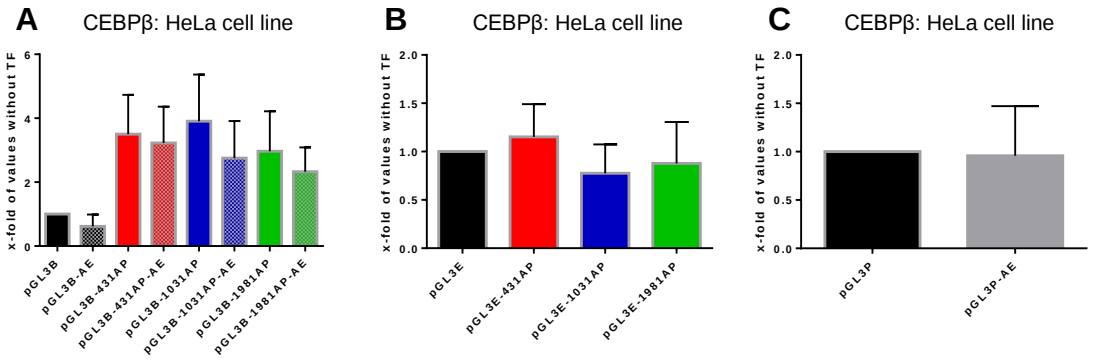

Fig. 19. Luciferase assays demonstrating the change of transcriptional activity through CEBPβ overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The effect of CEBPβ on *ADGB* promoter and enhancer dependent gene expression is neither reproducible in MCF7, nor in HEK293T cells. In both cell lines no significant change of gene expression level is detectable.

In presence of REST and all three promoter lengths of the *ADGB* promoter the transcription is increased to an approximate four-fold (Figure 20A). A higher transcription rate is additionally detectable in combination with the *ADGB* enhancer (Figure 20A).

The constructs with the pGL3Enhancer Vector backbone present an increasing gene transcription, up to 2-fold in case of the 1981bp *ADGB* promoter (Figure 20B).

No transcription factor dependent change of gene expression is observed in the construct of the SV40 promoter vector, with or without the *ADGB* enhancer (Figure 20C).

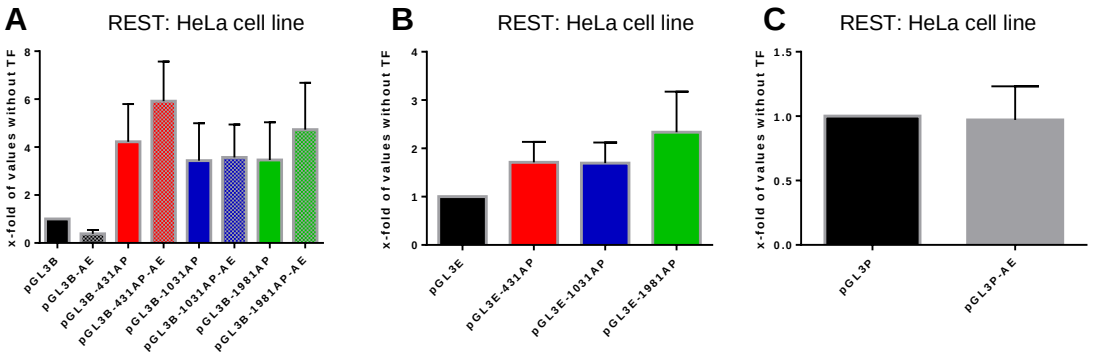

Fig. 20. Luciferase assays demonstrating the change of transcriptional activity through REST overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

In MCF7 cells the increasing effect on gene transcription of REST in presence of *ADGB* is primarily seen in the constructs containing the *ADGB* promoter only (Figure 21A). In comparison with the transfections containing the *ADGB* enhancer as well, smaller gene expression levels are detectable (Figure 21A).

In case of the transfections that include *ADGB* gene sections plus the viral promoter or enhancer, the observed results in HeLa cells can be confirmed in MCF7 cells (Figures 21B, C).

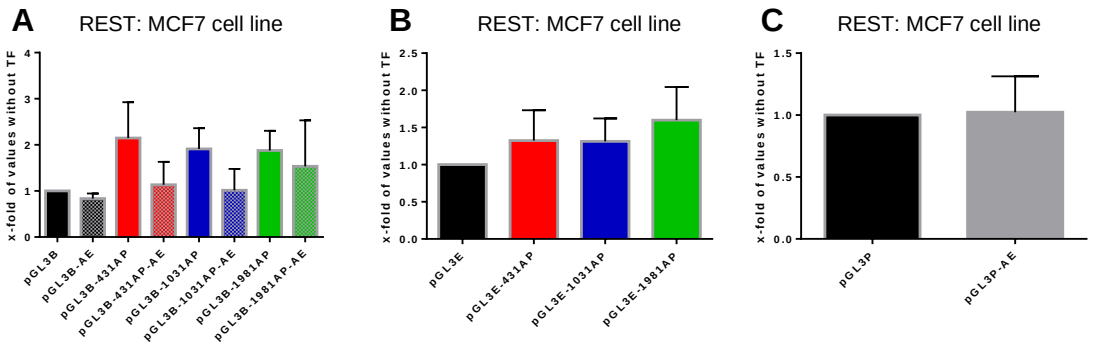

Fig. 21. Luciferase assays demonstrating the change of transcriptional activity through REST overexpression in MCF7 cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The same transcriptional effect, as observed in MCF7 cells is confirmed in HEK293T cells (Figures 22A, B, C). Hence, the detected effect is more pronounced than in MCF7 cells.

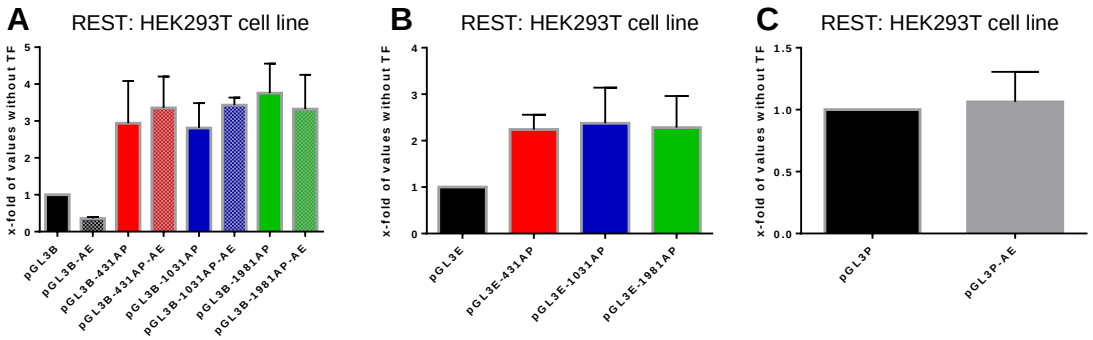

Fig. 22. Luciferase assays demonstrating the change of transcriptional activity through REST overexpression in HEK293T cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The relative normalized endogenous expression of REST demonstrates the highest expression in MCF7 and the lowest in HEK293T cells (Figure 23). This observation underlines the more pronounced effect in HEK293T cells.

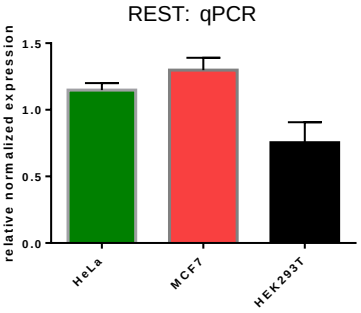

Fig. 23. qPCR demonstrating the endogenous Pax5 expression in HeLa, MCF7 and HEK293T cells.

The transcription factor GATA2 enhances the corresponding gene transcription in the experiments containing the *ADGB* promoter and enhancer in the pGL3Basic Vector constructs (Figure 24A). The results display the highest expression level in the vector construct with the shortest *ADGB* promoter fragment and the *ADGB* enhancer (Figure 24A). In case of the longer *ADGB* promoter fragments a decreasing effect of the *ADGB* enhancer is detectable (Figure 24A).

The vector constructs containing the androglobin gene fragments and the SV40 promoter or enhancer show an increase of gene transcription in presence of GATA2 (Figures 24B, C).

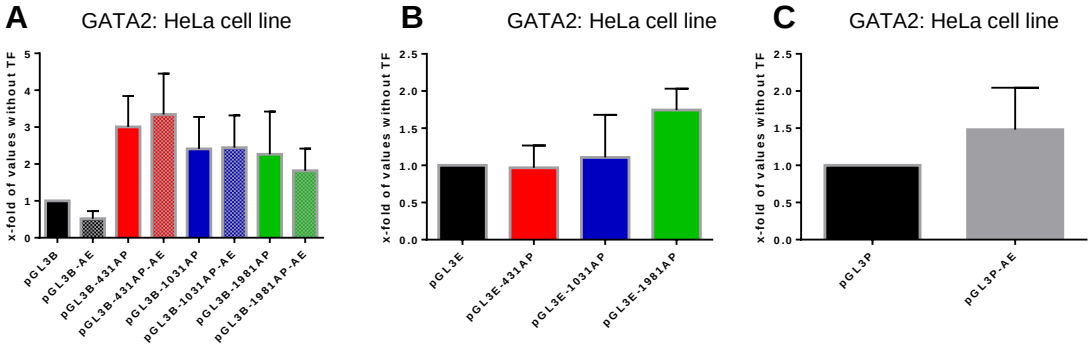

Fig. 24. Luciferase assays demonstrating the change of transcriptional activity through GATA2 overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The effect of GATA2 on *ADGB* promoter and enhancer dependent gene expression is neither reproducible in MCF7 cells nor in HEK293T cells. In both cell lines no significant change of gene expression level is detectable.

GATA3 displays an increasing impact on gene transcription, which is moderated by the *ADGB* promoter/enhancer (Figure 25A). The highest activity, up to a 7-fold increase is observed in case of the shortest *ADGB* promoter fragment (Figure 25A). The difference

between the transcriptional activity with or without the *ADGB* enhancer is comparable to the obtained results in presence of other transcription factors. An increase in presence of the 431bp *ADGB* promoter and a decrease in case of the longest 1981bp *ADGB* promoter (Figure 25A).

Given the vector construct of the *ADGB* promoter with the SV40 enhancer, a similar cluster of gene transcription alteration is observed as in presence of the *ADGB* enhancer (Figure 25B).

GATA3 enhances the gene transcription in presence of the viral promoter and the *ADGB* enhancer (Figure 25C).

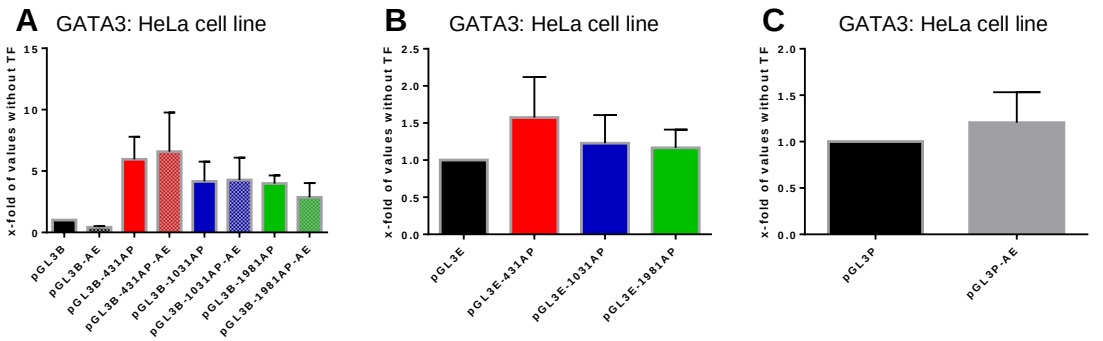

Fig. 25. Luciferase assays demonstrating the change of transcriptional activity through GATA3 overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The effect of GATA3 on *ADGB* promoter and enhancer dependent gene expression is neither reproducible in MCF7 cells nor in HEK293T cells. In both cell lines no significant change of gene expression level is detectable.

In regards of TEAD4, an overall enhancing effect is measured (Figures 26A, B, C).

The effect of TEAD4 on gene transcription in the pGL3Basic Vector constructs is pronounced in the experiments containing the *ADGB* promoter and enhancer, compared to without the *ADGB* enhancer (Figure 26A). The highest transcription is again observed in case of the pGL3B-431AP-AE vector construct (Figure 26A).

The transcriptional effect of TEAD4 on the *ADGB* promoter and the SV40 enhancer is increased up to an approximate 2.5-fold enhancement (Figure 26B).

TEAD4 yields to a more than 1.5-fold enhancement in combination of the viral promoter and the *ADGB* enhancer, compared to without the *ADGB* enhancer (Figure 26C).

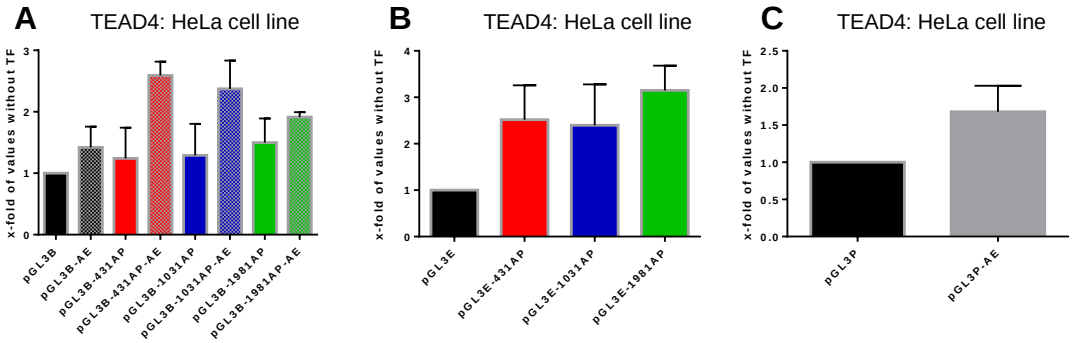

Fig. 26. Luciferase assays demonstrating the change of transcriptional activity through TEAD4 overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The effect of TEAD4 on *ADGB* promoter and enhancer dependent gene expression is neither reproducible in MCF7 cells nor in HEK293T cells. In both cell lines no significant change of gene expression level is detectable.

3.2.3 Transcription factor with a repressing transcriptional effect

The transfection experiments performed in presence of HP1Gamma and PU.1 display a repressing effect on the *ADGB* regulatory gene regions dependent luciferase transcription. Whereas the effect of HP1Gamma is supposedly dependent on the *ADGB* promoter, PU.1 seems to exert its repression through the *ADGB* enhancer (Figures 27A, 31A).

In case of the overexpression of HP1Gamma, the gene expression is more repressed in presence of the *ADGB* promoter only, compared to the transfection results including the *ADGB* enhancer (Figure 27A).

The observed effects can partly be seen in combination with viral gene regulatory regions as well. Whereas the effect of HP1Gamma on the *ADGB* promoter dependent gene expression is also diminishing in presence of the SV40 enhancer, the combination of the viral promoter and the *ADGB* enhancer leads to an increase in gene expression, compared to without the enhancer (Figures 27B, C).

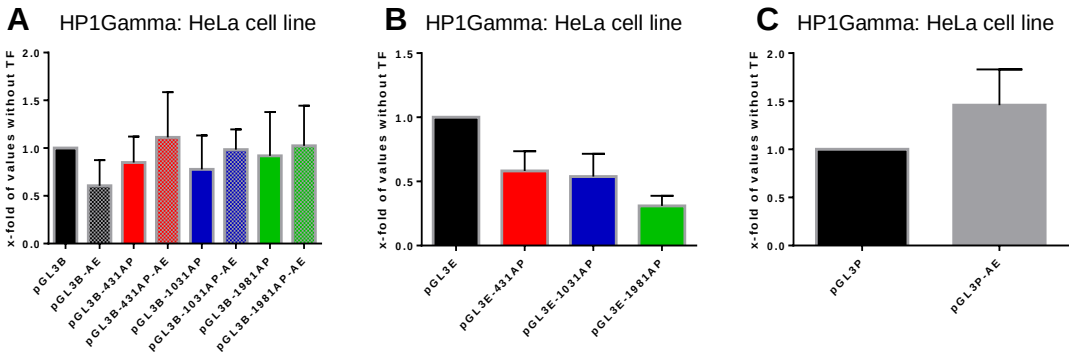

Fig. 27. Luciferase assays demonstrating the change of transcriptional activity through HP1Gamma overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The transcriptional effects of HP1Gamma are partially reproducible in MCF7 cells. As observed in HeLa cells, the *ADGB* promoter dependent transcription is also diminished in presence of HP1Gamma (Figures 28A, B). The presence of the *ADGB* promoter in combination with the enhancer, in co-overexpression with HP1Gamma induces the gene transcription, except for the longest 1981bp *ADGB* promoter (Figure 28A). The SV40 promoter and *ADGB* enhancer however expose a diminished gene transcription (Figure 28C).

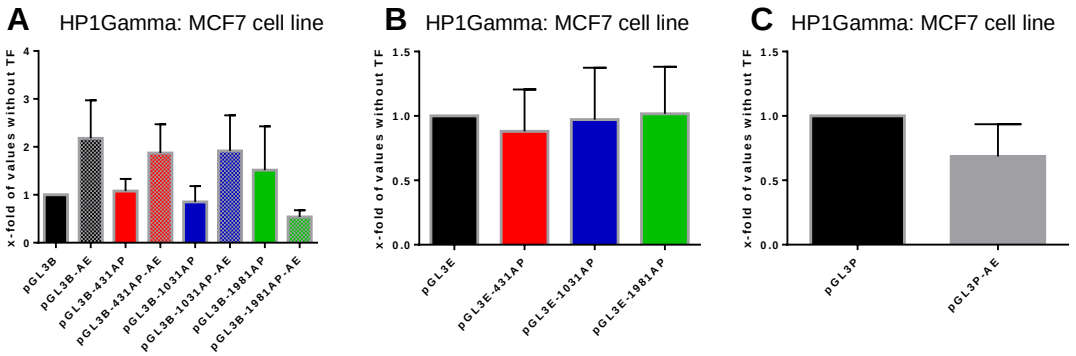

Fig. 28. Luciferase assays demonstrating the change of transcriptional activity through HP1Gamma overexpression in MCF7 cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

In HEK293T cells a HP1Gamma dependent transcriptional decrease via the interaction with the *ADGB* promoter is also observable (Figures 29A, B). Contrary to the observations in HeLa cells, the addition of the *ADGB* enhancer exposes an even more pronounced decline of transcription (Figure 29A).

The observed results in HeLa cells concerning the interaction of the SV40 promoter and the *ADGB* enhancer can be confirmed in HEK293T cells (Figure 29C).

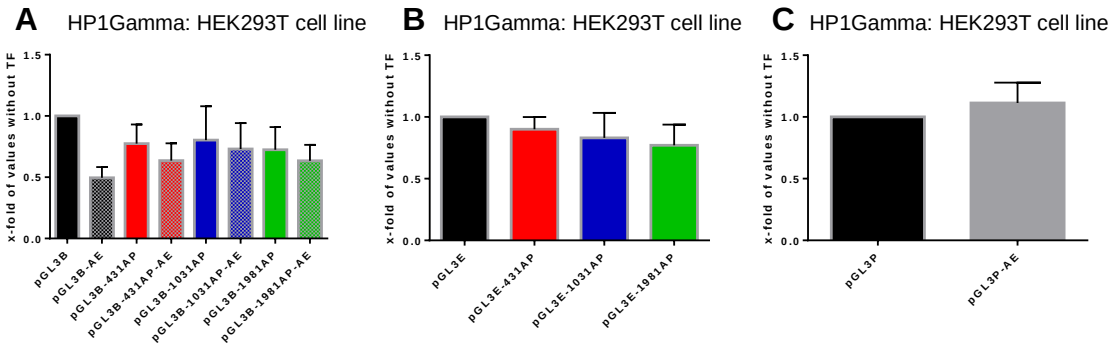

Fig. 29. Luciferase assays demonstrating the change of transcriptional activity through HP1Gamma overexpression in HEK293T cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The highest endogenous expression is observed in HeLa cells, followed by MCF7 and HEK293T cells (Figure 30). The relatively high endogenous expression of HP1Gamma in MCF7 might be the reason for the demonstrated impaired repression.

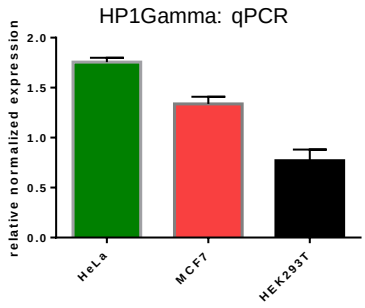

Fig. 30. qPCR demonstrating the endogenous HP1Gamma expression in HeLa, MCF7 and HEK293T cells.

PU.1 seems to exert a repressing function by binding to the *ADGB* enhancer (Figure 31A). Except for the transfections containing pGL3B-431AP-AE, the transcriptional activity is diminished in presence of the *ADGB* promoter and enhancer, compared to without the enhancer (Figure 31A). The most pronounced repression is demonstrated on the vector constructs containing the 1031bp *ADGB* promoter and the *ADGB* enhancer (Figure 31A). The impact of PU.1 on the vector constructs containing *ADGB* regulatory gene regions, in combination with the SV40 promoter or enhancer is not pronounced (Figures 31B, C). A transcriptional enhancing effect can be anticipated in both cases nonetheless (Figures 31B, C).

This finding displays that, in HeLa cells the transcriptional repression of PU.1 seems to be exclusive for the interaction of the *ADGB* promoter and enhancer.

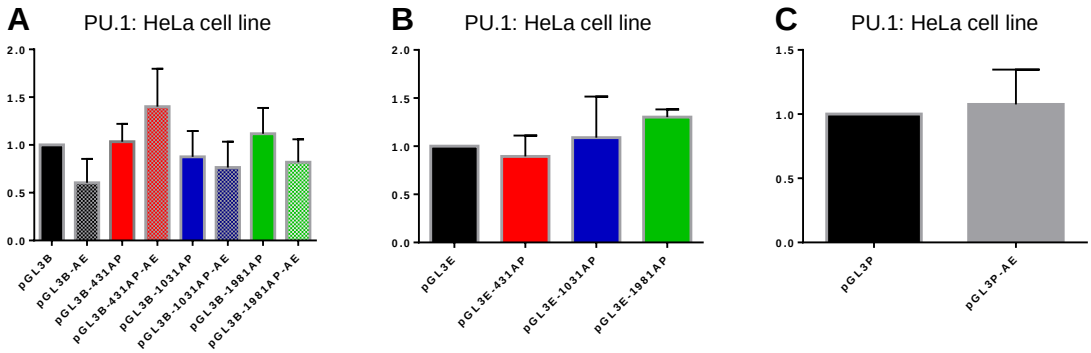

Fig. 31. Luciferase assays demonstrating the change of transcriptional activity through PU.1 overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

As observed in HeLa cells, PU.1 also exerts a repressing effect on *ADGB* promoter/enhancer dependent gene transcription in MCF7 and HEK293T cells (Figures 32A, 33A). Except for pGL3B-1981AP-AE in MCF7 cells, the presence of the *ADGB* enhancer also diminishes the gene transcription, compared to the transfections without (Figures 32A, 33A).

Concerning the transcriptional alteration of PU.1 in co-overexpression with the *ADGB* promoter/enhancer + SV40 promoter and enhancer, the same effect as in HeLa cells is observed in MCF7 cells (Figures 32B, C).

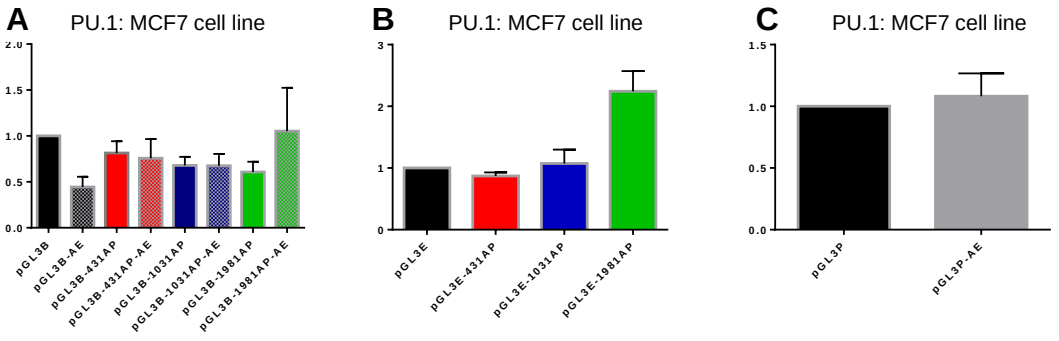

Fig. 32. Luciferase assays demonstrating the change of transcriptional activity through PU.1 overexpression in MCF7 cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

A slight decrease of gene transcription is noticeable in presence of the *ADGB* promoter and the SV40 enhancer in HEK293T cells (Figure 33B).

The *ADGB* enhancer cloned to the viral promoter does not seem to interact with the transcription factor in HEK293T cells (Figure 33C).

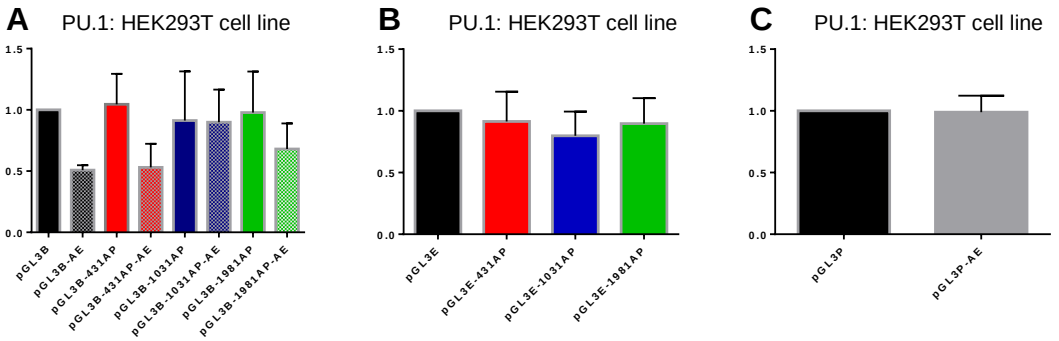

Fig. 33. Luciferase assays demonstrating the change of transcriptional activity through PU.1 overexpression in HEK293T cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The endogenous expression of PU.1 is most pronounced in HeLa cells (Figure 34). MCF7 and HEK293T display an approximate 50 percent lower expression (Figure 34).

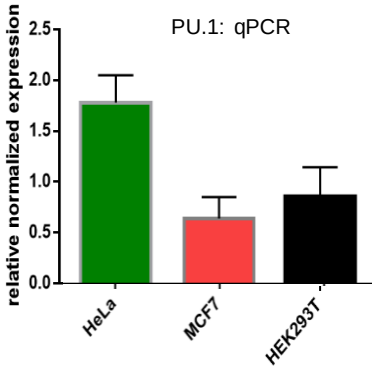

Fig. 34. qPCR demonstrating the endogenous PU.1 expression in HeLa, MCF7 and HEK293T cells.

3.2.4 Transcription factors with a pronounced transcriptional effect on the *ADGB* promoter + SV40 enhancer

The *ADGB* promoter, with or without the *ADGB* enhancer cloned to the pGL3Basic Vector backbone does not seem to alter the gene expression in presence of the transcription factors Rfx2 and Nrf3. Co-overexpressed with the *ADGB* promoter and the SV40 enhancer, an alteration is however displayed. Rfx2 enhances the corresponding gene transcription and Nrf3 reduces it.

In case of Rfx2, an increase of gene transcription can be observed for all three promoter lengths in all three cell lines (Figures 35A, B, C). The most pronounced effect is seen in presence of the longest, 1981bp *ADGB* promoter (Figures 35A, B, C).
The effect is most pronounced in HeLa cells (Figure 35A).

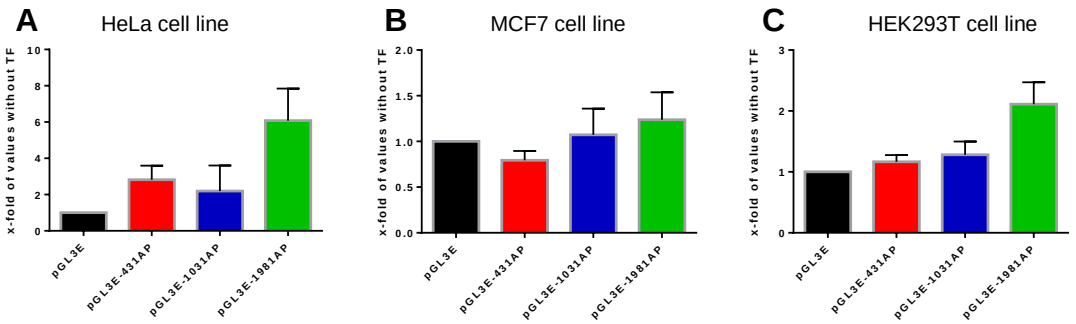

Fig. 35. Luciferase assays demonstrating the change of transcriptional activity through the *ADGB* promoter and the SV40 enhancer in presence of Rfx2. **(A):** Experiments performed in HeLa cells. **(B):** Experiments performed in MCF7 cells. **(C):** Experiments performed in HEK293T cells.

The endogenous expression of Rfx2 in HeLa and HEK293T cells is approximately the same (Figure 36). In MCF7 cells however, the corresponding expression is about 3-fold smaller (Figure 36).

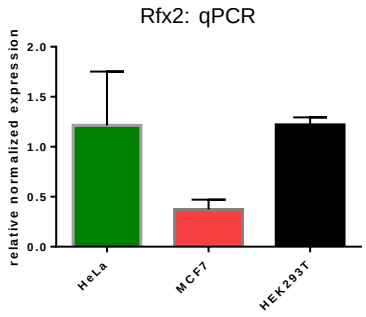

Fig. 36. qPCR demonstrating the endogenous RFX2 expression in HeLa, MCF7 and HEK293T cells.

The transcriptional repression resulting from the binding of Nrf3 is pronounced in the setup containing the *ADGB* promoter and the SV40 enhancer in HeLa cells (Figure 37B). In this case, the base pair lengths correlate with the repressive impact (Figure 37B). As observed in the other transfections, the transcriptional repression seems to be dependent on the interaction with the *ADGB* promoter and an enhancer other than the *ADGB* enhancer (Figures 37A, C).

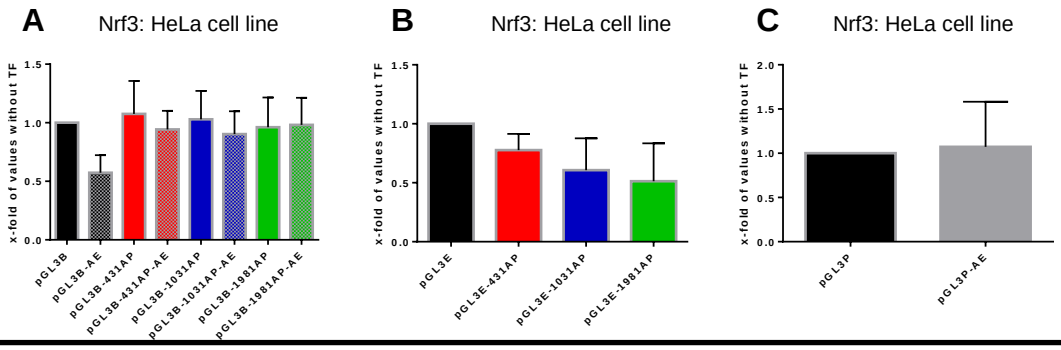

Fig. 37. Luciferase assays demonstrating the change of transcriptional activity through PU.1 overexpression in HeLa cells. **(A):** Experiments performed on pGL3B-x vector backbone. **(B):** Experiments performed on pGL3E-x vector backbone. **(C):** Experiments performed on pGL3P-x vector backbone.

The effect of Nrf3 on *ADGB* promoter and enhancer dependent gene expression is neither reproducible in MCF7, nor in HEK293T cells. In both cell lines no significant change of gene expression level is detectable.

3.2.5 Transcription factors with insignificant effects on gene expression levels

The transcription factors GABPα, USF1, Nrf2, PML-IV, Trim28 and P54 were furthermore examined but did not exhibit a distinguishable effect on the *ADGB* promoter or enhancer dependent gene transcription in the transfection experiments.

Whereas stimulating trends are suggested prior to transfection of Nrf2, PML-IV and Trim28, the overexpression of GABPα, P54 and USF1 display decreasing potentials. On account of the performed transfections only tendencies about the actual effect of the transcription factors are noticeable. For more pronounced results an adaption of the experimental settings and further research needs to be completed.

3.3 Chromatin immunoprecipitation

GATA3 and Pax5 were chosen as representatives of the examined transcription factors to investigate the corresponding binding sequences on behalf of the *ADGB* promoter and enhancer. The transcription factors were chosen due to the promising results in the transfection experiments to alter corresponding gene expression via binding to the regulatory *ADGB* gene sequences.

In the experimental setup, DNA samples of the cell lines MCF7 and Hep3B were used as templates to analyze the binding of the transcription factors. *In silico* analysis employing

the R2 Genomics Analysis and Visualization Platform (http://r2.amc.nl) revealed high endogenous expression levels in those cell lines. The performed qPCRs could experimentally validate these assumptions.
All ChIP experiments were at least performed in triplicates.

The DNA of the employed cells was sonicated to reach a desired average DNA fragment length of 500bp. The successful fragmentation was verified using a 1.8% agarose gel (Figures 38, 39).

In case of MCF7 cell DNA, 45 repetitions of the described sonication cycle were needed to obtain the desired size (Figure 38).

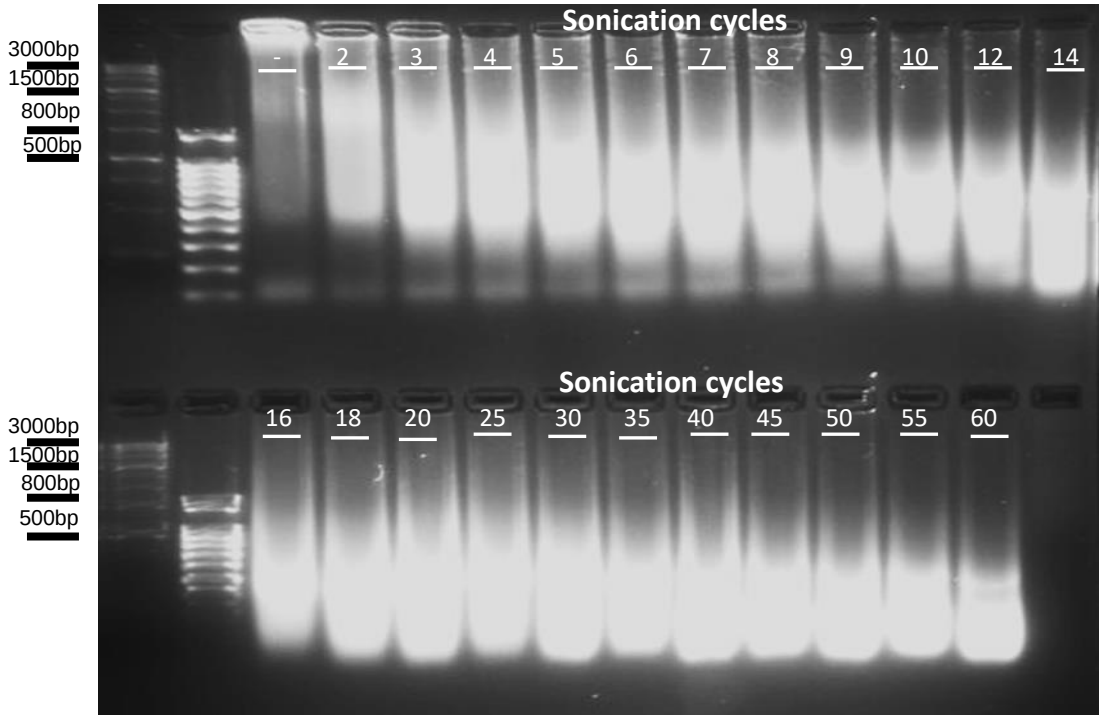

Fig. 38. 1.8% agarose gel presenting the sheared **MCF7 cell** DNA after different numbers of sonication cycles. 250µl DNA sample was subjected to hydrodynamic shearing. One cycle consists of 30 seconds sonication (100% machine efficiency), followed by 60 seconds cool down on ice.
Columns from **upper** part, left to right: (1) 1Kbp DNA ladder; (2) 100bp DNA ladder; (3) Control (unsonicated DNA); (4 - 14) DNA samples of increasing repetitions of sonication cycles.
Columns from **lower** part, left to right: (1) 1Kbp DNA ladder; (2) 100bp DNA ladder; (3 - 14) DNA samples of increasing repetitions of sonication cycles.

13 cycles of the described procedure for Hep3B cell DNA were needed to obtain the requested size (Figure 39).

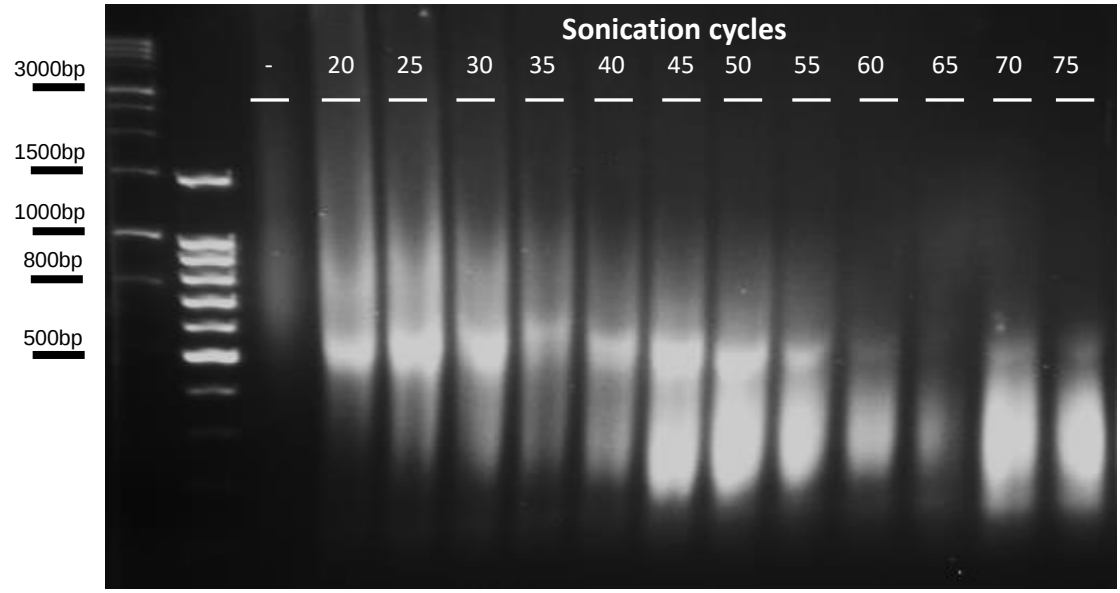

Fig. 39. 1.8% agarose gel presenting the sheared **Hep3B cell** DNA after different numbers of sonication cycles. 250µl DNA sample was subjected to hydrodynamic shearing. On cycle consists of 15 seconds sonication (100% machine efficiency), followed by 30 seconds cool down on ice.
Columns from left to right: (1) 1Kbp DNA ladder; (2) 100bp DNA ladder; (3) Control (unsonicated DNA); (4 - 14) DNA samples of increasing repetitions of sonication cycles.

The sonicated DNA was immunoprecipitated with specific antibodies against GATA3 and Pax5. An unspecific antibody against actin was used as an internal, negative control of the antibody binding. As an antibody, negative control ddH_2O instead of an antibody was included in the immunoprecipitation process.

The obtained DNA was used as a template in PCRs to amplify the specific transcription factor binding regions in the *ADGB* promoter or enhancer sequences. For the PCR amplification of the input, positive control the sonicated DNA, before immunoprecipitation was used as a template.

The DNA amplification was investigated on a 1.8% agarose gel.

The experiments illustrate the GATA3 antibody binding to the predicted transcription factor binding sites 1, 2, 5 and 6 in the *ADGB* promoter, using DNA of MCF7 cells (Figure 40A). The binding of the GATA3 antibody to the genomic regions 2 and 6 is verified using the DNA of Hep3B-cells (Figure 40B).

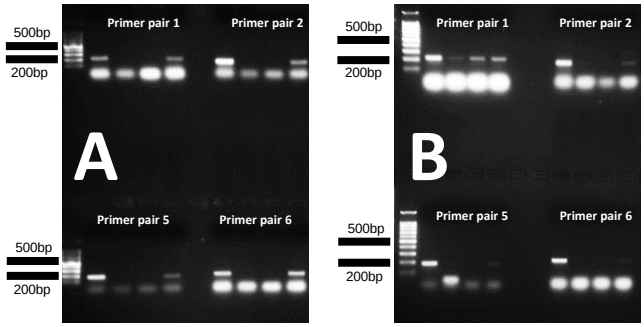

Fig. 40. 1.8% agarose gels presenting ChIP assays to evaluate the binding capacity of anti-**GATA3** to putative TF-binding sites in the ***ADGB* promoter** region.
Four different primer pairs (Primer pair 1, 2, 5 and 6) were used to amplify the TF-binding regions. **(A):** MCF-7 cell DNA. **(B):** Hep3B cell DNA.
Columns from left to right **(2 – 5 / examined primer pair):** (1) 100bp DNA ladder; (2) Input control; (3) Antibody control; (4) Internal control; (5) Amplified DNA with 10µl anti-GATA3.

The binding sites 3, 4 and 6 were additionally established as proof to GATA3 antibody binding in the *ADGB* enhancer, using MCF7 cell DNA and the binding sites 3 and 6 using Hep3B-cell DNA (Figures 41A, B).

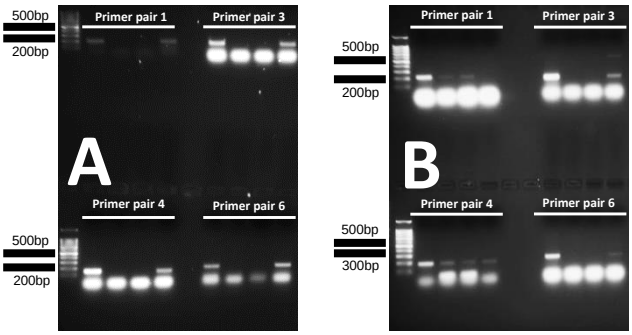

Fig. 41. 1.8% agarose gels presenting ChIP assays to evaluate the binding capacity of anti-**GATA3** to putative TF-binding sites in the ***ADGB* enhancer** region.
Four different primer pairs (Primer pair 1, 3, 4 and 6) were used to amplify the TF-binding regions. **(A):** MCF-7 cell DNA. **(B):** Hep3B cell DNA.
Columns from left to right **(2 – 5 / examined primer pair):** (1) 100bp DNA ladder; (2) Input control; (3) Antibody control; (4) Internal control; (5) Amplified DNA with 10µl anti-GATA3.

In the *ADGB* promoter, binding site 4 was identified in both examined DNA samples to bind Pax 5 antibody (Figures 42A, B).

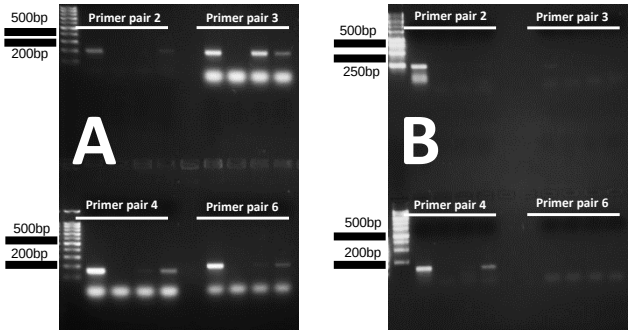

Fig. 42. 1.8% agarose gels presenting ChIP assays to evaluate the binding capacity of anti-**Pax5** to putative TF-binding sites in the ***ADGB* promoter** region.
Four different primer pairs (Primer pair 2, 3, 4 and 6) were used to amplify the TF-binding regions. **(A):** MCF-7 cell DNA. **(B):** Hep3B cell DNA.
Columns from left to right **(2 – 5 / examined primer pair):** (1) 100bp DNA ladder; (2) Input control; (3) Antibody control; (4) Internal control; (5) Amplified DNA with 10µl anti-Pax5.

Concerning the *ADGB* enhancer, the DNA region 3 in the analyzed MCF7 cell DNA and binding site 2 and 4 in Hep3B cell DNA were established to bind Pax5 antibody (Figures 43A, B).

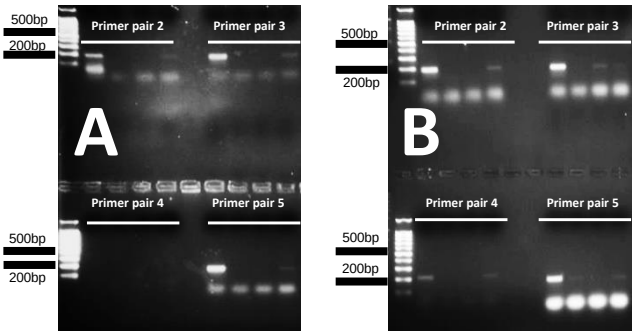

Fig. 43. 1.8% agarose gels presenting ChIP assays to evaluate the binding capacity of anti-**Pax5** to putative TF-binding sites in the ***ADGB* enhancer** region.
Four different primer pairs (Primer pair 1, 3, 4 and 6) were used to amplify the TF-binding regions. **(A):** MCF-7 cell DNA. **(B):** Hep3B cell DNA.
Columns from left to right **(2 – 5 / examined primer pair):** (1) 100bp DNA ladder; (2) Input control; (3) Antibody control; (4) Internal control; (5) Amplified DNA with 10µl anti-Pax5.

3.4 Site directed mutagenesis

The GATA3 binding site in the 1981bp *ADGB* promoter was chosen by in silico research and confirmed by the performed ChIP experiments. The successful mutation in two vector constructs of pGL3B-1981AP was verified by external sequencing (Figure 44).

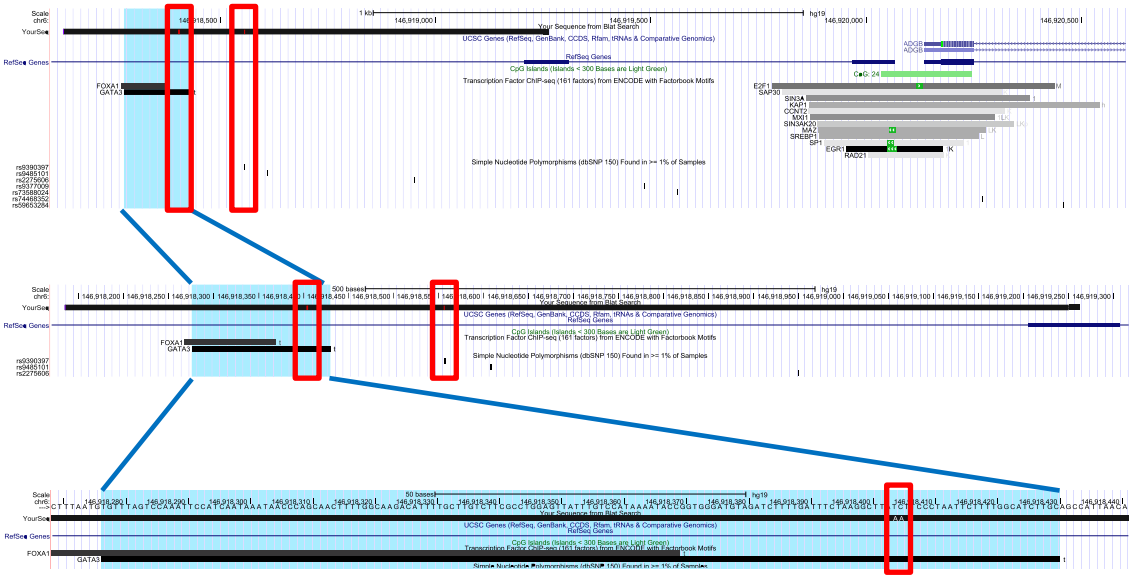

Fig. 44. The 1981bp *ADGB* promoter sequence aligned to the 1981bp *ADGB* promoter with the mutated GATA3 transcription factor binding site, as displayed by the UCSC Genome Browser (http://genome.ucsc.edu/).
The black horizontal bar, labeled as "Your Sequence from BLAT search" shows the sequence of the mutated 1981bp *ADGB* promoter. The red vertical bars in between symbolize deviations from the original genomic sequence. Overall, three mismatches are detected: two deviations belong to the desired mutation (TC → AA), overlapping with a GATA3 binding site, the third belongs to simple nucleotide polymorphism (SNP).
The mismatch at the end of the sequence is due to imprecision of the sequencing method after 1000 bp.

The two mutated vector constructs were further transfected, in comparison to the *ADGB* promoter and the pGL3B-1981AP-AE vector in HeLa, MCF7 and HEK293T cells. Mutation 1 displays similar gene expression levels in HeLa cells, compared to the unmutated vector containing the 1981bp *ADGB* promoter (Figure 45A). The expression of Mutation 2 is even elevated (Figure 45A).

In the other cell lines, the mutated vector constructs demonstrate increased gene transcription levels, compared to the unmutated version (Figures 45B, C).

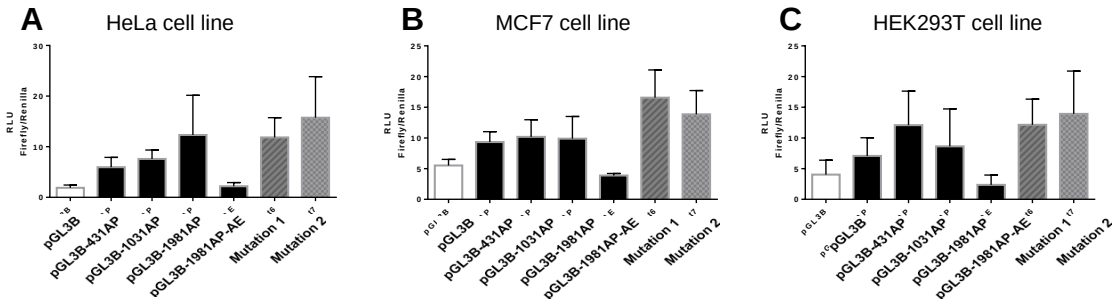

Fig. 45. Luciferase assays demonstrating the change of transcriptional activity through the *ADGB* promoter and/or enhancer in the pGL3Basic Vector, in comparison to the mutated pGL3B-1981AP vector in Mutation 1/2. **(A):** Experiments performed in HeLa cells. **(B):** Experiments performed in MCF7 cells. **(C):** Experiments performed in HEK293T cells.

The supplementation of 300ng GATA3 exhibits higher gene expression levels in all *ADGB* promoter containing vector constructs, in all three cell lines (Figures 46A, B, C). In presence of the transcription factor, the vector constructs containing the mutation of the GATA3 binding site in the 1981bp *ADGB* promoter show similar and even higher gene expression levels, compared to the vectors containing the unmutated 1981bp *ADGB* promoter (Figures 46A, B, C).

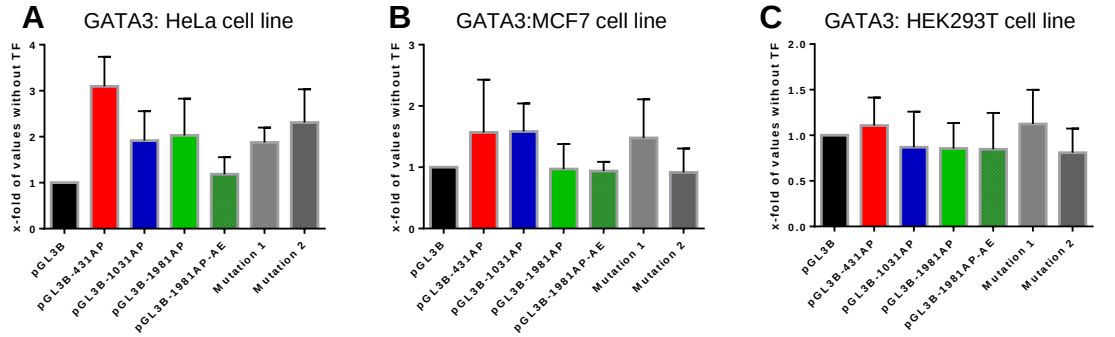

Fig. 46. Luciferase assays demonstrating the change of transcriptional activity through GATA3 overexpression on the unmutated and mutated *ADGB* promoter. **(A):** Experiments performed in HeLa cells. **(B):** Experiments performed in MCF7 cells. **(C):** Experiments performed in HEK293T cells.

4 DISCUSSION

4.1 Expression regulation of androglobin

As outlined in the introduction, a variety of essential functions could be linked to the protein family of globins in humans. Due to the relatively short scientific history androglobin has, the attributed functions still need to be specified.
Androglobin was first described in 2012 and termed as such because of its heme-incorporating globin properties and mostly abundant expression in testis (Hoogewijs et al., 2012b). The regulation of the gene has not been reported thus far.

The obtained results of this research display distinct regulation patterns of the *ADGB* promoter and enhancer dependent gene expression in presence of transcription factors, thus suggesting a close interaction. A comparison of the influence of the transcription factors on the alteration of *ADGB* promoter/enhancer dependent gene expression levels and the background study of these transcription factors enables us to draw additional conclusions about new potential functions, attributions and fields of androglobin expression.

Three different sizes of the *ADGB* promoter are examined in this thesis. The highest expression in the observed cell lines is detected in the overexpression of pGL3B-1031AP, suggesting the purest androglobin promoting element, or rather the gene region with the highest transcription enhancing probability in between 431 and 1031 base pairs. An element that needs a cooperative action with an element between 431 and 1031 could lay as an alternative explanation within the first 431bp. Furthermore, a potential gene silencing element can be anticipated in between base pairs 1031 and 1981 of the *ADGB* promoter, due to a diminishing gene transcription of the 1981bp *ADGB* promoter compared to the 1031bp *ADGB* promoter. This observation is also confirmed by the transfection experiments containing the viral SV40 enhancer.

The *ADGB* enhancer however has a decreasing effect on the gene expression level, suggesting a repressing function of this intronic gene region. Whereas the shortest promoter fragment combined with the *ADGB* enhancer demonstrates similar gene

expression patterns as without the enhancer, the reducing effect is highlighted in presence of the 1031bp and 1981bp *ADGB* promoter fragments.
This data suggests a repressive function of the enhancer in presence of the *ADGB* promoter, as observed in between gene regulatory regions via tracking, linking, relocating or looping (Kolovos et al., 2012). The direct interaction between the *ADGB* promoter and enhancer gene region could further be examined via long-range interaction tools such as ChIA-PET, 3C, 5C or Hi-C.

The observed effects of the *ADGB* promoter on gene expression are underlined in presence of the viral SV40 enhancer of the pGL3-Enhancer Vector. The expression pattern of enhancing the corresponding gene expression in between 431bp and 1031bp and reducing it in between 1031bp and 1981bp is even more pronounced in this case.
The *ADGB* enhancer, in presence of the SV40 promoter of the pGL3-Promoter Vector however displays an enhancing effect in HeLa and MCF7 cells and thus not the expected silencing feature. This observation exhibits a potential exclusive interplay between the *ADGB* promoter and enhancer.

In several transfection experiments the *ADGB* enhancer displays a relatively higher transcriptional activity, compared to the reference vector without the enhancer gene region, when cloned to the pGL3Basic Vector. This suggests a partial promoting transcriptional effect of the enhancer itself, as it has been observed for other enhancer elements as well (Kim and Shiekhattar, 2015, Andersson et al., 2015).

Some of the displayed transfection experiments, performed in the examined cell lines HeLa, MCF7 and HEK293T demonstrate similar gene expression patterns. In other cell lines they differ. The dissimilarity can be explained by the individual presence of the examined transcription factor in the cell line itself, as seen in the respective qPCRs. The experimental setting of overexpression of a transcription factor in a cell line with, with for example already high endogenous expression might thus not present pronounced expression patterns in the luciferase experiments. However, this also does not implicate that it has no effect on the *ADGB* regulatory gene fraction dependent transcription in

general. Some transcription factors need additional co-activation which, again, is not present to the same degree in the different cell lines (Latchman, 1993).

A selection of the observed transcription factors not only influence the androglobin expression, but also has been annotated in the expression regulation of globin chains in general, displaying a common globin property.

The GATA factors bind to *ADGB* gene regulatory regions to exert a stimulating effect. They are furthermore known as master regulatory factors of hemoglobin expression, influencing alpha-, beta- and gamma-globin chain expression (Harigae et al., 2006, Katsumura et al., 2017). Especially GATA2 regulates the alpha-globin chain expression inevitably (Anguita et al., 2004, Brecht et al., 2005, Ikonomi et al., 2000).
Nrf2 activates γ-globin gene expression, as also suggested to be the case for androglobin expression (Zhu et al., 2017, Macari and Lowrey, 2011).
In presence of HP1Gamma, the expression of γ- and the embryonic ϵ-globin is negatively influenced. Furthermore, the androglobin expression is diminished (Wang et al., 2017).
In addition to this, REST displays a mutual regulator of androglobin and neuroglobin (Ngb) expression. GATA3 has further been annotated to influence mammoglobin expression as well (Sangoi et al., 2016, Van Acker et al., 2019).
Trim28 represses and Nrf3 induces β-globin expression (Kobayashi et al., 1999, Hosoya et al., 2013). For those two transcription factors, an opposite expression regulation concerning androglobin could be observed, thus representing a potential involvement in the limitation of androglobin expression.
Even though not all the transcription factors, discussed above seem to mutually influence the gene transcription of the *ADGB* promoter or enhancer and other globin chains, the common globin backbone might represent a similar binding site. Specific research concerning transcription factor binding sites in globin chains might yield to additional points of interaction and further insights in the expression regulation of globins in general.

4.2 Expression regulation of androglobin in the process of spermatogenesis

The process of spermatogenesis, from a germ cell to a mature, haploid spermatozoa can be summarized in proliferation of diploid germ cells, meiosis, and finally extensive differentiation of the haploid cell. The whole process requires multiple, complex, developmental steps and, additionally, optimal surrounding conditions to ensure a proper sperm formation (Cheng and Mruk, 2010). Multiple transcriptional regulators, epigenetic modifications, such as DNA methylation or chromatin remodeling and endocrine, paracrine and autocrine factors have already been identified to be involved in the process (Sassone-Corsi, 1997, Grootegoed et al., 2000, Gunes and Kulac, 2013). Further details about DNA and cell interactions and especially the role of androglobin expression still lack to be specified.

As outlined in the introduction, the expression of androglobin in the process of spermatogenesis is tightly regulated and limited to specific late stages, focusing on post meiotic spermatids at the elongating stage. Neither of the tested mouse cell lines, derived from Sertoli cells (TM4), Leydig cells (TM3), spermatogonia (GC-1) and spermatocytes (GC-2) express abundant levels of androglobin (Hoogewijs et al., 2012b). The expression of androglobin however was proven to be essential for proper spermatogenesis. Infertile men display a diminished androglobin expression and corresponding genetic knockout yield to infertility in male mice. The following analysis about the regulation of androglobin expression in the process of spermatogenesis might yield to greater insights when, where and why the protein is expressed. Further information might consequently lead to future therapeutic goals dealing with male infertility or even path a way to establish a reliable male contraception by stage dependent spermatogenesis inhibition.

The transcription silencing effect of the *ADGB* enhancer underlines an overall tight expression regulation. The current study suggests potential inhibiting regulation of androglobin expression in spermatogenesis related cells via the interaction of the *ADGB* promoter and enhancer with contributions of binding transcription factors.

The cooperation between Rfx2 and the *ADGB* promoter dependent alteration of gene transcription is displayed by an enhancing effect of Rfx2 in the transfections experiments containing the *ADGB* promoter and a viral SV40-enhancer. The corresponding most abundant expression is detected in presence of the 1981bp *ADGB* promoter, which goes along with the presence of a H1t-specific 5´-CCTAGG-3´ palindrome motif in the 1981 *ADGB* promoter base sequence, that has proven to bind Rfx2 (Horvath et al., 2004).
A further correlation between the transcription factor and androglobin can be seen in a *Rfx2* gene knockout experiment in mice, demonstrating a spermatogenesis arrest at the same stage of morphological sperm development as the first androglobin expression is observed (Shawlot et al., 2015, Horvath et al., 2004). Additionally, the knockout mice demonstrate significantly smaller testes, overlapping with the same age (day 25) when *ADGB* expression levels are peaking (Hoogewijs et al., 2012b, Wu et al., 2016). Other studies outline the importance of Rfx2 for the expression of yet other ciliogenic genes (Chung et al., 2012).
The concordance concerning the significance in male fertility, the occurrence at the same stage of spermatogenesis and similar expression profiles suggest a collaboration between the transcription factor and androglobin. The potential direct influence of Rfx2 on *ADGB* gene expression might lack a yet to be discovered *ADGB* enhancer motif to demonstrate a more pronounced binding to the 1981bp *ADGB* promoter. A mutation of the palindrome motif sequence in the *ADGB* promoter, followed by an analysis of the alteration of gene expression might also give further evidence for the binding of the transcription factor.

The transfection experiments including Trim28 display a stimulating tendency on *ADGB* promoter/enhancer dependent gene expression. Accordingly, Trim28 expression has been described to be restricted to the same spermatogenesis stage as androglobin during the development from a spermatocyte to an elongating spermatid (Weber et al., 2002).
The absent androglobin expression in earlier stages of spermatogenesis might be induced by the reducing transcriptional effect of HP1Gamma on the *ADGB* promoter dependent gene expression. The transcription factor is mainly expressed in germ cells and during these stages shown to be essential for proper spermatogenesis (Brown et al., 2010).
A cooperating influence of the transcription factors, such as Trim28 being a co-repressor of HP1Gamma to repress specific expression of genes has been described as well (Weber

et al., 2002). Co-overexpression experiments including both transcription factors might lead to a more substantial transcriptional repression, mediated by HP1Gamma and give further insights regarding the specific androglobin expression regulation in general and specific androglobin expression limitation in germ cells.

p54 is primarily expressed in testis, more specifically in spermatogonia and primary spermatocytes but displays a decreased expression in spermatids (Matsumoto et al., 2005). PU.1 was proven to accumulate in spermatogonic stem cells (Olive et al., 2007). Both transcription factors exhibit decreasing tendencies on the expression level of the *ADGB* promoter and enhancer, displaying a potential transcriptional regulation of androglobin expression in the mentioned spermatogenesis related cells types.

The low androglobin expression levels in Sertoli cells might be due to the observed reducing tendencies of the *ADGB* promoter and enhancer dependent gene expression in the experiments including USF1. The ubiquitous protein USF1 is a transcriptional activator that is also associated with several regulatory key points in spermatogenesis. In perinatal development, USF1 is expressed in the context of Sertoli cell differentiation, determining subsequentially the final number of Sertoli cells in adults (Orth et al., 1988, Wood and Walker, 2009). Furthermore, *usf*-depleted mice present an age-related decline in overall sperm production, possibly caused by a reduction of spermatogonial stem cells related to Sertoli cell loss (Faisal et al., 2019).
Another correlation between the transcription factor and androglobin expression in terms of profound impact on continuous spermatogenesis is drawn by the binding and activation of the Miwi promoter.
Miwi belongs to the P-element-induced wimpy testis (PIWI) family and controls the expression of small non-coding RNAs (Suárez Alonso et al., 2019). *Miwi*-gene knockout results in a spermatogenic arrest at the beginning of the round spermatid stage (Deng and Lin, 2002). USF1 interacts with the Miwi promoter during meiosis of male germ cells (Hou et al., 2012). Our research group could also establish proof of a robust association between androglobin and Miwi in terms of epigenetic modifications (Suárez Alonso et al., 2019). Overall, Miwi displays a potential common expression regulator of USF1 and androglobin.

The examined CEBPβ is present in high levels in Sertoli cells and is involved in immunological and inflammatory processes in testes (Lardenois et al., 2010, Pal et al., 2009, Yao et al., 2011). Surprisingly, an enhancing transcriptional effect is observed in the corresponding transfection experiments, especially in presence of the *ADGB* promoter. The correlation between the transcription factor and the expression regulation of androglobin could suggest a potential additional presence of CEBPβ in spermatids, stimulating androglobin expression.

Another explanation is a potential interaction in the associated field of control of immunological and inflammatory events in testis (Lardenois et al., 2010). The suggested role of androglobin expression in the context of immunology is outlined in the next section.

4.3 The potential influence of androglobin in immunological processes, focused on the lymphatic cell line

No direct correlation between androglobin and its role in the immunological response has been stated until now. This research topic, nonetheless, involves several transcription factors that, on one hand have a regulating effect on *ADGB* gene expression and on the other hand have been annotated to be involved in B-cell lymphopoiesis. This leads to the hypothesis that, unlike hemoglobin, which is commonly involved in the expression of erythrocytes, therefore representing the myeloid branch androglobin could play part in the lymphoid branch, particularly in the B-cell differentiation. Publicly available single cell RNA-sequencing data from mice indicate expression in this branch and support this hypothesis.

The acquired, opposed to the innate immune system is represented by the lymphatic cell line. The corresponding specific immunology response of B- and T-lymphocytes against pathogens is unique in the human organism. Even though the morphological appearance of the two lymphocyte types is similar, their responsibilities differ.

B-lymphocytes and corresponding B-plasma cells represent the humoral, adaptive immune system, which builds specific antibodies against antigens. T-lymphocytes can roughly be divided into CD4+ and CD8+ cells. They are responsible for the primary

immune answer, hence oversee the cell immunity. The T-cell receptors in this context recognizes alien antigens, which then leads to the accompanying elimination.

As part of the hematopoiesis, the lymphoid cells derive from hematopoietic stem cell in the fetal liver and later in the bone marrow. After the evolution from multipotent progenitors into myeloid or lymphoid multipotent progenitor cells, further specification of the lymphoid branch is divided into the formation of B- and T-lymphocytes.
Whereas the following differentiation of T-lymphocytes is characterized by the development of the T-cell receptor and the migration to the thymus, the B-cell lymphopoiesis takes place in the bone marrow (Cano, Lopera, 2013).

The path from a hematopoietic stem cell (HSC) to a mature B-cell lymphocyte, that is being released to the periphery, involves several intermediate stages and a complex network of transcription factors that influence the developmental progress (Somasundaram et al., 2015). HSC along the way evolve into multipotent progenitors (MPP), lymphoid-primed multipotent progenitors (LMPP), B lymphoid progenitors (BLP), followed by pre-pro-B-cells and B-cell fractions (Figure 47) (Hardy et al., 2007). The transcription factors PU.1, early B-cell factor (EBF) and Pax5 have been shown to be, among other transcription factors essential for early B-cell lymphopoiesis (Figure 47) (Xue et al., 2007). This discussion focuses on the transcription factors that overlap with the *ADGB* gene regulatory network.

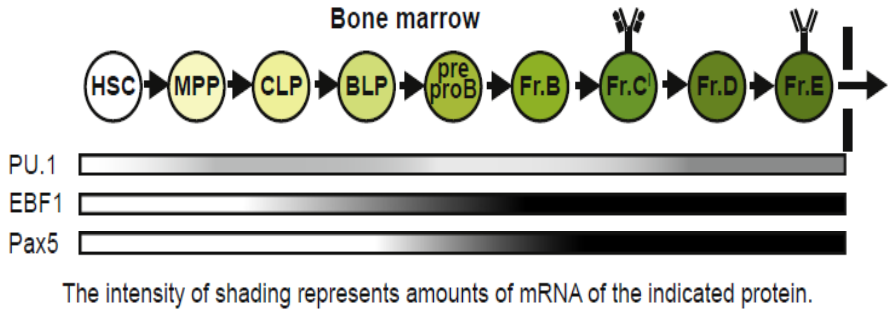

Fig. 47. Progression of B-cell development and stage dependent expression of key transcription factors.
The development to LMPP would be presented between MPP and CLP. The dashed line at the end of the development indicates the transit to the periphery. The corresponding mRNA amounts for each transcription factor were estimated using data from Gene Commons and the Immunological Genome Project (ImmGen).
(Figure edited from Hagman, 2015)

The fundamental differentiation of MPP cells between the myeloid and the lymphoid lineage is regulated by PU.1 in a dose-dependent manner (Zhang et al., 2013, Umazume

et al., 2018). Experiments performed in mutant hematopoietic progenitor cells present a B-cell induction in presence of low PU.1 concentration, whereas B-cell development is blocked under high concentrations (DeKoter and Singh, 2000). Further lymphoid differentiation is shown to be PU.1 independent (Pang et al., 2018). For the proceeding cell differentiation Erythroblast Transformation Specific (Ets) family proteins, such as Ets-1 or GABPα, fulfil the function of PU.1 (Hagman and Lukin, 2006).

The induction of B-cell development is, among other factors regulated by the generation of signaling receptors (Busslinger, 2004). Total *Pu.1* knockout results in failure to express the essential interleukin 7 α (IL-7) receptors and a corresponding total loss of B-cells expression (DeKoter et al., 2002 Hagman, 2015).

The function of the IL-7 receptor is further explained by the essential recruitment of the transcription factor EBF1, which is discussed in the section below (Tsapogas et al., 2011, Hagman, 2015).

Alongside, PU.1 is described to recruit the transcription factor p300, a coactivator of early B-cell progenitors (Hagman and Lukin, 2006).

Regarding an association with androglobin expression regulation, PU.1 mutually displays a silencing effect in the transcription factor overexpressing transfections. The recruited Ets-1, GABPα and p300 have shown to bind to the *ADGB* enhancer sequence, as revealed by UCSC-integrated ENCODE ChIP-sequencing data.

EBF1 is exclusively expressed in B-cells in the hematopoietic system (Hagman et al., 1993). *Ebf1* knockout experiments demonstrate a differentiational arrest at a common lymphoid progenitor cells (CLP) to early pre-pro-B stage (Hagman, 2015).

Additionally, EBF1 appears to act as a central coordinator of several target genes at distinctive developmental steps (O'Riordan and Grosschedl, 1999). The explicit binding and regulation of the Pax5 promoter, as well as another annotation of p300 recruitment are noteworthy in this context (Zhao et al., 2003, Hagman and Lukin, 2006, O'Riordan and Grosschedl, 1999).

EBF1 further blocks the expression of GATA3, which is essential for T-cell differentiation (Banerjee et al., 2013, Liu et al., 2016).

Apart from the recruitment or expression inhibition of the overlapping transcription factors, unpublished data of our research group could also demonstrate an EBF1-dependent stimulating effect on *ADGB* gene expression itself.

The transcription factors Pax5 and CEBPβ represent another crosslink between the gene regulatory network of androglobin and B-cell differentiation. Compared to the previously discussed PU.1 and EBF1, their impact is mainly focused on later stages of B-cell differentiation (Medina and Singh, 2005, Adams et al., 1992).
Pax5 has a moderating role, influencing surrounding circumstances to enable the expression of B-lineage genes or repress the transcription of myeloid genes and formation of natural killer (NK) or T-cells (Zhang et al., 2013, Hagman et al., 2012, Delogu et al., 2006). A more than 800-fold gene enhancement is further stated between the cooperation of the mentioned Ets-1 and Pax5 (Fitzsimmons et al., 2009). Gene knockout experiments result in loss of mature B-cells and a developmental arrest at a stage between pre-pro-B and pro-B-cell (Horcher et al., 2001, Hagman, 2015).
CEBPβ contribution in the field of B-cell differentiation focuses on two phases. On one hand, *Cebpβ* knockout mice display a reduced responsiveness to the discussed IL-7, suggesting a trans-activating role of the transcription factor (Chen et al., 1997). On the other hand, studies suspect its presence at later stages, corresponding to immunoglobulin (Ig) chain expression in mature B-cells (Cooper et al., 1994). Noteworthy is the transdifferentiation of B-cell precursors into macrophages and antagonization of Pax-5 activity in presence of augmented CEBPβ expression, making a precise control of the transcription factor expression inevitable (Xie et al., 2004, Medina and Singh, 2005).
Overexpression of the transcription factors Pax5 and CEBPβ in presence of *ADGB* gene fragments display a clear induction of gene expression.
The *ADGB* promoter/enhancer-dependent gene expression exhibit a mutual enhanced expression in presence of Pax5 and CEBPβ. Our ChIP experiments could further confirm Pax5 binding to the *ADGB* promoter region.

As mentioned above, GABPα is associated with decisive interaction between PU.1, EBF and Pax5, or respectively their target genes (Xue et al., 2007, Rosmarin et al., 2004).

Consecutive gene knockout leads to an overall reduction of pro-B-cells in bone marrow and a blockage of further B-cell differentiation (Xue et al., 2007, Yang et al., 2011).
Apart from its impact on other hematopoietic cell lines, Trim28 is described to mediate essential epigenetic regulations for B-cell differentiation (Chikuma et al., 2012, Santoni de Sio et al., 2012).
Although the overexpression transfection experiments involving GABPα and Trim28 in presence of *ADGB* gene regulatory elements do not alter the corresponding gene expression significantly, both transcription factors have been annotated by ENCODE data to bind to the *ADGB* enhancer region. The two transcription factors further display *ADGB* gene regulatory tendencies that are in favor of the presented hypothesis. Overall, this could suggest a stronger collaboration between androglobin and the transcription factors under alternative circumstances.

A closer look on the transcription factor induced gene regulatory network of B-cells and androglobin expression reveals supplementary overlapping transcription factors. However, they have not been allocated to specific B-cell developmental stages.

Analysis of peripheral blood and bone marrow of GATA2 deficient patients reveal reduced B-progenitors and B-cell counts, meanwhile relative T-cell counts are increased (Novakova et al., 2016, Ganapathi et al., 2015). In this context a potential direct repression of PU.1 expression is considered as accountable (Kitajima et al., 2006).
In the context of immune response, Rfx2 is associated with the regulation of MHC II receptor gene expression which are, among other antigen presenting cells also distinguishable in B-lymphocytes (Reith et al., 1988).
Both transcription factors also display a mutual enhancing effect on *ADGB* gene regulatory regions dependent transcription.

Next to the transcriptional influence, a selection of the mentioned transcription factors has also been annotated to be involved in B-cell derived malignancies. CEBPβ is associated with proliferation and survival of malignant plasma cells in the context of multiple myelomas and Trim28 functions as a tumor promoter in B-cell non-Hodgkin lymphomas (Zhang et al., 2018, Pal et al., 2009). An additional crosslink is displayed by

unpublished data of our research group that establishes proof of HDLM-2 cells, a Hodgkin lymphoma-derived cell line expressing a transcriptional variant of androglobin.

In future experiments it would be interesting to investigate the impact of other essential transcription factors for B-cell development, such as E2A or Ikaros on androglobin gene expression as well. To draw further conclusions, the exposure of transcription factors under more physiological conditions, such as an adapted amount of PU.1 could give further insight into the exact expression regulation of androglobin. An analysis of the behavior of HSC or B-lymphocytes, such as the M12 cell line in presence of androglobin might give additional evidence for a concrete interaction.

In conclusion, the discussed transcription factors exhibit the same expression patterns on B-cells compared to *ADGB* gene expression. The overlapping transcriptional behavior could suggest a mutual transcription regulation and hence similar transcription conditions and potentially overlapping fields of expression. These results could suggest a potential, yet to be discovered function of *ADGB* gene expression in the complex network B-cell development or general expression. An *actual* involvement of androglobin in immunological processes, however, still lacks to be elucidated.

4.4 Androglobin – a counterpart of the androgen receptor?

The hormonal regulation represents an additional, non-redundant factor to ensure proper male fertility. Testosterone and the corresponding dihydrotestosterone are the most established hormones in this context. They exert their influence via binding to a ligand-activated transcription factor, the androgen receptor (AR). The ligand-receptor complex further binds to regulatory regions of genes, androgen response elements to influence specific translation (O'Hara and Smith, 2015). Disruption of the AR can result in a feminized phenotype, display as cryptorchidism or as an impaired spermatogenesis (Kaftanovskaya et al., 2012). In particular, aberrations of AR expression do not disrupt spermatogenesis completely, but rather interfere with sperm maturation (Hiort et al., 2000). AR mutations in Leydig cells cause a detachment of spermatids from Sertoli cells, arresting spermatogenesis at the round to elongating transition stage, as it is observed in

Adgb$^{-/-}$ knockout mice (O'Hara and Smith, 2015, Chang et al., 2013, Hoogewijs et al., 2012b).

The impact of the AR on transactivation of target genes is altered by binding coregulators that either enhance or repress its activity through, for example histone modification or chromatin remodeling (Davey and Grossmann, 2016, Heinlein and Chang, 2002). A comparison between the gene regulatory network of androglobin and the AR reveals several overlapping patterns.
Interestingly, the overlapping transcription factors are associated with opposite effects on the expression regulation of the AR, compared to the observed results of androglobin. GABPα, p54 and PU.1 stimulate AR signaling at the respective target genes (Ishitani et al., 2003, Lin et al., 2017, Sharma et al., 2014). REST, CEBPβ, Nrf2 and PML-IV are in contrast annotated to surpress AR-mediated transactivation (Wang et al., 2007, Barakat et al., 2015, Svensson et al., 2014, Schultz et al., 2014, Yang et al., 2004).
The discrepancy displays androglobin as a pontential counterpart of AR expression, mediating opposite reactions at the respective targets.

4.4.1 Androglobin – a potential counterpart of the androgen receptor in B-cell lymphopoiesis

Next to a predominantly expression in the reproductive system, the AR mediates the function of androgens in a broad spectrum of different body compartments as well. Bone and muscle tissue or the cardiovascular system are annotated in the literature as examples (Rana et al., 2014). Furthermore, an androgen dependent attenuation of inflammatory processes is observed (Traish et al., 2018, Rana et al., 2014). In relation to the aspects discussed above, the opposite alterations of the immune and hematopoietic system of androglobin and the AR are noteworthy.

Among the AR-dependent modification of other anti-inflammatory pathways, the production of B-lymphocytes is modulated by androgen and corresponding AR signaling (Viselli et al., 1997).
A dependency was first shown in an elevated B-lymphocyte population of hypogonadal and castrated male mice (Smithson et al., 1998). AR and B-cell specific AR knockout mice, as well as wild type castrated mice display an increased proliferation and a reduced

apoptosis of B-cells, resulting in an overall increase of pro-B, pre-B, and immature B-cell populations in the bone marrow (Ellis et al., 2001, Altuwaijri et al., 2009).
Olsen et al. propose lacking androgen stimulation of bone marrow stromal cells as a possible explanation for this observation. However, the exact pathomechanisms remain unclear (Olsen and Kovacs, 2001).

In this thesis, potential androglobin involvement in specific B-cell expression and further differentiation is outlined, as opposed to the androgen dependent reduction of the B-cell population / increase of B-cells under androgen deficient conditions. Furthermore, the described transcription factor dependent expression regulation of androglobin and the AR are contradicting for the most part.
The observed increase of B-cell population in the context of androgen deficiency could therefore be related to an increased androglobin expression as an additional contrary observation between androglobin and AR expression and underline the hypothetical role of androglobin as AR counterpart.

4.4.2 Androglobin – a potential future target in prostate cancer

In addition to the tissue differentiations outlined above, the development and maintenance of function of the prostate is also AR and consequently hormone expression dependent.
In the context of the prostate however, not only the physiological role of hormones, but also the pathological role in cancers is under extensive investigation. Operative or chemical hormone deprivation, using antiandrogens or other endogenous hormone therapies are a common supplementary or palliative therapy for the treatment of prostate cancer (PC). The alteration of the hormone concentration is supposed to inhibit the growth stimulus and consequently reduce the tumor mass and stop further progression.
In some patients however, the applied treatment lacks to have an effect, or the desired therapy success diminishes over the time. The PC is consequently labeled castration resistant or hormone refractory. The majority of therapy resistant PCs nonetheless maintain AR expression throughout cancer progression (Heinlein and Chang, 2004).
Fujita and Nonomura reviewed multiple mechanisms, ranging from AR point mutations to alternative androgen biosynthesis and altered AR cofactor expression that facilitate the adaption of the PC to the diminished levels of hormones under therapy (Fujita and

Nonomura, 2019). The exact pathways of tumor adaption however still lack to be elucidated.

As observed in other tissues, the expression regulation of the AR in PC and the corresponding hormone dependency reveal overlapping transcription factors with androglobin expression as well. Except for GATA2 and TRIM28, the transcription factors also exhibit a reverse impact on the expression levels. This suggests androglobin as a potential counterpart of the AR in PC, thus representing androglobin expression regulation as a potential future target to prevent tumor growth.

In the context of hormone dependent PC growth, a transcriptional elevation of GABPα has been annotated and HP1Gamma was shown to distinctly promote androgen dependent cancer growth (Itsumi et al., 2013, Sharma et al., 2014). The corresponding *Hp1gamma* knockout reveals reduced AR expression and improved clinical outcome (Itsumi et al., 2013, Slezak et al., 2013). The transcription factor CEBPβ downregulates the expression of the AR and REST is linked to hormone resistance in PC tissue (Wang et al., 2007, Liang et al., 2014). The corresponding expression of both transcription factors is however associated to poorer clinical outcome. Chronic inflammation in case of CEBPβ expression and the expression of neuroendocrine tumor cells in presence of REST are surmised as putative precursors for the respective tumor proliferation (Kim and Fields, 2008, Adamo et al., 2019, Flores-Morales et al., 2019, Svensson et al., 2014).

So far, no direct correlation was established between the expression regulation of the AR and androglobin. The results of this thesis however display transcriptional controversies. The opposite results in the transcriptional regulation, as well as the additional overlapping transcription factors are noteworthy and reveal a potential future field of interest.
This thesis could only generate speculations but opens novel avenues for future extensive research.
To shed some light into potential interactions, experiments with AR negative PC cell lines PC3 or DU-145 and hormone dependent PC cell lines, such as LNCaP cells in presence of androglobin could be carried out. Additionally, it would be interesting to investigate androglobin expression in presence of androgen expression and *vice-versa*.

Other approaches could study the androgen level of androglobin deficient patients *in vivo* or the androglobin level of androgen deprived PC patients.

4.5 Unclassified transcription factors and potential correlation to androglobin expression

The following part of the discussion includes transcription factors that have been annotated by UCSC genome browser integrated ENCODE data to bind to the *ADGB* enhancer region. The corresponding transfection experiments further display a distinct expression regulation of the *ADGB* dependent gene transcription.
Even though they neither have been associated in any known context of androglobin expression, nor have they been linked to any of the research topics outlined above, the exploration of the transcription factors could reveal potential reasons why they bind to the *ADGB* gene region and outline potential fields of androglobin expression for future studies.

As suggested by the names, Nrf2 and Nrf3 belong to the same family of transcription factors. Furthermore, they are both involved in the cellular defense against oxidative stress. Concerning the antioxidative response element (ARE) mediated gene expression, they however can be considered as counterparts. In response to an elevated level of oxidative stress Nrf2 expression is annotated to be increased, as observed in the context of spermatogenesis (Nakamura et al., 2010, Nguyen et al., 2003). Nrf3 expression is contrary considered as a negative regulator, maintaining the ARE dependent gene expression at a normal level (Sankaranarayanan and Jaiswal, 2004).
Referring to the results of androglobin expression, they also display converse results. Nrf2 functions as transcriptional inductor and Nrf3 as repressor.
In the context of male fertility, *Nrf2* knockout reveals infertility, caused by an age increasing damage in the seminiferous tubules, decreased sperm counts and motility (Nakamura et al., 2010). *Nrf3* knockout mice however appear normal and do not present any corresponding restrictions (Derjuga et al., 2004).
Androglobin could in this context potentially display one of the Nrf2 dependent expressed genes, in response to elevated levels of oxidative stress supposedly in the region of testis. Globins in general have often been associated with the oxidative metabolism, as described

above. The examined expression in the hypoxic lumen of the seminiferous tubules represents furthermore an increased likelihood for oxidative stress and the corresponding genetic knockouts display a similar stage of spermatogenesis arrest.
Contrary to this observation, Suárez et al. display that androglobin expression is not elevated in presence of ROS, suggesting a different role other than detoxification or sensing (Suárez Alonso et al., 2019).

PML-IV is an isoform of the PML protein, which is expressed by the PML nuclear bodies (NBs). Localized in the cell nucleus, the PML NBs have been linked to the regulation of multiple cellular processes such as DNA transcription or repression (Bernardi and Pandolfi, 2007, Hsu and Kao, 2018, Lallemand-Breitenbach and de The, 2010).
A transcriptional regulation of PML NBs has been observed in presence of cellular or oxidative stress (Terris et al., 1995, Lallemand-Breitenbach and de The, 2010).
The connection to androglobin expression regulation could in this context represent another example of androglobins involvement in cellular defense against oxidative stress.
PU.1, Trim28 and the GATA factors have furthermore been demonstrated to be mutually involved in the respective expression regulations of PML-IV (Tsuzuki et al., 2000, Kepkay et al., 2011, Yoshida et al., 2007).
The obtained PML-IV transfection results however do not display a pronounced effect. On the one hand, this could be explained by a missing co-activation, supposedly of one of the overlapping transcription factors. On the other hand, it has also been demonstrated that the specific isoform PML-IV is minor quantitatively expressed in many cell lines and, compared to the other isoforms evolutionary not conserved (Condemine et al., 2006). This proposes a potentially higher expression regulation of androglobin by other isoforms of the PML protein. Future studies including different co-transfections and the other PML isoformes could yield to further insights.

The transcription factor TEAD4 is predominately annotated to early embryonic developmental stages. The corresponding gene knockout experiments reveal a specific transcription factor dependency during the implantation into the uterine endometrium (Yagi et al., 2007, Nishioka et al., 2008). TEAD4 is also regulating the expression of the transcription factor GATA3 in this context (Ralston et al., 2010).

Disruption of gene expression in later embryonic stages display no developmental impact (Nishioka et al., 2008).
The observation about TEAD4 dependent androglobin expression regulation displays a hypothesis of androglobin expression in early stages of embryonic development. A scientific investigation of *Adgb*$^{-/-}$ knockout mice could shed light into a potential impact. The binding of the transcription factor could furthermore be explained by a potential TEAD4 dependent expression regulation and recruitment of GATA3, exceeding the environment of embryonic development. The potential impact of GATA3 on androglobin expression is further outlined below. In relation to the obtained transfection results, the bias of general TEAD4 binding and activation of SV40 enhancers should also be taken into consideration (Pobbati and Hong, 2013).

The GATA factor family is a group of transcription factors that overall consists of six transcription variants. Whereas most annotations about expression of the first three, GATA1 – 3, correspond to the context of hematopoietic cell differentiation, a broader spectrum of tissues are associated with GATA4 – 6 expression (Molkentin, 2000, Weiss and Orkin, 1995).
A close expression regulation among the individual GATA factors is further annotated such as the GATA switch (e.g. among GATA1 and GATA2 in the erythroid lineage differentiation) or the cofactor dependent regulation (e.g. through FOG1: friend of GATA1) (Ray et al., 2009, Bresnick et al., 2010, Pal et al., 2004).

In the context of androglobin expression, the pronounced impact of GATA factors concerning the expression regulation of other globin chains, as outlined above and the involvement in the gonadal development and maintenance of the reproductive function are noteworthy (Tevosian et al., 2002, Tevosian, 2014, Yomogida et al., 1994).
ChIP-sequencing experiments, accessible via the UCSC genome browser integrated ENCODE data present GATA2 and GATA3 binding to multiple *ADGB* gene regions. The *ADGB* promoter symbolizes in particular a target for both transcription factors and the *ADGB* enhancer for GATA2. The corresponding transfections further underline distinct gene expression regulations and both, the *ADGB* promoter and enhancer nucleotide sequences contain the proposed GATA binding motif 5´-WGATAR-3´ (in

which W indicates A/T and R indicates A/G). Other studies indicate the 5´-GATC-3´ motif, which is also present in both sequences as preferential binding site of GATA2 and GATA3 (Ko and Engel, 1993). The ChIP experiments could furthermore validate a concrete GATA3 antibody binding to specific *ADGB* promoter and enhancer sequences.

In addition to the attributed features that are mentioned in other sections of this discussion, GATA2 is mainly required for the proliferation and survival of early hematopoietic cells and GATA3 is expressed in T-cells and controls T-cell lymphopoiesis (Ting et al., 1996, Hosoya et al., 2009, Tsai and Orkin, 1997).

Despite the common association to the globin family, none of the mentioned GATA2 or GATA3 characteristics seem to overlap with annotated *ADGB* gene expression or the outlined potential future fields of research. This leaves the actual circumstances of interaction between the transcription factors and *ADGB* gene expression *in vivo* yet to be established.

On one hand, the observed results could display new potential fields of androglobin expression or GATA2 and GATA3 involvement in additional regulatory processes, beyond those that are annotated. On the other hand, the performed experiments could display an example of transcriptional regulation *in vitro*, that is not observable *in vivo*. On this behalf it should be mentioned that only a small fraction of the potential more than 7 million GATA motifs in the human genome are actually occupied *in vivo*, representing a redundancy of GATA binding motifs (Fujiwara et al., 2009, Yu et al., 2009, Linnemann et al., 2011). Furthermore, approximately 90% of the GATA occupied sites are localized away from promoters, suggesting a rather long-range transcriptional control involving several co-factors (Fujiwara et al., 2009). In the observed transfections, GATA3 could bind indirectly via another, not identified transcription factor to the *ADGB* gene region, that is present endogenously in the cell model.

The transfection results of the mutated GATA3 binding site sequence of the 1981bp *ADGB* promoter underline the gap between anticipation and obtained results. Even though the sequence represents a tremendously high probability of binding, the actual mutation does not prevent the GATA3 dependent enhancement of gene transcription as presumed. The results much rather display an overall higher transcriptional level,

suggesting a mutation dependent prevention of transcriptional repression or repressor binding.

In conclusion, additional research is required before further assumptions about the correlation between the GATA transcription factors and androglobin expression can be made. In this context, transfection experiments including the other GATA transcription factors could give insights about a potential GATA dependent androglobin expression regulation in the field of andrology or hematopoiesis. A mutation of a larger GATA binding site sequence could prevent the transcription factor binding and potentially inhibit the observed transcriptional enhancement.

At the start of this project, the *ADGB* promoter and the *ADGB* enhancer sequences were chosen solely based on assumptions carried out by mining epigenomic data from ENCODE. The cloning in luciferase reporter gene vectors could for the first time establish proof of an actual gene transcription regulation by these regions *in vitro* and reveal the presumed enhancer as a transcriptional repressor.
The consecutive exposure to transcription factors established a more distinguished picture of the actual transcriptional regulation of androglobin and the ChIP experiments revealed concrete binding sites for two of them in the corresponding gene regions.
The following analysis of interactions gave new insights about the complex network of transcriptional regulation of androglobin expression and exposed potential androglobin involvement in the field of hematopoiesis and as counterpart of the AR expression.
Our side-directed mutagenesis experiments could at last refute the assumptions about the GATA3 binding site sequence in the 1981bp *ADGB* promoter.

Conclusion

Overall, this project exposes several new potential research targets and future fields for therapeutic intervention. Some of the observations complement each other and others contradict each other, as observed in the discussion about the GATA factors.
Regarding this point, one must take into consideration that the annotations concerning the transcription factor binding display an overview of all transcription factors. A sequential

expression of single transcription factors can be anticipated, depending on the respective situation of need for androglobin expression.

Furthermore, the collected and annotated data displays results from experiments that were performed *in vitro*. The exact mechanisms remain to be specified and verified *in vivo*, before concrete conclusions about processes or potential fields of androglobin expression can be drawn.

Given the relatively short scientific period since its discovery, the precise function of androglobin gene expression in the complex network of the human body remains unclear. Current knowledge indicates that androglobin is an essential protein with astonishing diverse features. Therefore, androglobin expression and its corresponding regulation represent a truly exciting and innovative research topic, also for future research.

5 SUMMARY

5.1 English summary

Androglobin was discovered in 2012 as most recent member of the globin protein family. Most abundant mRNA expression levels were detected in testes and the corresponding protein expression was proven to be essential for male reproduction and influence in particular the process of sperm maturation. The strong evolutionary conservation of androglobin suggests a strongly conserved function. The protein structure displays a genetic chimerism with a central, characteristic globin domain.

Despite extensive research the exact circumstances where, when and why the protein is expressed however still remain unclear. This study aims to explore the expression regulation of the androglobin gene locus and reveal expression patterns and interactions between gene regulatory regions.

In this context, three sizes of the androglobin promoter and the androglobin enhancer were identified by *in silico* analysis of the androglobin gene region and cloned into three different vector constructs. To investigate ADGB promoter- and enhancer-dependent gene regulation, the various constructs were overexpressed in different cell lines with and without co-transfection of numerous transcription factors. For two specific transcription factors, GATA Binding Protein 3 (GATA3) and Paired Box 5 (Pax5) chromatin immunoprecipitation (ChIP) experiments were performed to identify the specific binding sites in the corresponding gene regions. Additionally, a GATA3 binding site sequence was exemplary mutated and the construct transfected to evaluate the specific binding capacity.

This research project could overall establish proof of a distinct gene regulatory network between the androglobin promoter and the androglobin enhancer and display corresponding interactions with transcription factors.

The interpretation of the outlined interactions hint for an androglobin expression limitation to specific stages of spermatogenesis and corresponding cells. Furthermore, this thesis could link androglobin expression for the first time to B-cell lymphopoiesis and the expression regulation of the androgen receptor, expanding the knowledge about the astonishing diversity of globin expression in general and, in particular of androglobin.

5.2 German summary

Das im Jahr 2012 entdeckte Androglobin ist das jüngste Mitglied in der Gruppe der Globine. Die höchste Expressionsrate konnte in Testis in Zusammenhang mit der Spermatogenese nachgewiesen werden und sie korreliert mit der männlichen Fruchtbarkeit. Die hohe evolutionäre Konservierung von Androglobin weist zusätzlich auf eine essenzielle Rolle im Organismus hin. Der molekulare Aufbau des Proteins zeigt eine chimäre Struktur, wobei die charakteristische Globindomäne zentral gelegen ist.

Das vorliegende Forschungsprojekt fokussiert sich auf die Expressionsregulation des Androglobin Genlokus. Das Forschungsprojekt hat das Ziel, Expressionsmuster und Interaktionen zwischen genregulatorischen Regionen darzustellen.

In diesem Zusammenhang wurden drei verschiedene Varianten des Androglobin Promoters und der Androglobin Enhancer durch in silico Analyse des Androglobin Genlokus identifiziert und jeweils in Gen Vektoren kloniert. Im Anschluss wurden die entstandenen Produkte in drei Zelllinien transfiziert und die Transkription in Ab- und Anwesenheit von Transkriptionsfaktoren gemessen und verglichen. Für zwei der Transkriptionsfaktoren, GATA Binding Protein 3 (GATA3) und Paired Box 5 (Pax5) wurde die den Transkriptionsfaktor bindende Gensequenz der untersuchten Androglobinabschnitte mittels Chromatin-Immunpräzipitations-Sequenzierungsexperimente (ChIP-seq) ermittelt. Exemplarisch wurde zusätzlich eine der GATA3 bindenden Gensequenzen mutiert, das Produkt transfiziert und die veränderte Transkription gemessen.

Zusammenfassend konnte dieses Forschungsprojekt ein enges genregulatorisches Netzwerk zwischen dem Androglobin Promoter, dem Androglobin Enhancer und den beteiligten Transkriptionsfaktoren abbilden und damit verbunden Erklärungen für die gezielte Expression des Proteins in bestimmten Phasen, sowie in ausgewählten Zellen der Spermatogenese geben. Des Weiteren konnte erstmalig ein Zusammenhang zwischen einer Androglobin Expression und der B Zell Lymphopoese, sowie der Expressionsregulation des Androgen Rezeptors hergestellt werden.

Insgesamt dient die vorliegende Arbeit somit der Erweiterung des Verständnisses über die erstaunliche Vielfalt der Globin Funktionen im Allgemeinen, mit besonderem Fokus auf die des Androglobins.

6 REFERENCES

1.) ADAMO, H., HAMMARSTEN, P., HAGGLOF, C., DAHL SCHERDIN, T., EGEVAD, L., STATTIN, P., HALIN BERGSTROM, S. & BERGH, A. 2019. Prostate cancer induces C/EBPbeta expression in surrounding epithelial cells which relates to tumor aggressiveness and patient outcome. *Prostate,* 79**,** 435-445.

2.) ADAMS, B., DORFLER, P., AGUZZI, A., KOZMIK, Z., URBANEK, P., MAURER-FOGY, I. & BUSSLINGER, M. 1992. Pax-5 encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev,* 6**,** 1589-607.

3.) ALTUWAIJRI, S., CHUANG, K. H., LAI, K. P., LAI, J. J., LIN, H. Y., YOUNG, F. M., BOTTARO, A., TSAI, M. Y., ZENG, W. P., CHANG, H. C., YEH, S. & CHANG, C. 2009. Susceptibility to autoimmunity and B cell resistance to apoptosis in mice lacking androgen receptor in B cells. *Mol Endocrinol,* 23**,** 444-53.

4.) ANDERSSON, R., SANDELIN, A. & DANKO, C. G. 2015. A unified architecture of transcriptional regulatory elements. *Trends Genet,* 31**,** 426-33.

5.) ANGUITA, E., HUGHES, J., HEYWORTH, C., BLOBEL, G. A., WOOD, W. G. & HIGGS, D. R. 2004. Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *EMBO J,* 23**,** 2841-52.

6.) BAHLER, M. & RHOADS, A. 2002. Calmodulin signaling via the IQ motif. *FEBS Lett,* 513**,** 107-13.

7.) BAI, S. N. 2015. The concept of the sexual reproduction cycle and its evolutionary significance. *Front Plant Sci,* 6**,** 11.

8.) BANERJEE, A., NORTHRUP, D., BOUKARABILA, H., JACOBSEN, S. E. & ALLMAN, D. 2013. Transcriptional repression of Gata3 is essential for early B cell commitment. *Immunity,* 38**,** 930-42.

9.) BARAKAT, D. J., ZHANG, J., BARBERI, T., DENMEADE, S. R., FRIEDMAN, A. D. & PAZ-PRIEL, I. 2015. CCAAT/Enhancer binding protein beta controls androgen-deprivation-induced senescence in prostate cancer cells. *Oncogene,* 34**,** 5912-22.

10.) BERNARDI, R. & PANDOLFI, P. P. 2007. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol,* 8**,** 1006-16.

11.) BIRNBOIM, H. C. & DOLY, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res,* 7**,** 1513-23.

12.) BRECHT, K., SIMONEN, M., KAMKE, M. & HEIM, J. 2005. Hematopoietic transcription factor GATA-2 promotes upregulation of alpha globin and cell death in FL5.12 cells. *Apoptosis,* 10**,** 1063-78.

13.) BRESNICK, E. H., LEE, H. Y., FUJIWARA, T., JOHNSON, K. D. & KELES, S. 2010. GATA switches as developmental drivers. *J Biol Chem,* 285**,** 31087-93.

14.) BROWN, J. P., BULLWINKEL, J., BARON-LUHR, B., BILLUR, M., SCHNEIDER, P., WINKING, H. & SINGH, P. B. 2010. HP1gamma function is required for male germ cell survival and spermatogenesis. *Epigenetics Chromatin,* 3**,** 9.

15.) BURMESTER, T., EBNER, B., WEICH, B. & HANKELN, T. 2002. Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Mol Biol Evol,* 19**,** 416-21.

16.) BURMESTER, T. & HANKELN, T. 2009. What is the function of neuroglobin? *J Exp Biol,* 212**,** 1423-8.

17.) BURMESTER, T. & HANKELN, T. 2014. Function and evolution of vertebrate globins. *Acta Physiol (Oxf),* 211**,** 501-14.

18.) BURMESTER, T., WEICH, B., REINHARDT, S. & HANKELN, T. 2000. A vertebrate globin expressed in the brain. *Nature,* 407**,** 520-3.

19.) BUSSLINGER, M. 2004. Transcriptional control of early B cell development. *Annu Rev Immunol,* 22**,** 55-79.

20.) CANO, R.L.E., LOPERA, H.D.E. 2013. Chapter 5. Introduction to T and B lymphocytes. In: ANAYA, J.M., SHOENFELD, Y., ROJAS-VILLARRAGA, A., et al., editors. Autoimmunity: From Bench to Bedside. Bogota (Colombia): *El Rosario University Press*; 2013 Jul 18.

21.) CHANG, C., LEE, S. O., WANG, R. S., YEH, S. & CHANG, T. M. 2013. Androgen receptor (AR) physiological roles in male and female reproductive systems: lessons learned from AR-knockout mice lacking AR in selective cells. *Biol Reprod,* 89**,** 21.

22.) CHEN, X., LIU, W., AMBROSINO, C., RUOCCO, M. R., POLI, V., ROMANI, L., QUINTO, I., BARBIERI, S., HOLMES, K. L., VENUTA, S. & SCALA, G. 1997. Impaired generation of bone marrow B lymphocytes in mice deficient in C/EBPbeta. *Blood,* 90**,** 156-64.

23.) CHENG, C. Y. & MRUK, D. D. 2010. The biology of spermatogenesis: the past, present and future. *Philos Trans R Soc Lond B Biol Sci,* 365**,** 1459-63.

24.) CHIKUMA, S., SUITA, N., OKAZAKI, I. M., SHIBAYAMA, S. & HONJO, T. 2012. TRIM28 prevents autoinflammatory T cell development in vivo. *Nat Immunol,* 13**,** 596-603.

25.) CHOMCZYNSKI, P. & SACCHI, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem,* 162**,** 156-9.

26.) CHUNG, M. I., PEYROT, S. M., LEBOEUF, S., PARK, T. J., MCGARY, K. L., MARCOTTE, E. M. & WALLINGFORD, J. B. 2012. RFX2 is broadly required for ciliogenesis during vertebrate development. *Dev Biol,* 363**,** 155-65.

27.) COHEN, S. N., CHANG, A. C. & HSU, L. 1972. Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA. *Proc Natl Acad Sci U S A,* 69**,** 2110-4.

28.) COLEGRAVE, N. 2012. The evolutionary success of sex. Science & Society Series on Sex and Science. *EMBO Rep,* 13**,** 774-8.

29.) COLLINS, F. S., GREEN, E. D., GUTTMACHER, A. E., GUYER, M. S. & INSTITUTE, U. S. N. H. G. R. 2003. A vision for the future of genomics research. *Nature,* 422**,** 835-47.

30.) CONDEMINE, W., TAKAHASHI, Y., ZHU, J., PUVION-DUTILLEUL, F., GUEGAN, S., JANIN, A. & DE THE, H. 2006. Characterization of endogenous human promyelocytic leukemia isoforms. *Cancer Res,* 66**,** 6192-8.

31.) CONSORTIUM, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature,* 489**,** 57-74.

32.) COOPER, C. L., BERRIER, A. L., ROMAN, C. & CALAME, K. L. 1994. Limited expression of C/EBP family proteins during B lymphocyte development. Negative regulator Ig/EBP predominates early and activator NF-IL-6 is induced later. *J Immunol,* 153**,** 5049-58.

33.) DAHM, R. 2005. Friedrich Miescher and the discovery of DNA. *Dev Biol,* 278**,** 274-88.

34.) DAVEY, R. A. & GROSSMANN, M. 2016. Androgen Receptor Structure, Function and Biology: From Bench to Bedside. *Clin Biochem Rev,* 37**,** 3-15.

35.) DE SANCTIS, D., PESCE, A., NARDINI, M., BOLOGNESI, M., BOCEDI, A. & ASCENZI, P. 2004. Structure-function relationships in the growing hexa-coordinate hemoglobin sub-family. *IUBMB Life,* 56**,** 643-51.

36.) DEKOTER, R. P., LEE, H. J. & SINGH, H. 2002. PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. *Immunity,* 16**,** 297-309.

37.) DEKOTER, R. P. & SINGH, H. 2000. Regulation of B lymphocyte and macrophage development by graded expression of PU.1. *Science,* 288**,** 1439-41.

38.) DELOGU, A., SCHEBESTA, A., SUN, Q., ASCHENBRENNER, K., PERLOT, T. & BUSSLINGER, M. 2006. Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity,* 24**,** 269-81.

39.) DENG, W. & LIN, H. 2002. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell,* 2**,** 819-30.

40.) DERJUGA, A., GOURLEY, T. S., HOLM, T. M., HENG, H. H., SHIVDASANI, R. A., AHMED, R., ANDREWS, N. C. & BLANK, V. 2004. Complexity of CNC transcription factors as revealed by gene targeting of the Nrf3 locus. *Mol Cell Biol,* 24**,** 3286-94.

41.) ELLIS, T. M., MOSER, M. T., LE, P. T., FLANIGAN, R. C. & KWON, E. D. 2001. Alterations in peripheral B cells and B cell progenitors following androgen ablation in mice. *Int Immunol,* 13**,** 553-8.

42.) ERNST, C., ODOM, D. T. & KUTTER, C. 2017. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nat Commun,* 8**,** 1411.

43.) FAISAL, I., CISNEROS-MONTALVO, S., HAMER, G., TUOMINEN, M. M., LAURILA, P. P., TUMIATI, M., JAUHIAINEN, M., KOTAJA, N., TOPPARI, J., MAKELA, J. A. & KAUPPI, L. 2019. Transcription Factor USF1 Is Required for Maintenance of Germline Stem Cells in Male Mice. *Endocrinology,* 160**,** 1119-1136.

44.) FISHILEVICH, S., NUDEL, R., RAPPAPORT, N., HADAR, R., PLASCHKES, I., INY STEIN, T., ROSEN, N., KOHN, A., TWIK, M., SAFRAN, M., LANCET, D. & COHEN, D. 2017. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford),* 2017.

45.) FITZSIMMONS, D., LUKIN, K., LUTZ, R., GARVIE, C. W., WOLBERGER, C. & HAGMAN, J. 2009. Highly cooperative recruitment of Ets-1 and release of autoinhibition by Pax5. *J Mol Biol,* 392**,** 452-64.

46.) FLORES-MORALES, A., BERGMANN, T. B., LAVALLEE, C., BATTH, T. S., LIN, D., LERDRUP, M., FRIIS, S., BARTELS, A., KRISTENSEN, G., KRZYZANOWSKA, A., XUE, H., FAZLI, L., HANSEN, K. H., RODER, M. A., BRASSO, K., MOREIRA, J. M., BJARTELL, A., WANG, Y., OLSEN, J. V., COLLINS, C. C. & IGLESIAS-GATO, D. 2019. Proteogenomic Characterization of Patient-Derived Xenografts Highlights the Role of REST in Neuroendocrine Differentiation of Castration-Resistant Prostate Cancer. *Clin Cancer Res,* 25**,** 595-608.

47.) FUCHS, C., BURMESTER, T. & HANKELN, T. 2006. The amphibian globin gene repertoire as revealed by the Xenopus genome. *Cytogenet Genome Res,* 112**,** 296-306.

48.) FUJITA, K. & NONOMURA, N. 2019. Role of Androgen Receptor in Prostate Cancer: A Review. *World J Mens Health,* 37**,** 288-295.

49.) FUJIWARA, T., O'GEEN, H., KELES, S., BLAHNIK, K., LINNEMANN, A. K., KANG, Y. A., CHOI, K., FARNHAM, P. J. & BRESNICK, E. H. 2009. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell,* 36**,** 667-81.

50.) GANAPATHI, K. A., TOWNSLEY, D. M., HSU, A. P., ARTHUR, D. C., ZERBE, C. S., CUELLAR-RODRIGUEZ, J., HICKSTEIN, D. D., ROSENZWEIG, S. D., BRAYLAN, R. C., YOUNG, N. S., HOLLAND, S. M. & CALVO, K. R. 2015. GATA2 deficiency-associated bone marrow disorder differs from idiopathic aplastic anemia. *Blood,* 125**,** 56-70.

51.) GOODENOUGH, U. & HEITMAN, J. 2014. Origins of eukaryotic sexual reproduction. *Cold Spring Harb Perspect Biol,* 6.

52.) GROOTEGOED, J. A., SIEP, M. & BAARENDS, W. M. 2000. Molecular and cellular mechanisms in spermatogenesis. *Baillieres Best Pract Res Clin Endocrinol Metab,* 14**,** 331-43.

53.) GUNES, S. & KULAC, T. 2013. The role of epigenetics in spermatogenesis. *Turk J Urol,* 39**,** 181-7.

54.) HAGMAN, J. 2015. Chapter 3. Transcriptional Regulation of Early B Cell Development. 10.1016/B978-0-12-397933-9.00003-5.

55.) HAGMAN, J., BELANGER, C., TRAVIS, A., TURCK, C. W. & GROSSCHEDL, R. 1993. Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. *Genes Dev,* 7**,** 760-73.

56.) HAGMAN, J. & LUKIN, K. 2006. Transcription factors drive B cell development. *Curr Opin Immunol,* 18**,** 127-34.

57.) HAGMAN, J., RAMIREZ, J. & LUKIN, K. 2012. B lymphocyte lineage specification, commitment and epigenetic control of transcription by early B cell factor 1. *Curr Top Microbiol Immunol,* 356**,** 17-38.

58.) HARDY, R. R., KINCADE, P. W. & DORSHKIND, K. 2007. The protean nature of cells in the B lymphocyte lineage. *Immunity,* 26**,** 703-14.

59.) HARIGAE, H., OKITSU, Y., YOKOYAMA, H., FUJIWARA, T., INOMATA, M., TAKAHASHI, S., MINEGISHI, N., KAKU, M. & SASAKI, T. 2006. Induction of erythroid-specific genes by overexpression of GATA-2 in K562 cells. *Int J Hematol,* 84**,** 38-42.

60.) HEINLEIN, C. A. & CHANG, C. 2002. Androgen receptor (AR) coregulators: an overview. *Endocr Rev,* 23**,** 175-200.

61.) HEINLEIN, C. A. & CHANG, C. 2004. Androgen receptor in prostate cancer. *Endocr Rev,* 25**,** 276-308.

62.) HIORT, O., HOLTERHUS, P. M., HORTER, T., SCHULZE, W., KREMKE, B., BALS-PRATSCH, M., SINNECKER, G. H. & KRUSE, K. 2000. Significance of mutations in the androgen receptor gene in males with idiopathic infertility. *J Clin Endocrinol Metab,* 85**,** 2810-5.

63.) HIRSH, A. E. & FRASER, H. B. 2001. Protein dispensability and rate of evolution. *Nature,* 411**,** 1046-9.

64.) HOLM, L. & SANDER, C. 1993. Structural alignment of globins, phycocyanins and colicin A. *FEBS Lett,* 315**,** 301-6.

65.) HOOD, L. & ROWEN, L. 2013. The Human Genome Project: big science transforms biology and medicine. *Genome Med,* 5**,** 79.

66.) HOOGEWIJS, D., DEWILDE, S., VIERSTRAETE, A., MOENS, L. & VINOGRADOV, S. N. 2012a. A phylogenetic analysis of the globins in fungi. *PLoS One,* 7**,** e31856.

67.) HOOGEWIJS, D., EBNER, B., GERMANI, F., HOFFMANN, F. G., FABRIZIUS, A., MOENS, L., BURMESTER, T., DEWILDE, S., STORZ, J. F., VINOGRADOV, S. N. & HANKELN, T. 2012b. Androglobin: a chimeric globin in metazoans that is preferentially expressed in Mammalian testes. *Mol Biol Evol,* 29**,** 1105-14.

68.) HOOGEWIJS, D., HOUTHOOFD, K., MATTHIJSSENS, F., VANDESOMPELE, J. & VANFLETEREN, J. R. 2008. Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in C. elegans. *BMC Mol Biol,* 9**,** 9.

69.) HORCHER, M., SOUABNI, A. & BUSSLINGER, M. 2001. Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity,* 14**,** 779-90.

70.) HORVATH, G. C., KISTLER, W. S. & KISTLER, M. K. 2004. RFX2 is a potential transcriptional regulatory factor for histone H1t and other genes expressed during the meiotic phase of spermatogenesis. *Biol Reprod,* 71**,** 1551-9.

71.) HOSOYA, T., CLIFFORD, M., LOSSON, R., TANABE, O. & ENGEL, J. D. 2013. TRIM28 is essential for erythroblast differentiation in the mouse. *Blood,* 122**,** 3798-807.

72.) HOSOYA, T., KUROHA, T., MORIGUCHI, T., CUMMINGS, D., MAILLARD, I., LIM, K. C. & ENGEL, J. D. 2009. GATA-3 is required for early T lineage progenitor development. *J Exp Med,* 206**,** 2987-3000.

73.) HOU, Y., YUAN, J., ZHOU, X., FU, X., CHENG, H. & ZHOU, R. 2012. DNA demethylation and USF regulate the meiosis-specific expression of the mouse Miwi. *PLoS Genet,* 8**,** e1002716.

74.) HSU, K. S. & KAO, H. Y. 2018. PML: Regulation and multifaceted function beyond tumor suppression. *Cell Biosci,* 8**,** 5.

75.) IDEKER, T., GALITSKI, T. & HOOD, L. 2001. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet,* 2**,** 343-72.

76.) IKONOMI, P., NOGUCHI, C. T., MILLER, W., KASSAHUN, H., HARDISON, R. & SCHECHTER, A. N. 2000. Levels of GATA-1/GATA-2 transcription factors modulate expression of embryonic and fetal hemoglobins. *Gene,* 261**,** 277-87.

77.) ISHITANI, K., YOSHIDA, T., KITAGAWA, H., OHTA, H., NOZAWA, S. & KATO, S. 2003. p54nrb acts as a transcriptional coactivator for activation function 1 of the human androgen receptor. *Biochem Biophys Res Commun,* 306**,** 660-5.

78.) ITSUMI, M., SHIOTA, M., YOKOMIZO, A., KASHIWAGI, E., TAKEUCHI, A., TATSUGAMI, K., INOKUCHI, J., SONG, Y., UCHIUMI, T. & NAITO, S. 2013. Human heterochromatin protein 1 isoforms regulate androgen receptor signaling in prostate cancer. *J Mol Endocrinol,* 50**,** 401-9.

79.) JIANG, Y. Y., KONG, D. X., QIN, T. & ZHANG, H. Y. 2010. How does oxygen rise drive evolution? Clues from oxygen-dependent biosynthesis of nuclear receptor ligands. *Biochem Biophys Res Commun,* 391**,** 1158-60.

80.) KAFTANOVSKAYA, E. M., HUANG, Z., BARBARA, A. M., DE GENDT, K., VERHOEVEN, G., GORLOV, I. P. & AGOULNIK, A. I. 2012. Cryptorchidism in mice with an androgen receptor ablation in gubernaculum testis. *Mol Endocrinol,* 26**,** 598-607.

81.) KAKAR, S., HOFFMAN, F. G., STORZ, J. F., FABIAN, M. & HARGROVE, M. S. 2010. Structure and reactivity of hexacoordinate hemoglobins. *Biophys Chem,* 152**,** 1-14.

82.) KASTING, J. F. 2001. Earth history. The rise of atmospheric oxygen. *Science,* 293**,** 819-20.

83.) KATSUMURA, K. R., BRESNICK, E. H. & GROUP, G. F. M. 2017. The GATA factor revolution in hematology. *Blood,* 129**,** 2092-2102.

84.) KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

85.) KEPKAY, R., ATTWOOD, K. M., ZIV, Y., SHILOH, Y. & DELLAIRE, G. 2011. KAP1 depletion increases PML nuclear body number in concert with ultrastructural changes in chromatin. *Cell Cycle,* 10**,** 308-22.

86.) KIM, M. H. & FIELDS, J. 2008. Translationally regulated C/EBP beta isoform expression upregulates metastatic genes in hormone-independent prostate cancer cells. *Prostate,* 68**,** 1362-71.

87.) KIM, T. K. & SHIEKHATTAR, R. 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell,* 162**,** 948-59.

88.) KITAJIMA, K., TANAKA, M., ZHENG, J., YEN, H., SATO, A., SUGIYAMA, D., UMEHARA, H., SAKAI, E. & NAKANO, T. 2006. Redirecting differentiation of hematopoietic progenitors by a transcription factor, GATA-2. *Blood,* 107**,** 1857-63.

89.) KNOLL, A. H. & CARROLL, S. B. 1999. Early animal evolution: emerging views from comparative biology and geology. *Science,* 284**,** 2129-37.

90.) KO, L. J. & ENGEL, J. D. 1993. DNA-binding specificities of the GATA transcription factor family. *Mol Cell Biol,* 13**,** 4011-22.

91.) KOBAYASHI, A., ITO, E., TOKI, T., KOGAME, K., TAKAHASHI, S., IGARASHI, K., HAYASHI, N. & YAMAMOTO, M. 1999. Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3. *J Biol Chem,* 274**,** 6443-52.

92.) KOLOVOS, P., KNOCH, T. A., GROSVELD, F. G., COOK, P. R. & PAPANTONIS, A. 2012. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin,* 5**,** 1.

93.) KRISTIANSEN, G., HU, J., WICHMANN, D., STIEHL, D. P., ROSE, M., GERHARDT, J., BOHNERT, A., TEN HAAF, A., MOCH, H., RALEIGH, J., VARIA, M. A., SUBARSKY, P., SCANDURRA, F. M., GNAIGER, E., GLEIXNER, E., BICKER, A., GASSMANN, M., HANKELN, T., DAHL, E. & GORR, T. A. 2011. Endogenous myoglobin in breast cancer is hypoxia-

inducible by alternative transcription and functions to impair mitochondrial activity: a role in tumor suppression? *J Biol Chem,* 286**,** 43417-28.

94.) KUGELSTADT, D., HABERKAMP, M., HANKELN, T. & BURMESTER, T. 2004. Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. *Biochem Biophys Res Commun,* 325**,** 719-25.

95.) KUMP, L. R. 2008. The rise of atmospheric oxygen. *Nature,* 451**,** 277-8.

96.) LALLEMAND-BREITENBACH, V. & DE THE, H. 2010. PML nuclear bodies. *Cold Spring Harb Perspect Biol,* 2**,** a000661.

97.) LANDER, E. S., LINTON, L. M., BIRREN, B., NUSBAUM, C., ZODY, M. C., BALDWIN, J., DEVON, K., DEWAR, K., DOYLE, M., FITZHUGH, W., FUNKE, R., GAGE, D., HARRIS, K., HEAFORD, A., HOWLAND, J., KANN, L., LEHOCZKY, J., LEVINE, R., MCEWAN, P., MCKERNAN, K., MELDRIM, J., MESIROV, J. P., MIRANDA, C., MORRIS, W., NAYLOR, J., RAYMOND, C., ROSETTI, M., SANTOS, R., SHERIDAN, A., SOUGNEZ, C., STANGE-THOMANN, Y., STOJANOVIC, N., SUBRAMANIAN, A., WYMAN, D., ROGERS, J., SULSTON, J., AINSCOUGH, R., BECK, S., BENTLEY, D., BURTON, J., CLEE, C., CARTER, N., COULSON, A., DEADMAN, R., DELOUKAS, P., DUNHAM, A., DUNHAM, I., DURBIN, R., FRENCH, L., GRAFHAM, D., GREGORY, S., HUBBARD, T., HUMPHRAY, S., HUNT, A., JONES, M., LLOYD, C., MCMURRAY, A., MATTHEWS, L., MERCER, S., MILNE, S., MULLIKIN, J. C., MUNGALL, A., PLUMB, R., ROSS, M., SHOWNKEEN, R., SIMS, S., WATERSTON, R. H., WILSON, R. K., HILLIER, L. W., MCPHERSON, J. D., MARRA, M. A., MARDIS, E. R., FULTON, L. A., CHINWALLA, A. T., PEPIN, K. H., GISH, W. R., CHISSOE, S. L., WENDL, M. C., DELEHAUNTY, K. D., MINER, T. L., DELEHAUNTY, A., KRAMER, J. B., COOK, L. L., FULTON, R. S., JOHNSON, D. L., MINX, P. J., CLIFTON, S. W., HAWKINS, T., BRANSCOMB, E., PREDKI, P., RICHARDSON, P., WENNING, S., SLEZAK, T., DOGGETT, N., CHENG, J. F., OLSEN, A., LUCAS, S., ELKIN, C., UBERBACHER, E., FRAZIER, M., et al. 2001. Initial sequencing and analysis of the human genome. *Nature,* 409**,** 860-921.

98.) LARDENOIS, A., CHALMEL, F., BARRIONUEVO, F., DEMOUGIN, P., SCHERER, G. & PRIMIG, M. 2010. Profiling spermatogenic failure in adult testes bearing Sox9-deficient Sertoli cells identifies genes involved in feminization, inflammation and stress. *Reprod Biol Endocrinol,* 8**,** 154.

99.) LATCHMAN, D. S. 1993. Transcription factors: an overview. *Int J Exp Pathol,* 74**,** 417-22.

100.) LIANG, H., STUDACH, L., HULLINGER, R. L., XIE, J. & ANDRISANI, O. M. 2014. Down-regulation of RE-1 silencing transcription factor (REST) in advanced prostate cancer by hypoxia-induced miR-106b~25. *Exp Cell Res,* 320**,** 188-99.

101.) LIN, J., LIU, W., LUAN, T., YUAN, L., JIANG, W., CAI, H., YUAN, W., WANG, Y., ZHANG, Q. & WANG, L. 2017. High expression of PU.1 is associated with Her-2 and shorter survival in patients with breast cancer. *Oncol Lett,* 14**,** 8220-8226.

102.) LINNEMANN, A. K., O'GEEN, H., KELES, S., FARNHAM, P. J. & BRESNICK, E. H. 2011. Genetic framework for GATA factor function in vascular biology. *Proc Natl Acad Sci U S A,* 108**,** 13641-6.

103.) LIU, S., CHAN, H. L., BAI, F., MA, J., SCOTT, A., ROBBINS, D. J., CAPOBIANCO, A. J., ZHU, P. & PEI, X. H. 2016. Gata3 restrains B cell proliferation and cooperates with p18INK4c to repress B cell lymphomagenesis. *Oncotarget,* 7**,** 64007-64020.

104.) MACARI, E. R. & LOWREY, C. H. 2011. Induction of human fetal hemoglobin via the NRF2 antioxidant response signaling pathway. *Blood,* 117**,** 5987-97.

105.) MATSUMOTO, K., KWON, O. Y., KIM, H. & AKAO, Y. 2005. Expression of rck/p54, a DEAD-box RNA helicase, in gametogenesis and early embryogenesis of mice. *Dev Dyn,* 233**,** 1149-56.

106.) MAXAM, A. M. & GILBERT, W. 1977. A new method for sequencing DNA. *Proc Natl Acad Sci U S A,* 74**,** 560-4.

107.) MEDINA, K. L. & SINGH, H. 2005. Genetic networks that regulate B lymphopoiesis. *Curr Opin Hematol,* 12**,** 203-9.

108.) MOLKENTIN, J. D. 2000. The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *J Biol Chem,* 275**,** 38949-52.

109.) NAKAMURA, B. N., LAWSON, G., CHAN, J. Y., BANUELOS, J., CORTES, M. M., HOANG, Y. D., ORTIZ, L., RAU, B. A. & LUDERER, U. 2010. Knockout of the transcription factor NRF2 disrupts spermatogenesis in an age-dependent manner. *Free Radic Biol Med,* 49**,** 1368-79.

110.) NGUYEN, T., SHERRATT, P. J. & PICKETT, C. B. 2003. Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu Rev Pharmacol Toxicol,* 43**,** 233-60.

111.) NISHIOKA, N., YAMAMOTO, S., KIYONARI, H., SATO, H., SAWADA, A., OTA, M., NAKAO, K. & SASAKI, H. 2008. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev,* 125**,** 270-83.

112.) NOVAKOVA, M., ZALIOVA, M., SUKOVA, M., WLODARSKI, M., JANDA, A., FRONKOVA, E., CAMPR, V., LEJHANCOVA, K., ZAPLETAL, O., POSPISILOVA, D., CERNA, Z., KUHN, T., SVEC, P., PELKOVA, V., ZEMANOVA, Z., KERNDRUP, G., VAN DEN HEUVEL-EIBRINK, M., VAN DER VELDEN, V., NIEMEYER, C., KALINA, T., TRKA, J., STARY, J., HRUSAK, O. & MEJSTRIKOVA, E. 2016. Loss of B cells and their precursors is the most constant feature of GATA-2 deficiency in childhood myelodysplastic syndrome. *Haematologica,* 101**,** 707-16.

113.) O'HARA, L. & SMITH, L. B. 2015. Androgen receptor roles in spermatogenesis and infertility. *Best Pract Res Clin Endocrinol Metab,* 29**,** 595-605.

114.) O'RIORDAN, M. & GROSSCHEDL, R. 1999. Coordinate regulation of B cell differentiation by the transcription factors EBF and E2A. *Immunity,* 11**,** 21-31.

115.) OLIVE, V., WAGNER, N., CHAN, S., KASTNER, P., VANNETTI, C., CUZIN, F. & RASSOULZADEGAN, M. 2007. PU.1 (Sfpi1), a pleiotropic regulator expressed from the first embryonic stages with a crucial function in germinal progenitors. *Development,* 134**,** 3815-25.

116.) OLSEN, N. J. & KOVACS, W. J. 2001. Effects of androgens on T and B lymphocyte development. *Immunol Res,* 23**,** 281-8.

117.) ORTH, J. M., GUNSALUS, G. L. & LAMPERTI, A. A. 1988. Evidence from Sertoli cell-depleted rats indicates that spermatid number in adults depends on numbers of Sertoli cells produced during perinatal development. *Endocrinology,* 122**,** 787-94.

118.) PAL, R., JANZ, M., GALSON, D. L., GRIES, M., LI, S., JOHRENS, K., ANAGNOSTOPOULOS, I., DORKEN, B., MAPARA, M. Y., BORGHESI, L., KARDAVA, L., ROODMAN, G. D., MILCAREK, C. & LENTZSCH, S. 2009. C/EBPbeta regulates transcription factors critical for proliferation and survival of multiple myeloma cells. *Blood,* 114**,** 3890-8.

119.) PAL, S., CANTOR, A. B., JOHNSON, K. D., MORAN, T. B., BOYER, M. E., ORKIN, S. H. & BRESNICK, E. H. 2004. Coregulator-dependent facilitation of chromatin occupancy by GATA-1. *Proc Natl Acad Sci U S A,* 101**,** 980-5.

120.) PANG, S. H. M., DE GRAAF, C. A., HILTON, D. J., HUNTINGTON, N. D., CAROTTA, S., WU, L. & NUTT, S. L. 2018. PU.1 Is Required for the Developmental Progression of Multipotent Progenitors to Common Lymphoid Progenitors. *Front Immunol,* 9**,** 1264.

121.) PAYNE, J. L., MCCLAIN, C. R., BOYER, A. G., BROWN, J. H., FINNEGAN, S., KOWALEWSKI, M., KRAUSE, R. A., JR., LYONS, S. K., MCSHEA, D. W., NOVACK-GOTTSHALL, P. M., SMITH, F. A., SPAETH, P., STEMPIEN, J. A. & WANG, S. C. 2011. The evolutionary consequences of oxygenic photosynthesis: a body size perspective. *Photosynth Res,* 107**,** 37-57.

122.) PERRIN, B. J. & HUTTENLOCHER, A. 2002. Calpain. *Int J Biochem Cell Biol,* 34**,** 722-5.

123.) PERUTZ, M. F. 1979. Regulation of oxygen affinity of hemoglobin: influence of structure of the globin on the heme iron. *Annu Rev Biochem,* 48**,** 327-86.

124.) PESCE, A., BOLOGNESI, M., BOCEDI, A., ASCENZI, P., DEWILDE, S., MOENS, L., HANKELN, T. & BURMESTER, T. 2002. Neuroglobin and cytoglobin. Fresh blood for the vertebrate globin family. *EMBO Rep,* 3**,** 1146-51.

125.) POBBATI, A. V. & HONG, W. 2013. Emerging roles of TEAD transcription factors and its coactivators in cancers. *Cancer Biol Ther,* 14**,** 390-8.

126.) RALSTON, A., COX, B. J., NISHIOKA, N., SASAKI, H., CHEA, E., RUGG-GUNN, P., GUO, G., ROBSON, P., DRAPER, J. S. & ROSSANT, J. 2010. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development,* 137**,** 395-403.

127.) RANA, K., DAVEY, R. A. & ZAJAC, J. D. 2014. Human androgen deficiency: insights gained from androgen receptor knockout mouse models. *Asian J Androl,* 16**,** 169-77.

128.) RANDI, E. B., VERVAET, B., TSACHAKI, M., PORTO, E., VERMEYLEN, S., LINDENMEYER, M. T., THUY, L. T. T., COHEN, C. D., DEVUYST, O., KISTLER, A. D., SZABO, C., KAWADA, N., HANKELN, T., ODERMATT, A., DEWILDE, S., WENGER, R. H. & HOOGEWIJS, D. 2020. The Antioxidative Role of Cytoglobin in Podocytes: Implications for a Role in Chronic Kidney Disease. *Antioxid Redox Signal,* 32**,** 1155-1171.

129.) RAY, S., DUTTA, D., RUMI, M. A., KENT, L. N., SOARES, M. J. & PAUL, S. 2009. Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. *J Biol Chem,* 284**,** 4978-88.

130.) REITH, W., SATOLA, S., SANCHEZ, C. H., AMALDI, I., LISOWSKA-GROSPIERRE, B., GRISCELLI, C., HADAM, M. R. & MACH, B. 1988. Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. *Cell,* 53**,** 897-906.

131.) ROADMAP EPIGENOMICS, C., KUNDAJE, A., MEULEMAN, W., ERNST, J., BILENKY, M., YEN, A., HERAVI-MOUSSAVI, A., KHERADPOUR, P., ZHANG, Z., WANG, J., ZILLER, M. J., AMIN, V., WHITAKER, J. W., SCHULTZ, M. D., WARD, L. D., SARKAR, A., QUON, G., SANDSTROM, R. S., EATON, M. L., WU, Y. C., PFENNING, A. R., WANG, X., CLAUSSNITZER, M., LIU, Y., COARFA, C., HARRIS, R. A., SHORESH, N., EPSTEIN, C. B., GJONESKA, E., LEUNG, D., XIE, W., HAWKINS, R. D., LISTER, R., HONG, C., GASCARD, P., MUNGALL, A. J., MOORE, R., CHUAH, E., TAM, A., CANFIELD, T. K., HANSEN, R. S., KAUL, R., SABO, P. J., BANSAL, M. S., CARLES, A., DIXON, J. R., FARH, K. H., FEIZI, S.,

KARLIC, R., KIM, A. R., KULKARNI, A., LI, D., LOWDON, R., ELLIOTT, G., MERCER, T. R., NEPH, S. J., ONUCHIC, V., POLAK, P., RAJAGOPAL, N., RAY, P., SALLARI, R. C., SIEBENTHALL, K. T., SINNOTT-ARMSTRONG, N. A., STEVENS, M., THURMAN, R. E., WU, J., ZHANG, B., ZHOU, X., BEAUDET, A. E., BOYER, L. A., DE JAGER, P. L., FARNHAM, P. J., FISHER, S. J., HAUSSLER, D., JONES, S. J., LI, W., MARRA, M. A., MCMANUS, M. T., SUNYAEV, S., THOMSON, J. A., TLSTY, T. D., TSAI, L. H., WANG, W., WATERLAND, R. A., ZHANG, M. Q., CHADWICK, L. H., BERNSTEIN, B. E., COSTELLO, J. F., ECKER, J. R., HIRST, M., MEISSNER, A., MILOSAVLJEVIC, A., REN, B., STAMATOYANNOPOULOS, J. A., WANG, T. & KELLIS, M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature,* 518**,** 317-30.

132.) ROESNER, A., FUCHS, C., HANKELN, T. & BURMESTER, T. 2005. A globin gene of ancient evolutionary origin in lower vertebrates: evidence for two distinct globin families in animals. *Mol Biol Evol,* 22**,** 12-20.

133.) ROSMARIN, A. G., RESENDES, K. K., YANG, Z., MCMILLAN, J. N. & FLEMING, S. L. 2004. 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.

134.) SANDELIN, A., ALKEMA, W., ENGSTROM, P., WASSERMAN, W. W. & LENHARD, B. 2004. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res,* 32**,** D91-4.

135.) SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

136.) SANGER, F., THOMPSON, E. O. & KITAI, R. 1955. The amide groups of insulin. *Biochem J,* 59**,** 509-18.

137.) SANGOI, A. R., SHRESTHA, B., YANG, G., MEGO, O. & BECK, A. H. 2016. The Novel Marker GATA3 is Significantly More Sensitive Than Traditional Markers Mammaglobin and GCDFP15 for Identifying Breast Cancer in Surgical and Cytology Specimens of Metastatic and Matched Primary Tumors. *Appl Immunohistochem Mol Morphol,* 24**,** 229-37.

138.) SANKARANARAYANAN, K. & JAISWAL, A. K. 2004. Nrf3 negatively regulates antioxidant-response element-mediated expression and antioxidant induction of NAD(P)H:quinone oxidoreductase1 gene. *J Biol Chem,* 279**,** 50810-7.

139.) SANTONI DE SIO, F. R., MASSACAND, J., BARDE, I., OFFNER, S., CORSINOTTI, A., KAPOPOULOU, A., BOJKOWSKA, K., DAGKLIS, A., FERNANDEZ, M., GHIA, P., THOMAS, J. H., PINSCHEWER, D., HARRIS, N. & TRONO, D. 2012. KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. *Blood,* 119**,** 4675-85.

140.) SASSONE-CORSI, P. 1997. Transcriptional checkpoints determining the fate of male germ cells. *Cell,* 88**,** 163-6.

141.) SCHULTZ, M. A., HAGAN, S. S., DATTA, A., ZHANG, Y., FREEMAN, M. L., SIKKA, S. C., ABDEL-MAGEED, A. B. & MONDAL, D. 2014. Nrf1 and Nrf2 transcription factors regulate androgen receptor transactivation in prostate cancer cells. *PLoS One,* 9**,** e87204.

142.) SHARMA, N. L., MASSIE, C. E., BUTTER, F., MANN, M., BON, H., RAMOS-MONTOYA, A., MENON, S., STARK, R., LAMB, A. D., SCOTT, H. E., WARREN, A. Y., NEAL, D. E. & MILLS, I. G. 2014. The ETS family member GABPalpha modulates androgen receptor signalling and mediates an aggressive phenotype in prostate cancer. *Nucleic Acids Res,* 42**,** 6256-69.

143.) SHAWLOT, W., VAZQUEZ-CHANTADA, M., WALLINGFORD, J. B. & FINNELL, R. H. 2015. Rfx2 is required for spermatogenesis in the mouse. *Genesis,* 53**,** 604-611.

144.) SLEZAK, J., TRUONG, M., HUANG, W. & JARRARD, D. 2013. HP1gamma expression is elevated in prostate cancer and is superior to Gleason score as a predictor of biochemical recurrence after radical prostatectomy. *BMC Cancer,* 13**,** 148.

145.) SMITHSON, G., COUSE, J. F., LUBAHN, D. B., KORACH, K. S. & KINCADE, P. W. 1998. The role of estrogen receptors and androgen receptors in sex steroid regulation of B lymphopoiesis. *J Immunol,* 161**,** 27-34.

146.) SOMASUNDARAM, R., PRASAD, M. A., UNGERBACK, J. & SIGVARDSSON, M. 2015. Transcription factor networks in B-cell

differentiation link development to acute lymphoid leukemia. *Blood,* 126**,** 144-52.

147.) SPERLING, E. A., FRIEDER, C. A., RAMAN, A. V., GIRGUIS, P. R., LEVIN, L. A. & KNOLL, A. H. 2013. Oxygen, ecology, and the Cambrian radiation of animals. *Proc Natl Acad Sci U S A,* 110**,** 13446-51.

148.) SUÁREZ ALONSO, M. 2019. Exploring the regulation and molecular function of androglobin.

149.) SVENSSON, C., CEDER, J., IGLESIAS-GATO, D., CHUAN, Y. C., PANG, S. T., BJARTELL, A., MARTINEZ, R. M., BOTT, L., HELCZYNSKI, L., ULMERT, D., WANG, Y., NIU, Y., COLLINS, C. & FLORES-MORALES, A. 2014. REST mediates androgen receptor actions on gene repression and predicts early recurrence of prostate cancer. *Nucleic Acids Res,* 42**,** 999-1015.

150.) TERRIS, B., BALDIN, V., DUBOIS, S., DEGOTT, C., FLEJOU, J. F., HENIN, D. & DEJEAN, A. 1995. PML nuclear bodies are general targets for inflammation and cell proliferation. *Cancer Res,* 55**,** 1590-7.

151.) TEVOSIAN, S. G. 2014. Transgenic mouse models in the study of reproduction: insights into GATA protein function. *Reproduction,* 148**,** R1-R14.

152.) TEVOSIAN, S. G., ALBRECHT, K. H., CRISPINO, J. D., FUJIWARA, Y., EICHER, E. M. & ORKIN, S. H. 2002. Gonadal differentiation, sex determination and normal Sry expression in mice require direct interaction between transcription partners GATA4 and FOG2. *Development,* 129**,** 4627-34.

153.) THURMAN, R. E., RYNES, E., HUMBERT, R., VIERSTRA, J., MAURANO, M. T., HAUGEN, E., SHEFFIELD, N. C., STERGACHIS, A. B., WANG, H., VERNOT, B., GARG, K., JOHN, S., SANDSTROM, R., BATES, D., BOATMAN, L., CANFIELD, T. K., DIEGEL, M., DUNN, D., EBERSOL, A. K., FRUM, T., GISTE, E., JOHNSON, A. K., JOHNSON, E. M., KUTYAVIN, T., LAJOIE, B., LEE, B. K., LEE, K., LONDON, D., LOTAKIS, D., NEPH, S., NERI, F., NGUYEN, E. D., QU, H., REYNOLDS, A. P., ROACH, V., SAFI, A., SANCHEZ, M. E., SANYAL, A., SHAFER, A., SIMON, J. M., SONG, L., VONG, S., WEAVER, M., YAN, Y., ZHANG, Z., ZHANG, Z., LENHARD, B., TEWARI, M., DORSCHNER, M. O., HANSEN, R. S., NAVAS, P. A., STAMATOYANNOPOULOS, G., IYER, V. R., LIEB, J. D., SUNYAEV, S. R.,

AKEY, J. M., SABO, P. J., KAUL, R., FUREY, T. S., DEKKER, J., CRAWFORD, G. E. & STAMATOYANNOPOULOS, J. A. 2012. The accessible chromatin landscape of the human genome. *Nature,* 489**,** 75-82.

154.) TING, C. N., OLSON, M. C., BARTON, K. P. & LEIDEN, J. M. 1996. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature,* 384**,** 474-8.

155.) TRAISH, A., BOLANOS, J., NAIR, S., SAAD, F. & MORGENTALER, A. 2018. Do Androgens Modulate the Pathophysiological Pathways of Inflammation? Appraising the Contemporary Evidence. *J Clin Med,* 7.

156.) TSAI, F. Y. & ORKIN, S. H. 1997. Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood,* 89**,** 3636-43.

157.) TSAPOGAS, P., ZANDI, S., AHSBERG, J., ZETTERBLAD, J., WELINDER, E., JONSSON, J. I., MANSSON, R., QIAN, H. & SIGVARDSSON, M. 2011. IL-7 mediates Ebf-1-dependent lineage restriction in early lymphoid progenitors. *Blood,* 118**,** 1283-90.

158.) TSUZUKI, S., TOWATARI, M., SAITO, H. & ENVER, T. 2000. Potentiation of GATA-2 activity through interactions with the promyelocytic leukemia protein (PML) and the t(15;17)-generated PML-retinoic acid receptor alpha oncoprotein. *Mol Cell Biol,* 20**,** 6276-86.

159.) UMAZUME, A., KEZUKA, T., MATSUDA, R., USUI, Y., TAKAHASHI, H., YAMAKAWA, N., YASHIRO, T., NISHIYAMA, C. & GOTO, H. 2018. Role of PU.1 Expression as an Inflammatory Marker in Experimental Autoimmune Uveoretinitis. *Ocul Immunol Inflamm,* 26**,** 951-963.

160.) VAN ACKER, Z. P., DECLERCK, K., LUYCKX, E., VANDEN BERGHE, W. & DEWILDE, S. 2019. Non-Methylation-Linked Mechanism of REST-Induced Neuroglobin Expression Impacts Mitochondrial Phenotypes in a Mouse Model of Amyotrophic Lateral Sclerosis. *Neuroscience,* 412**,** 233-247.

161.) VINOGRADOV, S. N., BAILLY, X., SMITH, D. R., TINAJERO-TREJO, M., POOLE, R. K. & HOOGEWIJS, D. 2013a. Microbial eukaryote globins. *Adv Microb Physiol,* 63**,** 391-446.

162.) VINOGRADOV, S. N., FERNANDEZ, I., HOOGEWIJS, D. & ARREDONDO-PETER, R. 2011. Phylogenetic relationships of 3/3 and 2/2 hemoglobins in Archaeplastida genomes to bacterial and other eukaryote hemoglobins. *Mol Plant,* 4**,** 42-58.

163.) VINOGRADOV, S. N., HOOGEWIJS, D., BAILLY, X., ARREDONDO-PETER, R., GOUGH, J., DEWILDE, S., MOENS, L. & VANFLETEREN, J. R. 2006. A phylogenomic profile of globins. *BMC Evol Biol,* 6**,** 31.

164.) VINOGRADOV, S. N., HOOGEWIJS, D., BAILLY, X., ARREDONDO-PETER, R., GUERTIN, M., GOUGH, J., DEWILDE, S., MOENS, L. & VANFLETEREN, J. R. 2005. Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc Natl Acad Sci U S A,* 102**,** 11385-9.

165.) VINOGRADOV, S. N. & MOENS, L. 2008. Diversity of globin function: enzymatic, transport, storage, and sensing. *J Biol Chem,* 283**,** 8773-7.

166.) VINOGRADOV, S. N., TINAJERO-TREJO, M., POOLE, R. K. & HOOGEWIJS, D. 2013b. Bacterial and archaeal globins - a revised perspective. *Biochim Biophys Acta,* 1834**,** 1789-800.

167.) VISELLI, S. M., REESE, K. R., FAN, J., KOVACS, W. J. & OLSEN, N. J. 1997. Androgens alter B cell development in normal male mice. *Cell Immunol,* 182**,** 99-104.

168.) WANG, W., BERGH, A. & DAMBER, J. E. 2007. Increased expression of CCAAT/enhancer-binding protein beta in proliferative inflammatory atrophy of the prostate: relation with the expression of COX-2, the androgen receptor, and presence of focal chronic inflammation. *Prostate,* 67**,** 1238-46.

169.) WANG, Y., WANG, Y., MA, L., NIE, M., JU, J., LIU, M., DENG, Y., YAO, B., GUI, T., LI, X., GUO, C., MA, C., TAN, R. & ZHAO, Q. 2017. Heterochromatin Protein 1gamma Is a Novel Epigenetic Repressor of Human Embryonic -Globin Gene Expression. *J Biol Chem,* 292**,** 4811-4817.

170.) WEBER, P., CAMMAS, F., GERARD, C., METZGER, D., CHAMBON, P., LOSSON, R. & MARK, M. 2002. Germ cell expression of the transcriptional co-repressor TIF1beta is required for the maintenance of spermatogenesis in the mouse. *Development,* 129**,** 2329-37.

171.) WEISS, M. J. & ORKIN, S. H. 1995. GATA transcription factors: key regulators of hematopoiesis. *Exp Hematol,* 23**,** 99-107.

172.) WILKINS, A. S. & HOLLIDAY, R. 2009. The evolution of meiosis from mitosis. *Genetics,* 181**,** 3-12.

173.) WOOD, M. A. & WALKER, W. H. 2009. USF1/2 transcription factor DNA-binding activity is induced during rat Sertoli cell differentiation. *Biol Reprod,* 80**,** 24-33.

174.) WU, Y., HU, X., LI, Z., WANG, M., LI, S., WANG, X., LIN, X., LIAO, S., ZHANG, Z., FENG, X., WANG, S., CUI, X., WANG, Y., GAO, F., HESS, R. A. & HAN, C. 2016. Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. *Sci Rep,* 6**,** 20435.

175.) XIE, H., YE, M., FENG, R. & GRAF, T. 2004. Stepwise reprogramming of B cells into macrophages. *Cell,* 117**,** 663-76.

176.) XUE, H. H., BOLLENBACHER-REILLEY, J., WU, Z., SPOLSKI, R., JING, X., ZHANG, Y. C., MCCOY, J. P. & LEONARD, W. J. 2007. The transcription factor GABP is a critical regulator of B lymphocyte development. *Immunity,* 26**,** 421-31.

177.) YAGI, R., KOHN, M. J., KARAVANOVA, I., KANEKO, K. J., VULLHORST, D., DEPAMPHILIS, M. L. & BUONANNO, A. 2007. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development,* 134**,** 3827-36.

178.) YANG, L., YEH, S. D., XIE, S., ALTUWAIJRI, S., NI, J., HU, Y. C., CHEN, Y. T., BAO, B. Y., SU, C. H. & CHANG, C. 2004. Androgen suppresses PML protein expression in prostate cancer CWR22R cells. *Biochem Biophys Res Commun,* 314**,** 69-75.

179.) YANG, Z. F., DRUMEA, K., CORMIER, J., WANG, J., ZHU, X. & ROSMARIN, A. G. 2011. GABP transcription factor is required for myeloid differentiation, in part, through its control of Gfi-1 expression. *Blood,* 118**,** 2243-53.

180.) YAO, P. L., LIN, Y. C. & RICHBURG, J. H. 2011. Transcriptional suppression of Sertoli cell Timp2 in rodents following mono-(2-ethylhexyl) phthalate exposure is regulated by CEBPA and MYC. *Biol Reprod,* 85**,** 1203-15.

181.) YOMOGIDA, K., OHTANI, H., HARIGAE, H., ITO, E., NISHIMUNE, Y., ENGEL, J. D. & YAMAMOTO, M. 1994. Developmental stage- and spermatogenic cycle-specific expression of transcription factor GATA-1 in mouse Sertoli cells. *Development,* 120**,** 1759-66.

182.) YOSHIDA, H., ICHIKAWA, H., TAGATA, Y., KATSUMOTO, T., OHNISHI, K., AKAO, Y., NAOE, T., PANDOLFI, P. P. & KITABAYASHI, I. 2007. PML-retinoic acid receptor alpha inhibits PML IV enhancement of PU.1-induced C/EBPepsilon expression in myeloid differentiation. *Mol Cell Biol,* 27**,** 5819-34.

183.) YU, M., RIVA, L., XIE, H., SCHINDLER, Y., MORAN, T. B., CHENG, Y., YU, D., HARDISON, R., WEISS, M. J., ORKIN, S. H., BERNSTEIN, B. E., FRAENKEL, E. & CANTOR, A. B. 2009. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell,* 36**,** 682-95.

184.) ZHANG, P. P., DING, D. Z., SHI, B., ZHANG, S. Q., GU, L. L., WANG, Y. C. & CHENG, C. 2018. Expression of TRIM28 correlates with proliferation and Bortezomib-induced apoptosis in B-cell non-Hodgkin lymphoma. *Leuk Lymphoma,* 59**,** 2639-2649.

185.) ZHANG, Q., IIDA, R., YOKOTA, T. & KINCADE, P. W. 2013. Early events in lymphopoiesis: an update. *Curr Opin Hematol,* 20**,** 265-72.

186.) ZHAO, F., MCCARRICK-WALMSLEY, R., AKERBLAD, P., SIGVARDSSON, M. & KADESCH, T. 2003. Inhibition of p300/CBP by early B-cell factor. *Mol Cell Biol,* 23**,** 3837-46.

187.) ZHU, X., LI, B. & PACE, B. S. 2017. NRF2 mediates gamma-globin gene regulation and fetal hemoglobin induction in human erythroid progenitors. *Haematologica,* 102**,** e285-e288.

188.) ZIMMER, C. 2009. Origins. On the origin of sexual reproduction. *Science,* 324**,** 1254-6.

7 LIST OF ABBREVATIONS

3C	Chromosome conformation capture
5C	Chromosome conformation capture carbon copy
ADGB	Human androglobin
Adgb	Mouse androglobin
AE	ADGB enhancer region
AMP	Adenosine monophosphate
AP	ADGB promoter region
AR	Androgen receptor
ARE	Antioxidative response element
ATP	Adenosine triphosphate
BLP	B lymphoid progenitors
bp	Base pairs
BSA	Bovine serum albumin
°C	Degree Celsius
cDNA	Complementary DNA
CEBPβ	CCAAT Enhancer Binding Protein Beta transcription factor
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag Sequencing
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation and subsequent sequencing
Chr6	Chromosome 6
CLP	Common lymphoid progenitor cells
Cygb	Cytoglobin
dATP	Deoxyribose adenosine triphosphate
dCTP	Deoxyribose cytidine triphosphate
ddH_2O	Double-distilled water
dGTP	Deoxyribose guanosine triphosphate
DH5α *E. coli*	*E. coli* cells designed by Douglas Hanahan
DMEM	Dulbecco´s Modified Eagle´s Medium

DNA	Deoxyribonucleic acid
dNTP	Deoxyribose nucleoside triphosphate
dTTP	Deoxyribose thymine triphosphate
EBF	Early B-cell factor
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetate
e.g.	For example
ENCODE	Encyclopedia of DNA Elements
Enh	Enhancer
ER	Endoplasmic reticulum
ETS	Erythroblast Transformation Specific
FCS	Fetal calf serum
FOG1	Friend of GATA1
g	Gram
GABPα	GA Binding Protein Transcription Factor Subunit Alpha transcription factor
GATA2	GATA Binding Protein 2 transcription factor
GATA3	GATA Binding Protein 3 transcription factor
gDNA	Genomic DNA
GbE	Globin E / eye-globin
GbX	Globin X
GbY	Globin Y
Hb	Hemoglobin
HEK293T	Human embryonic kidney 293 cells
HeLa	Henrietta Lacks cell
Hep3B	Human hepatoma 3 cells
HF	High fidelity
HGP	Human Genome Project
Hi-C	High-throughput sequencing
HP1Gamma	Chromobox 3 transcription factor
HPRT	Hypoxanthine-guanine phosphoribosyltransferase
HsaCAPN7	Human calpain-7

HSC	Hematopoietic stem cell
Ig chain	Immunoglobulin chain
IL-7	Interleukin 7
ImmGen	Immunological Genome Project
Kb	Kilobase
Lar II	Luciferase Assay Reagent II
LB	Lysogeny broth
LECA	Last common eukaryotes ancestor
LMPP	Lymphoid-primed multipotent progenitors
luc+	Luciferase gene
µg	Microgram
µl	Microliter
µM	Micromolar
M	Molar
Mb	Myoglobin
MCF7	Michigan Cancer Foundation 7 cells
mg	Milligram
min	Minute
ml	Milliliter
mm	Millimeter
mM	Millimolar
M-MLV	Moloney Murine Leukemia Virus
mRNA	Messenger RNA
MOPS	3-(N-morpholino) propane sulfonic acid
MPP	Multipotent progenitors
NB	Nuclear bodies
ng	Nanogram
Ngb	Neuroglobin
NK cells	Natural killer cells
NLS	Nuclear localization signals
Np40	Nonoxynol 40
Nrf2	Nuclear Factor, Erythroid 2 Like 2 transcription factor

Nrf3	Nuclear Factor, Erythroid 2 Like 3 transcription factor
P54	DEAD-Box Helicase 6 transcription factor
Pax5	Paired Box 5 transcription factor
PBS	Phosphate-buffered saline
PC	Prostate cancer
PCR	Polymerase chain reaction
pGL3B	pGL3-Basic Vector
pGL3E	pGL3-Enhancer Vector
pGL3P	pGL3-Promoter Vector
pH	Potential of hydrogen
PIWI	P-element-induced wimpy testis
PLB	Passive lysis buffer
PML-IV	Promyelocytic Leukemia 4 transcription factor
PMSF	Phenylmethylsulfonyl fluoride
Prom	Promoter
PU.1	Spi-1 Proto-Oncogene transcription factor
qPCR	Quantitative PCR
RbCl	Rubidium chloride
REST	RE1 Silencing transcription factor
Rfx2	Regulatory Factor X2 transcription factor
RLU	Relative luminescence units
RNA	Ribonucleic acid
rpm	Rounds per minute
RT-PCR	Reverse transcriptase PCR
SDS	Sodium dodecyl sulphate
sec	Seconds
Sp17	Sperm surface protein 17
SV40	Simian virus 40
Tead4	TEA Domain Transcription Factor 4
TF	Transcription factor
TFB	Transformation buffer
Trim 28	Tripartite Motif Containing 28

UCSC	University of California Santa Cruz
USF1	Upstream Transcription Factor 1
W	Watt

8 LIST OF FIGURES

9 LIST OF TABLES

10 ACKNOWLEDGEMENTS

Ich danke Herrn Prof. Dr. Hoogewijs für seine stets freundliche und äußerst motivierende Betreuung dieser Arbeit, für seine großartige Unterstützung und hilfreichen Anregungen.

Mein weiterer Dank gilt der Forschungsgruppe um Prof. Dr. Hoogewijs am Universitätsklinikum Essen, Dr. María Suárez Alonso, Dr. Miriam Cantore, Marteen Chantillon, Marcel Saarbeck und Ozan Karaman. Bedanken möchte ich mich auch bei Dr. María Suárez Alonso und Dr. Miriam Cantore, die mit Engelsgeduld und tatkräftigem Support Probleme beim Einstieg in die experimentelle Forschungsarbeit mit mir durchgestanden haben.

Mein Dank gilt ferner meinen Kollegen der Forschungsgruppen von Prof. Dr. Fandrey und Prof. Dr. Metzen aus dem Institut für Physiologie am Uniklinikum Essen, die mich sehr freundlich aufgenommen haben und mir stets gute Ratschläge gaben, um meine Forschungsexperimente voranzutreiben. Namentlich möchte ich mich insbesondere bei Dr. Ulf Brockmeier und Dr. Teresa Otto bedanken.

Ein großes Dankeschön gebührt auch dem Promotionskolleg ELAN, das mir unter Leitung von Prof. Dr. Rauen und der Else Kröner-Fresenius-Stiftung eine Forschungszeit unter außergewöhnlich gute Rahmenbedingungen ermöglicht hat.

Bedanken möchte ich mich auch bei meinem Bruder Dr. David Markworth und bei Dr. Henrik Dindas für die sprachliche Korrektur meiner Arbeit. Und nicht zuletzt danke ich meiner Familie, Paula Pirschtat und meinen Freunden, die stets einen optimalen privaten Ausgleich für mich geschaffen haben.

11 CURRICULUM VITAE

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.