



UPPSALA  
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Science and Technology 1285*

# The Human Y chromosome and its role in the developing male nervous system

MARTIN M. JOHANSSON



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2015

ISSN 1651-6214  
ISBN 978-91-554-9331-8  
urn:nbn:se:uu:diva-261789

Dissertation presented at Uppsala University to be publicly examined in Zootissalen (EBC 01.01006), Evolutionsbiologiskt centrum, EBC, Villavägen 9, Uppsala, Friday, 23 October 2015 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: David Skuse (University College London Behavioural Sciences Unit).

### Abstract

Johansson, M. M. 2015. The Human Y chromosome and its role in the developing male nervous system. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology* 1285. 63 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9331-8.

Recent research demonstrated that besides a role in sex determination and male fertility, the Y chromosome is involved in additional functions including prostate cancer, sex-specific effects on the brain and behaviour, graft-versus-host disease, nociception, aggression and autoimmune diseases. The results presented in this thesis include an analysis of sex-biased genes encoded on the X and Y chromosomes of rodents. Expression data from six different somatic tissues was analyzed and we found that the X chromosome is enriched in female biased genes and depleted of male biased ones. The second study described copy number variation (CNV) patterns in a world-wide collection of human Y chromosome samples. Contrary to expectations, duplications and not deletions were the most frequent variations. We also discovered novel CNV patterns of which some were significantly overrepresented in specific haplogroups. A substantial part of the thesis focuses on analysis of spatial expression of two Y-encoded brain-specific genes, namely *PCDH11Y* and *NLGN4Y*. The perhaps most surprising discovery was the observation that X and Y transcripts of both gene pairs are mostly expressed in different cells in human spinal cord and medulla oblongata. Also, we detected spatial expression differences for the *PCDH11X* gene in spinal cord. The main focus of the spatial investigations was to uncover genetically coded sexual differences in expression during early development of human central nervous system (CNS). Also, investigations of the expression profiles for 13 X and Y homolog gene pairs in human CNS, adult brain, testes and still-born chimpanzee brain samples were included. Contrary to previous studies, we found only three X-encoded genes from the 13 X/Y homologous gene pairs studied that exhibit female-bias. We also describe six novel non-coding RNAs encoded in the human MSY, some of which are polyadenylated and with conserved expression in chimpanzee brain. The description of dimorphic cellular expression patterns of X- and Y-linked genes should boost the interest in the human specific gene *PCDH11Y*, and draw attention to other Y-encoded genes expressed in the brain during development. This may help to elucidate the role of the Y chromosome in sex differences during early CNS development in humans.

*Keywords:* MSY, sex differences, CNV, SNP, palindrome, palindromes, gr/gr duplication, gr/gr deletion, b2/b3 deletion, b2/b3 duplication, blue-grey duplication, blue-grey like duplication, IR2, U3, STS, AZFa, AZFb, AZFc, Ovary nucleus, Medulla oblongata, spinal cord, white matter, Affymetrix 6.0, embryo, embryonal, haplogroup, haplogroups, R1a, R1b, R-M207, E-M96, I-M170, J-M304, G-M201, Ashkenazi, Bolivian, Chinese, SNP array, padlock probing, AMY-tree

*Martin M. Johansson, Department of Organismal Biology, Norbyvägen 18 A, Uppsala University, SE-75236 Uppsala, Sweden.*

© Martin M. Johansson 2015

ISSN 1651-6214

ISBN 978-91-554-9331-8

urn:nbn:se:uu:diva-261789 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-261789>)

*To our ancestors*

*“Be humble for you are made of earth.  
Be noble for you are made of stars.”  
Serbian Proverb*

Cover:  
*Probe intensity values for the AZFb/c region displaying the  
“Blue-grey like duplication pattern”.*

# List of Papers

This thesis is based on the following papers and manuscripts, which are referred to in the text by their Roman numerals.

- I Reinius B, **Johansson MM**, Radomska KJ, Morrow EH, Pandey GK, Kanduri C, Sandberg R, Williams RW, Jazin E. (2012) Abundance of female-biased and paucity of male-biased somatically expressed genes on the mouse X-chromosome. *BMC Genomics*. 2012, (13):607-625
- II **Johansson MM**, Van Geystelen A, Larmuseau MHD, Djurovic S, Andreassen OA, Agartz, and Jazin E. (2015) Microarray analysis of copy number variants on the human Y chromosome reveals novel and frequent duplications overrepresented in specific haplogroups. *PLoS One*. 2015, 10(8): e0137223
- III **Johansson MM\***, Lundin E,\*, Mirzazadeh M, Qian X, Halvardson J, Darj E, Feuk L, Nilsson M and Jazin E, (2015) Cellular sexual dimorphism of X and Y homolog gene expression in human central nervous system during early male development. *Submitted for publication*.
- IV **Johansson MM**, Suci P, Ahmad T, Zaghlool A, Halvardson J, Etemadikhah M, Darj E, Feuk L and Jazin E. (2015) Sex differences in expression of X and Y genes during development of the human and chimpanzee central nervous system. *Manuscript*.

Reprints were made with permission from the respective publishers.



# Contents

Introduction.....	9
Background.....	11
Why study the Y chromosome.....	11
Evolution of mammalian sex chromosomes.....	12
Selected genes located in MSY.....	13
Genetic structure of the Y chromosome.....	16
MSY sequence.....	17
Palindromes.....	17
Mechanisms generating genomic rearrangements.....	17
The sexual differentiation of the human brain.....	18
Biological implications of Y chromosome abnormalities.....	19
Y chromosome and male fertility.....	19
Y chromosome specific regions involved in male infertility.....	20
Mutational causes.....	22
Aneuploidy states in male infertility.....	23
PAR1 PAR2 related diseases.....	23
Aneuploidy states.....	23
Male biased diseases.....	25
Cancer.....	25
Psychiatric diseases.....	27
Chromosome abnormalities.....	27
Gene abnormalities.....	27
The Aims of the studies.....	29
Paper I.....	29
Paper II.....	29
Manuscript III.....	29
Manuscript IV.....	29
Results and discussion.....	30
Paper I.....	30
Paper II.....	32
Manuscript III.....	37
Manuscript IV.....	41
Conclusions.....	44
Summary in Swedish.....	46
Sammanfattning.....	46
Acknowledgements.....	48
References.....	50

# Abbreviations

MSY	Male specific region
SNP	Single nucleotide polymorphism
CN	Copy number
SNP/CN Probes	Single nucleotide polymorphism and copy number probes
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
HG	Haplogroup
RNAseq	RNA sequencing
UTR	Untranslated region
DRG	Dorsal root ganglion
WM	White matter
ON	Olivary nucleus
STS	Sequence tagged site
AID	Autoimmune disease
SLE	Systemic lupus erythematosus
HCC	Hepatocellular carcinoma
LCR	Low copy repeats
NAHR	Non-allelic homologous recombination
NHEJ	Non-homologous end joining
MMBIR	Microhomology-mediated break induced replication
EAE	Experimental autoimmune encephalomyelitis
GWAS	Genome-wide association study
BPD	Bipolar disorder

# Introduction

The strikingly different human sex chromosomes evolved from an ordinary pair of autosomes<sup>1</sup>. The loss of gene content on the Y chromosome during evolution made it to appear like a dwarf in comparison with its homologous chromosome X. Some scientists predicted future extinction of the Y chromosome<sup>2</sup>, which indeed has occurred during evolution in some species<sup>3</sup>. However recent studies show that the human Y chromosome is here to stay<sup>4</sup>. In comparison, the human X chromosome contains more than twenty times the amount of genes than the Y, of which many are involved in neurodevelopmental processes<sup>5</sup>. A substantial amount of research has focused towards the X chromosome and several reviews documented the results achieved<sup>6,7</sup>. Although this thesis contains a part about X-linked genes, the main focus is directed towards the Y chromosome.

The Y chromosome was not only subjected to loss of genes but also acquisition of autosomal genes occurred throughout its evolution. Recent research showed that export of ancestral X and Y encoded genes by transposition has taken place as well<sup>8</sup>. Those evolutionary events resulted in functional specialization of the Y chromosome mainly in sex determination, male fertility, but also to some extent in functions related to proliferation and connectivity of the central nervous system<sup>9-11</sup>. The Y chromosome affects not only the male developmental pattern of the organism but its presence contributes to several other characteristics including nociception, experimental autoimmune encephalomyelitis (EAE) and density of vasopressin fibers in the lateral septum in adult mice<sup>12</sup>. It is therefore important to further understand the function of Y encoded genes in males.

Humans and mice share a limited amount of orthologous Y genes, resulting in limited possibilities to study non-shared human Y-encoded genes such as *PCDH11Y* and *NLGN4Y*. For human Y-linked genes without homologs in rodents, research is limited to human blood cell expression data, genetic structure analysis, investigation of germ cells of human origin, and analysis of various cell cultures and post mortem tissues<sup>13</sup>. There is a substantial knowledge about Y chromosome gene expression in adult gonads and in the central nervous system, but nothing is known about how these genes affect the development of human embryonic central nervous system at early stages, when the only known genetic differences between males and females are the differently expressed genes encoded on the sex chromosomes<sup>14</sup>. It is therefore of general scientific interest to investigate the possible contribution of

X- and Y-encoded genes to pregonadal sex differences in the developing central nervous system (CNS).

# Background

## Why study the Y chromosome

While I was asking the Director of a sequencing facility at a symposium about the costs and possibilities for sequencing human Y chromosome samples, I was met by laughter and a question regarding who managed to trick me into studying the Y chromosome. Yes, it's tiny in size, it contains few genes compared to other chromosomes and it appears rather uninteresting when compared with its ancestral autosomal fellow X chromosome. Nevertheless, there are several reasons to study it. The first one is that very few research teams focus their attention to the Y chromosome, and there is therefore plenty of room for new discoveries. Another reason is that it is known that the Y chromosome orchestrates the male developmental pattern and the abbreviations in its structure or ploidy state lead to congenital developmental disorders of the reproductive system<sup>15</sup>. Very recent research shows that reduced presence of Y chromosome in some white blood cells, termed as mosaicism, might act as a lifespan predictor in human male populations but the mechanism affecting aging is not yet known<sup>16</sup>. Copy number variations (CNV) in the azoospermia factor (AZF) regions might affect the carrier negatively in terms of reproductive success and a substantial effort was spent on elucidating the actual contribution of CNVs to male infertility. However, the causative genes are not yet clearly established<sup>17</sup>. Studies of the mutation rates and patterns of mutation throughout the male-specific region of the Y chromosome (MSY) have helped to understand human evolution and will assist to uncover the routes used by early modern humans during their expansion on earth.

Even if the genes that survived gene loss caused by lack of recombination with the ancestral X chromosome are few in numbers, some of the survivors are known to have important functions related to human health. For example, recent data suggests that at least two of them *UTY* and *PRKY* has protective roles in coronary artery disease<sup>18</sup>. Effects on behavior are also known. For example, studies in rats showed that switching Y chromosomes between different animal lines had behavioral outcomes in terms of aggression<sup>19</sup>. Further understanding of Y encoded gene function, together with studies on the function of their homologs on the X chromosome, will help to uncover mechanisms of male and female biased diseases. Finally, due to the complicated structural nature of the Y chromosome, and the high sequence similarity

ty between the X and Y homologs, not much is known about the specific spatial and temporal expression of the Y genes at the cellular level. Also of concern, until very recently, the sex chromosomes and foremost the Y chromosome, has been left out of many genome wide studies. This has resulted in a vast amount of data that has not been surveyed and analyzed in different contexts, and awareness about this situation has recently surfaced at research meetings focusing on studies of sex differences.

## Evolution of mammalian sex chromosomes

Prior to the evolution of sex chromosomes in eutherians, sex determination was most likely regulated by environmental factors such as temperature, similar to the sex determination control system operating in some reptilians and fish today<sup>20</sup>. Sex chromosome differentiation in mammals was initiated by the emergence of a sex-determining region on one in a pair of two ancestral autosomes, approximately 180-320 million years ago<sup>1,21</sup>. Mutational events in the ancestor of the autosomal gene pair *SOX3* led to the acquisition of the sex determining region we know of today as the sex-determining region Y (SRY)<sup>22</sup>. The *SRY* gene encodes a transcription factor which activates a downstream cascade of genes and eventually hormones, which altogether direct the organism towards the male developmental pathway<sup>23,24</sup>. After the emergence of the sex determining region, an inversion on the *SRY* bearing chromosome, resulted in suppressed recombination between the emerging sex chromosomes in that particular stratum<sup>21</sup>. This event led to a depletion of genes on the Y chromosome due to accumulation of point mutations, insertions, deletions and repetitive sequences<sup>25</sup>. Meanwhile, loss of recombination allowed for accumulation and grouping of sex specific alleles responsible for male fitness<sup>26</sup>. Three additional inversion events took place during the evolution of the mammalian sex chromosomes leading to further reduction of genetic content on the Y chromosome. During these evolutionary stratification events, the human Y chromosome lost 97% of its ancestral gene content to reach the actual composition of 27 genes, while only 2% of the original genes were lost on its X homolog<sup>27,28</sup>.

In total, 17 of the ancestral genes survived the degeneration of the human Y chromosome sequence and are listed below. From the first strata that formed 180 million of years (Myr) ago, only four gene pairs were retained on the sex chromosomes<sup>4,29</sup>. *SRY/SOX3* together with *RMBX/Y*, *RPS4X/Y* and *HSFX/Y* genes, were all located at the most distal q-arm of the ancient sex chromosomes. The second strata according to the physical order on the present human chromosome retained 10 gene pairs: *KDM5C/D*, *TSPX/TSPY*, *UTX/UTY*, *DDX3X/Y*, *USP9X/Y*, *ZFX/Y*, *IEF1AX/Y*, *TXLNGX/Y*, *TMSB4X/Y* and *AMELX/Y*. The third strata along the X chromosome towards the p-arm retained only two genes: *TBLIX/Y* and

*NLGN4X/Y*. The fourth and last strata that inverted 30 Myr ago during the divergence between old world monkey and hominoids, contains only one functional gene pair, namely *PRKY/Y*<sup>4</sup>.

As a result of the gene loss on the ancestral pair of sex chromosomes, the male cells ended up with a singular X chromosome and Y chromosome (XY). Female cells on the other hand, contained two X chromosomes (XX). The monosomy state in male cells led to development of a dosage compensation mechanism which restored balanced expression between the X and the autosomal chromosomes by overexpression of X-linked genes in mammals<sup>30</sup>. This mechanism not only resulted in balance between autosomes and X but also in balanced expression between males and females. In female cells (XX), the overexpression of X was suppressed resulting in the evolution of the X-chromosome inactivation (XCI) mechanism, which acts epigenetically by inactivating the one X chromosome<sup>31</sup>.

At the same time, evolution took an opposite turn in the avian lineage. In birds, the females contain an heterogametic sex chromosome composition (ZW) while the males are homogametic, displaying two Z chromosomes (ZZ)<sup>32</sup>. Interestingly, in birds, sex determination is not directed by a dominant factor like *SRY*, but is instead based on dosage of a Z-linked gene called *Dmrt1*<sup>33</sup>. Similar to the evolutionary processes that affected Y chromosome in mammals, the W chromosome in birds has been subjected to decay of genetic content<sup>1</sup>.

## Selected genes located in MSY

It would be fair to present a description of all the protein coding genes located in the MSY. Nevertheless, due to limitations of space in this thesis I will only mention a couple of genes that I find interesting and worth knowing of in the perspective of neural development.

*SRY* - Sex determining region Y. Beside the sex differentiation pathway activation, the *SRY* transcription factor have also been found to be expressed in liver, heart, kidney and brain<sup>34</sup>. Brain regions exhibiting *SRY* expression are medial rostral hypothalamus, frontal and temporal cortex<sup>35</sup>.

*PCDH11Y* - Protocadherin 11Y. This gene belongs to the protocadherin family which is a subfamily of the cadherin superfamily. The protein consists of extracellular domain that contains variable number of cadherin repeats, a transmembrane domain and a tail residing in the cytoplasm that is variable in length. *PCDH11Y* is involved in cell to cell recognition during development of central nervous system. *PCDH11Y* is unique to humans while its homolog *PCDH11X* is present in our closest relatives, the chimpanzees<sup>36</sup>. This gene arrived from X chromosome (Xq21.3) by a duplicative

translocation and is currently located on the p-arm in one of the two X-transposed regions (Figure 1, Table 1). Expression has been documented in fetal neocortex, ganglionic eminences, cerebellum and inferior olive<sup>37</sup>.

*NLGN4Y* - Neuroligin 4Y. This gene is a member of the neuroligin family which has beside the X homolog *NLGN4X*, three additional gene variants in humans; *NLGN1,2* and *3*. This family of genes is involved in cell adhesion and the protein for *NLGN4Y* is expressed at the postsynaptic cleft. Its role is essential for formation of synapses but also their maturation<sup>38</sup>.

*UTY* - Ubiquitous TPR motif Y. (Histone Demethylase UTY). This gene encodes a protein which is involved in protein-protein interactions. It has also been described as a histocompatibility antigen which may induce graft rejection of male stem cell grafts<sup>39</sup>. *UTY* is also important during embryonic development of the mice since TALEN mediated editing showed that no pups survived editing of both *Utx* and *Uty*, but male pups with intact *Uty* survived in approximately 50% of the cases<sup>40</sup>.

*TSPY1* - Testis specific protein Y. *TSPY* gene array consists of 20-40 tandem repeated units in the proximal region of Yp11.2. Each of these units is 20.4kb in length harbors one copy of *CYorf16* pseudogene. TSPY is predominantly expressed in testis and is involved in male fertility<sup>41</sup>.

**Table 1. Genes and transcription units residing on the MSY part of the Y chromosome.**

Sequence class	Gene symbol	Gene name	Alternative name	Nr of copies	Tissue expression	X-linked homologue	
X-transposed	TGIF2LY	TGF $\beta$ -induced transcription factor 2-like Y		1	Testis	TGIF2LX	
	PCDH11Y	Protocadherin 11 Y	PCDH22	1	Fetal brain, brain, spinal cord	PCDH11X	
X-degenerate	SRY	Sex determining region Y	TDF	1	Predominantly testis	SOX3	
	RP54Y1	Ribosomal protein S4 Y		1	Ubiquitous	RP54X	
	ZFY	Zinc finger Y		1	Ubiquitous	ZFX	
	AMELY	Amelogenin Y	AMGL	1	Teeth	AMELX	
	TBL1Y	Transducin $\beta$ -like 1		1	Fetal brain, prostate	TBL1X	
	PRKY	Protein Kinase Y		1	Ubiquitous	PRKX	
	USP9Y	Ubiquitin-specific protease 9 Y		1	Ubiquitous	USP9X	
	DDX3Y	Dead box protein 3 Y		1	Ubiquitous	DDX3X	
	UTY	Ubiquitous TPR motif Y	KDM6C	1	Ubiquitous	UTX	
	TMSB4Y	Thymosin $\beta$ -4 Y		1	Ubiquitous	TMSB4X	
	NLGN4Y	Neurologin 4 Y		1	Brain, prostate, testis, spinal cord	NLGN4X	
	Cyorf15A	Chromosome Y open reading frame 15A	TXLINGY	1	Ubiquitous	CXorf15	
	Cyorf15B	Chromosome Y open reading frame 15B	TXLINGY	1	Ubiquitous	CXorf15	
	SMCY	SMC (mouse) homologue Y	KDM5D, JARID1D	1	Ubiquitous	SMCX	
	EIF1AY	Translation initiation factor 1A Y		1	Ubiquitous	EIF1AX	
	RP54Y2	Ribosomal protein S4 Y isoform 2		1	Ubiquitous	RP54X	
	Amplificonic	TSPY	Testis-specific protein Y		~35	Testis	-
		VCY	Variable charge Y	BPY1	2	Testis	VCX
		XKRY	XK related Y		2	Testis	-
CDY		Chromodomain Y		4	Testis	-	
HSFY		Heat shock transcription factor Y		2	Testis	-	
RBMY		RNA-binding motif Y		6	Testis	RBMX	
PRY		PTP-BL related Y	PTPN13-Like Y	2	Testis	-	
BPY2		Basic protein Y 2		3	Testis	-	
DAZ		Deleted in azoospermia	SPGY	4	Testis	-	

## Genetic structure of the Y chromosome

The human Y chromosome contains pseudoautosomal regions, meaning regions that are present in both X and Y chromosomes, and a region that is Y specific MSY. The human pseudoautosomal region (PAR) consists of two regions, PAR1 which is located at the terminal region of the short arms of the X and Y chromosomes, and PAR2 which is located at the distal end of the q-arms of sex chromosomes (Figure 1)<sup>42</sup>. PAR1 (2.6Mb) contains 24 genes while PAR 2 (0.32Mb) harbors only four<sup>43,44</sup>. Genes within the PAR regions undergo recombination during meiosis. This implies that recombination occurs according to classical Mendelian rules of inheritance, contrasting with the MSY region that does not follow these rules. Recombination rate in the PAR regions is higher in males compared to females and exceeds the recombination rate for autosomal chromosomes by 17 times<sup>45-47</sup>. The recombination frequencies for PAR1 and PAR2 differ as well, with the PAR2 having a lower rate by more than one order of magnitude<sup>48</sup>. Interestingly, a new hypothesis postulating recombination between the X-transposed region (XTR) on the Y chromosome and the X locus from which it originated (Xq21.3) has been proposed by Veerappa et al.<sup>49</sup>. This novel hypothesis is based on findings discovered by applying SNP array technology, and it will be interesting to find out whether next generation sequencing technologies will be able to confirm the hypothesis or reject it.

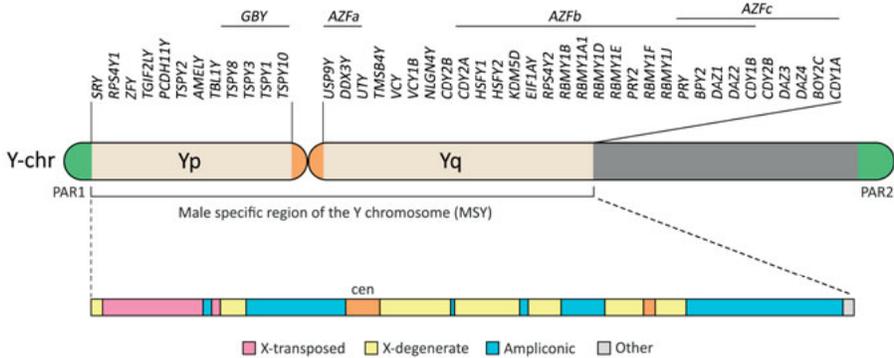


Figure 1. **Schematic representation of human Y chromosome.** Upper part of image show distribution of gene content throughout the MSY and definition of AZF regions. P- and q-arm is colored in pale yellow; PAR regions in green and the centromere in orange. Lower part of the image represents MSY region and the identity of different regions within. Pink color marks X-transposed, yellow for X-degenerate, orange for centromere, blue for ampliconic and gray for other sequences.

## MSY sequence

The MSY region is composed of three different classes of genomic sequences. X-transposed region shares 99% identity to Xq21 from which it was transposed 3-4 million years ago. Within this region two genes are to be found, *PCDH11Y* and *TGIF2LY* both of which have X homologs.

Second sequence class is made up of X-degenerate segments which harbor 16 single-copy genes and several pseudogenes. The sequence homology of these MSY genes and their X homologs reach 96% similarity at most.

The third class, the ampliconic region, is made up of long repeat units which share as much as 99.9% identity with each other. Another frequently used term for those scattered repeats is Amplicons, which exhibit the most gene dense regions of the MSY. Many multicopy genes and genes involved in male fertility are located within the ampliconic sequences<sup>27</sup>.

## Palindromes

Within the regions of ampliconic sequences, eight palindromes are to be found. A palindrome is a sequence which encodes the same order of genetic or alphabetic code independently of readout direction. Each palindrome has a linker sequence at its center which varies in length (2-170kb) and does not share any sequence similarity with the arms of palindromes. Six of eight palindromes carry genes which are expressed specifically in testes, and the very fundamental function of this specific genetic arrangement is to protect and preserve genetic sequences from decay<sup>27,50</sup>.

## Mechanisms generating genomic rearrangements

Genetic rearrangements can occur sporadically, so called *de novo*, or can arise repeatedly and even be transferred to the offspring through the germ line. Certain copy number variations are arising at the same genetic locus and do not vary in length or gene content, these variations are said to be recurrent. Variations that differ in length, locus and gene content are defined as non-recurrent.

For the majority of genetic rearrangements, certain genomic architectures are required. The recurrent rearrangements have breakpoints that cluster at regions enriched for low copy repeats (LCR) that flank the CNV region<sup>51</sup>. Non-recurrent rearrangements lack the clustering of their breakpoints but LCRs are frequently found close to one end of these rearrangements<sup>52</sup>. Short interspersed nuclear elements SINEs and Long interspersed nuclear elements LINEs are often observed at breakpoints of deletions<sup>53</sup>.

There are three major genetic mechanisms that are able to generate copy number alterations in the genome. Non allelic homologous recombination (NAHR) is responsible for generation of deletions, duplications and inver-

sions. The LCRs such as SINEs or LINEs need to be arranged in the same direction and exhibit at least 97% similarity with each other, in order to act as a substrate for the alteration<sup>54</sup>. Non-homologous end joining (NEHJ) is often generating non-recurrent rearrangements where no microhomology is seen at the flanking breakpoints. Instead, small deletions or insertions of bases are associated at the breakpoint regions<sup>55</sup>. A third mechanism capable of generating deletions, duplications, inversions and complex rearrangements is microhomology-mediated break induced replication (MMBIR). During the replication of the genome, a collapse of the replication fork generates one ended double-stranded DNA with the 3' end exposed. The broken end of the DNA can anneal to another replication fork resulting in new DNA synthesis leading to generation of new genomic rearrangements<sup>56</sup>.

In de novo NAHR mediated recurrent rearrangements the deletions on autosomal loci occur at a rate of 2:1, while the Y chromosome seem to exhibit a rate of 4.11:1 at the AZFa-HERV region<sup>57</sup>. It should be noted that this rate is not an average of all observed NAHR mediated rearrangements on Y chromosome.

## The sexual differentiation of the human brain

The male and female human brains do not only differ in terms of morphology but also in terms of behavioral phenotypes produced<sup>58</sup>. During early male fetal development the brain starts to differentiate in “organizational” manner due to hormonal changes that are initiated by the expression of the Y encoded gene *SRY*. This differentiation route starts around week 7 to 12 post gestation in humans<sup>9,59</sup>. The *SRY* protein which acts as a transcription factor, initiates a downstream cascade of gene expression among which *SOX9* is a key player which activates the development of the primordial gonads into testes. At this stage, Leydig cells in the developing gonads start to secrete testosterone which is spread throughout the embryo.

Testosterone is converted to estradiol by the enzyme P450 aromatase, and binds to nuclear oestrogen receptors in brain cells. Upon binding to its ligands, the receptors act as transcription factors which influence gene expression as well as DNA methylation of CpG islands<sup>60,61</sup>. These hormone induced “organizational” events masculinize the developing male brain which is further differentiated from the female brain pattern, by surges of androgen hormones at the pubertal age<sup>60</sup>. Female embryos which lack the *SRY* gene are not directed towards the above described path but instead express *WNT4* and other genes which influence the primordial gonads to develop into ovaries<sup>62</sup>.

The sexual dimorphism of the brain is not only dependent on hormonal surges during the development. Indeed, early studies in rats observed that male embryos were in average heavier than female rat embryos before any

hormonal differentiation had been initiated<sup>63</sup>. Also, other studies found evidence of sex differences in tyrosine hydroxylase-immunoreactive cells (TH), with tyrosine hydroxylase levels elevated in female rat brains prior to gonadal hormone release<sup>64</sup>. These observations clearly suggested that other factors influenced pregonadal sex differences, and the presence of X and Y chromosomes are the obvious candidates.

Due to difficulties in obtaining early human developing CNS samples, no controlled study has yet been carried out with the aim of discovering pregonadal sex differences in the human brain. Several recent studies included a few samples from early developing embryos, but none of them could match sex, tissue and age in an appropriate way<sup>65-67</sup>.

## Biological implications of Y chromosome abnormalities

### Y chromosome and male fertility

Inability to conceive or produce an offspring after one year of unprotected intercourse is defined as infertility<sup>68</sup>. Approximately 10-15% of couples are affected and males are responsible for infertility in 50% of the cases<sup>69</sup>. Female factors explain 50% of infertility causes, male factors contribute by 20-30% and the remaining 20-30% might be explained by interaction of male and female factors. Approximate total worldwide number of infertile men is between 30.625.000-30.641.000 based on self-reported cases. Global percentage of male infertility ranges from 2.5% to 12% varying between the continental regions with largest frequency in Central and Eastern Europe<sup>70</sup>.

Male fertility can be affected by pre-testicular factors such as smoking, alcohol consumption, excessive bicycle riding and complications such as hernia and orchiopexy (*undescended testis*)<sup>71,72</sup>. Post-testicular factors affecting fertility are age, seminoma, varicocele, mumps, chromosomal aneuploidy and microdeletions on the Y chromosome<sup>73,74</sup>.

Different levels of male infertility that directly corresponds to the semen quality are described as Oligozoospermia which implies that the spermatozoa in semen are below 15 million/mL and Azoospermia when there is a total absence of sperm cells in semen. Oligozoospermia is then subdivided into three categories ranging from mild (10-15 million/mL), moderate (5-10 million/mL) to severe (<5 million/mL)<sup>75</sup>.

If the Y chromosome alone was responsible for genetic factors resulting in inability to produce viable spermatids, it would then be easy to find the genes responsible for male infertility. Unfortunately, estimations points towards 2300 genes being involved in rodent male fertility, suggesting a similar amount of genes to be involved in human male fertility<sup>76</sup>.

## Y chromosome specific regions involved in male infertility

Several Y chromosome regions have been associated with infertility of which the most commonly described are summarized below. The first investigations concerning deletions on q-arm of the Y chromosome linked three regions to the infertility phenotype<sup>77</sup>. These regions were denoted azoospermia factor (AZF) a, b and c, each being made up of different structure and containing different set up of genes<sup>78</sup>. AZFa, at Yp11.21, spans 792kb and harbors two genes, *USP9Y* and *DDX3Y*<sup>79</sup>. AZFb is located at Yq11, spans 6.23Mb and contains five single copy genes: *SMCY (KDM5D)*, *EIF1AY*, *RPS4Y2*, *CYorf15A*, *CYorf15B*, and 7 multicopy genes: *XKRY*, *HSFY*, *RMBY1A1*, *PRY*, *CDY*, *BPY2* and *DAZ*<sup>27,80</sup>. Several different limits for the AZFb regions have been suggested, but I will use the definition proposed by Repping et al. 2002, where the AZFb interval spans into the proximal part of the AZFc region (Figure 1). AZFc is located at the very distal end of the MSY region of Yp11.23 and is a highly complex segment constituted by amplicons arranged in a stretch of 3.5Mb. As mentioned previously, the AZFb and AZFc regions overlap partly, sharing the *BPY2*, *CDY* and *DAZ* genes. Within the AZFc region, many transcription units are found, such as non-coding *TTY3* and *TTY4*. Beside these genes, an extensive array of pseudogenes, homologous to several AZFb and AZFc located genes, reside in the region<sup>81</sup>.

Copy number variations within AZF regions have been observed in infertile men, which led to closer investigations of these domains in order to resolve which genes therein are playing a role in male infertility. The results have been intriguing with varying phenotypical outcomes linked to the same copy number variants.

The most proximal of the AZF regions and perhaps the only deletion pattern that results in total lack of germ cells in the testis is the AZFa deletion<sup>82</sup>. This variant is very rare and constitutes 5% of all AZF deletions on the Y chromosome. Interestingly, AZFa duplications are four times more frequent than the deletions<sup>79</sup>. These duplications have no negative impact on fertility since several cases of duplication transfer between fathers and sons have been reported<sup>83</sup>.

Most of the AZFb deletions are called from palindrome 5 to proximal palindrome 1, and they affect no less than 19 genes in a region of 6.2Mb<sup>81</sup>. These deletions are mediated by NAHR but the other variations in this region such as deletion of palindrome 4 appears to be caused by other mechanisms<sup>84</sup>. Interestingly, no reports of AZFb duplications have been published until our recent work described in this thesis<sup>85,86</sup>.

A complete AZFc deletion ranges from amplicon b2 to b4 and removes 3.5Mb of the genome, eliminating completely all *DAZ* and *BPY2* genes and reducing the *CDY* gene copies. Hence, it's designated b2/b4 deletion (Table 2). In contrast with complete AZFa and AZFb deletions, this variant is com-

patible with production of functional gametes, although it leads to decreased sperm production, usually below 1 million/mL<sup>87</sup>. AZFc deletions are the most common of the AZF variants and occur at a frequency of 10-20% in infertile men in different populations<sup>10,88,89</sup>.

**Table 2 Copy number of genes and transcript families in reference sequence and different CNV variants.**

Gene or transcript family	Reference	b2/b4 del	gr/gr del	gr/gr del + b2/b4 dupl	b1/b3 del	b2/b3 del
RBMV	6	6	6	6	4	6
BPY2	3	0	2	4	2	1
DAZ	4	0	2	4	2	2
CDY1	2	0	1	2	2	1
PRY	2	2	2	2	0	2
CSPG4LY	2	0	1	2	2	1
GOLGA2LY	2	0	1	2	2	1
TTY3	2	0	1	2	2	1
TTY4	3	0	2	4	2	1
TTY5	1	1	1	1	0	1
TTY6	2	2	2	2	0	2
TTY17	3	0	2	4	2	1
Total	32	11	23	35	20	20

Due to large amount of different ampliconic sequences in the AZFc region, several different rearrangements are possible. These variants affect a fraction of the complete AZFc interval and are designated according to the substrate amplicons used during the rearrangement. As described below, the phenotypic outcomes are extremely variable.

The most common among AZFc variants is the gr/gr deletion, found in both fertile and infertile men, with average frequency of 5.0% in controls vs 6.1% in cases, based on 24 fertility studies<sup>90</sup>. Interestingly, the frequencies vary greatly between populations, and when the data from a Japanese population study was removed from analysis, the frequencies changed to 3.6% vs 6.2% respectively, increasing the difference between fertile and infertile men<sup>91</sup>. In several studies, the CNV pattern of gr/gr deletion was associated with infertility, (Repping et al. 2003, Ferlin et al. 2005, Navarro-Costa et al. 2007, Yang et al. 2008 and Rosen et al. 2012). On the other hand, other studies reported lack of association (Machev et al. 2004, Hucklenbroich et al. 2005, de Carvalho et al. 2006, Zhang et al. 2006, Wu et al. 2007 and Imken et al 2007).

Presence of gr/gr deletion in both father and offspring indicate that this variant is compatible with fertility. Nevertheless, if gr/gr deletion was selectively neutral, then its presence should reach a frequency of 40% within male population according to population genetic theory<sup>92</sup>. Therefore, this variant

has some functional effect on male phenotypes that remains to be clarified. For example, it has been suggested that these varying effects of gr/gr deletion on male fertility are dependent on some other Y chromosome derived factors or the ancestral forms of genes involved, namely *DAZL* on chromosome 3 and *CDYL* on chromosome 6<sup>93,94</sup>. Duplications of gr/gr appear in some studies at the same rate in control as in cases, 2.6% vs 3.8% respectively, while other studies find duplications significantly overrepresented in cases (0.9% in controls vs 7.0% in cases)<sup>95,96</sup>.

Another CNV pattern frequently detected within the AZFc region is the b2/b3 deletion. This variant does not affect the *RMBY* gene copies but has reducing effect on the copy number of *BPY*, *DAZ*, and *CDY* among others (Table 2). Initial studies did not provide any link between the b2/b3 deletion and infertility. Instead, this variant was described as a fixed CNV within haplogroup N<sup>97-99</sup>. However, subsequent studies with larger sample sizes identified associations between b2/b3 deletions and infertility, mainly in men bearing Y chromosomes outside of the haplogroup N<sup>100-102</sup>. Furthermore, investigations of b2/b3 duplications in Han Chinese in China and Taiwan have found an association between the duplication variant and impaired spermatogenesis<sup>95</sup>. The main outcome from b2/b3 duplication observations is that *DAZ* and *BPY2* seem be the prominent players in spermatogenesis<sup>103</sup>. These recent findings expand the focus in male infertility research from being mainly directed to the effects of deletions, to also include analysis of duplications.

Aside from the AZF regions on the Yq, there is an array of *TSPY* genes located at the distal Yp in proximity to the centromere. The array exhibits variation in amount of the 20.4kb repeats ranging from 18 to 64 copies<sup>104</sup>. Risk of low sperm production and infertility has been associated with both low copy number (below 21) and high copy number (above 55) of *TSPY* repeats in men<sup>105</sup>. Variation in *TSPY* array content has also been described in different haplogroups, for example haplogroup P\*(xR1a) exhibits significantly lower copy number compared to haplogroups J and DE<sup>106</sup>. In a study surveying the infertility in Chinese men, the authors suggested that the *TSPY* array might be an independent factor that significantly affects the phenotypic expression of AZFc deletions in spermatogenesis<sup>105</sup>.

## Mutational causes

In contrast with the traditional approach of associating CNVs on the Y chromosome with male infertility, Sato et al., directed their study into investigations of SNPs associated with the phenotype. From a relatively large sample (n=917) of azoospermic, oligospermic and controls, four loci were identified in the autosomal chromosomes 6,9,13 and 15<sup>107</sup>.

## Aneuploidy states in male infertility

Several pieces of indirect evidence suggest that the Y chromosome may be involved in sex reversal phenotypes but research in this area is not clear. For example, Ross et al.<sup>108</sup> claimed that XYY males had normal spermatogenesis compared with XXY individuals which suffered from testicular failure. This statement contradicts recent results suggesting that duplication and not deletions are responsible for infertility in men from Yi population in China<sup>109</sup>. One possibility to explain these opposing results might be that the balance of factors between X and Y chromosomes is relevant for functional spermatogenesis.

In a study investigating the impact of CNVs of Y chromosomal multi-copy genes on parent-of-origin effects on autoimmune disease in female offspring, the group found evidence for differential microRNA packaging in sperm between two consomic strains and a B6 mouse strain<sup>110</sup>. This finding opens up for another area of research that should be taken into consideration upon investigation of causes of male infertility.

## PAR1 PAR2 related diseases

There are several associations between different diseases or phenotypes to the genes located in PAR1 and PAR2. Since this thesis focuses on the MSY, I will just mention briefly four of these genes. The first gene that attracted attention was the Short stature homeobox (*SHOX*) gene located in PAR1. Copy number variation and mutations within exon 6 of *SHOX* are foremost associated with short stature as well as Léri-Weill dyschondrosteosis which is a bone growth disorder<sup>111-113</sup>. Another gene within PAR1 that has been linked to familial pulmonary alveolar proteinosis is the colony stimulating factor 2 receptor alpha *CSF2RA*, with both deletions and point mutations as causative agents<sup>114</sup>. The Xg blood group antigen *XG* is localized at the distal border of PAR1 and the MSY region and has been associated with both autism and bipolar affective disorder (BPD) in GWAS studies<sup>115,116</sup>. Interestingly, the *VAMP7* on the PAR2 has been associated with BPD<sup>117,118</sup> and duplications of this gene are involved in disruption of male urogenital development<sup>119</sup>.

## Aneuploidy states

In men, the most frequent sex chromosome aneuploidy states are 47, XXY Klinefelter's Syndrome and 47, XYY Syndrome. These syndromes occur at rates of 1:500 to 1:1000 and 1:1000 live male births respectively<sup>120,121</sup>. Male with Klinefelter's genetic setup are in most cases infertile<sup>122</sup> and have smaller testes<sup>123</sup> among other phenotypical characteristics. Interestingly, Klinefelter males show delayed development of speech as well as reading

difficulties<sup>124</sup>. For this thesis, a more interesting genetic setup is the XYY Syndrome which has a subtle, or no obvious phenotypical trait, suggesting that dosage of the Y chromosome is of minor importance compared to X<sup>125</sup>. In general, no congenital anomalies are observed in XYY bearing boys. However, birth weight, height and head circumference have been observed to be above average. Also, mild hypertelorism and broad nasal bridge are seen in XYY children<sup>126</sup>. Tall stature has been commonly observed and it has been related to *SHOX* gene<sup>121,127</sup>. On the behavioral side, attention deficit has been shown in 82% of XYY boys and depressive reactions to stressful events were found to be more frequent in patients vs. controls<sup>124,128</sup>. In contradiction to above mentioned elevated head circumference, one case has been reported where early infantile microcephaly has been described in a boy with XYY Syndrome<sup>129</sup>.

The first triple-Y syndrome 48,XYYY was described in 1965 by Townes, Ziegler and Lenhard<sup>130</sup>. Case descriptions of XYYY individuals note upper-respiratory infections, delayed development milestones such as rolling over, walking, speech delay and mild retardation<sup>131,132</sup>. In additional cases, problem with erection and small testicular size have been described as well as atrophic seminiferous tubules without any spermatogenesis<sup>133,134</sup>. Even though the mice Y chromosome differs from the human Y in terms of arrangement, size and gene content, studies on mice aneuploidy can be used for comparisons with humans. Indeed, infertility has been reported in rodents with aneuploidy states of the Y chromosome<sup>135</sup>.

The absence of “the second X” (or Y chromosome) in humans is denoted as 45, X karyotype and was first described by Henry Turner and named Turner’s syndrome<sup>136</sup>. This aneuploidy state is more interesting in studies focusing on the role of X chromosome, but deserves to be mentioned here since it can give valuable perspectives on effects of Y chromosome loss.

This sex chromosome imbalance has often detrimental impact on the survival of the fetus. Only 1% of conceptions with 45, X genetic setup survive through term<sup>137</sup>. The mosaicism of entire or partial secondary X is often found in the Turner’s diagnosed individuals, affecting the severity of the phenotype<sup>138</sup>. Mosaicism for Y chromosome has been detected in Turner patients as well. One indication of possible mosaicism for Y chromosome sequences is virilization<sup>139</sup>. The mosaicism of Y derived sequences in Turner individuals has been detected in 2-30% of the cases<sup>140</sup>. The presence of partial Y chromosome sequences as well as entire Y chromosome has also been described<sup>141,142</sup>. Even if these types of mosaicism cases are extremely rare, there are several types described, including both 45,X/46,XY and 45,X/47,XYY. Presence of Y chromosome sequences in Turner individuals is associated with risk of developing gonadoblastoma and prophylactic gonadectomy is usually recommended<sup>139</sup>.

Aneuploidy of X and Y chromosome is found to be associated with depressed verbal IQ scores. Interestingly, the Y-aneuploidy as opposed to the

X-aneuploidy, is associated with lower pragmatic versus structural language scores <sup>143</sup>. Taken together, these results suggest that the balance of Y chromosome expression has implications on language, social functioning and autism spectrum disorders (ASD) <sup>144</sup>.

## Male biased diseases

The genetic composition of the chromosome Y might have protective functions or opposite effects in certain diseases. It's only during the course of recent years that more focus has been directed towards gender differences in diseases. Here I will briefly mention some diseases that exhibit overrepresentation in males.

During the course of several studies, it has been shown that men exhibit prevalence for hypertension compared to females, regardless of ethnicity <sup>145,146</sup>. Also, cardiovascular diseases (coronary death, myocardial infarction and stroke) are overrepresented in males <sup>147</sup>. In a comparison of ischemic stroke between male patients and controls, seven Y-linked genes showed differential expression. Five of these are PAR genes (*VAMP7*, *CSF2RA*, *SPRY3*, *DHRX* and *PLDXD*) and two MSY genes (*EIF1AY* and *DDX3Y*). The MSY-linked genes were up-regulated in cases affected by ischemic stroke and the most interesting results from this study indicated that the X homolog *EIF1AX* expression level was not changed while *DDX3X* was up-regulated only at 24h after stroke. This suggests that X and Y homologs of *EIF1A* might differ in their biological functions <sup>148</sup>.

Autoimmune diseases (AID) are affecting females more frequently than males, but when affected, the males show a more severe phenotype <sup>149</sup>. Rheumatological autoimmune diseases are despite the huge variation between countries, more prone to affect males in all geographical locations <sup>150,151</sup>. Prevalence of Crohn's disease in males and females is not considered as a sex biased disease in many countries, with the exception of overrepresentation in the Japanese male population <sup>152</sup>. In cases where males are less prone to certain AIDs it's intriguing to investigate whether the XY complement is protecting the male or if the insulating effect is a result of different hormonal setup. An hypothesis by Moorthy et al. stated that low prevalence of systemic lupus erythematosus (SLE) in males might be caused by negative selection of male fetuses carrying genetic risk factors for SLE <sup>153</sup>. This hypothesis was later confirmed by Affarwal et al. where an excess of fetal male loss was observed in SLE families compared to controls <sup>154</sup>.

## Cancer

In a large study focusing on cancer mortality in the United States between 1977 and 2006, Cook et al. concluded that 32 out of 36 cancer types showed male preference <sup>155</sup>. We know that men are more sensitive to X-linked dis-

eases, since men only carry one copy of X chromosome. At the same time, it is probable that ectopic or unbalanced expression of the Y-linked genes might act as disease causing agent.

Recently, a study used data from the Utah Population Database where the authors investigated 257,252 males diagnosed with prostate cancer and high resolution pedigree records available. The outcome of this large study was that specific Y chromosome subtypes were associated with increased risk for prostate cancer. Even if no specific haplogroups were named, the investigation points out certain Y chromosome configurations which are predisposed towards the disease. Whether it is the Y chromosome itself, or loci on the autosomes associated with a certain Y chromosome subtypes, that are responsible for the association is not yet clear. It would be interesting to extract information about “protective associations” from the study. In other words, it would be informative to investigate which copy number variants of Y-linked genes are associated with low risk for prostate cancer <sup>156</sup>.

Early studies suggested that loss of Y chromosome was frequent in prostate cancer. However, a recent study presents opposite results, indicating that loss of Y chromosome is a rare event in prostate cancer. In fact, only 12 of 2053 observed samples showed loss of Y chromosome <sup>157,158</sup>. Interestingly, a human prostate cancer cell line named PC-3 that did not contain the Y chromosome, was modified by incorporation of an exogenous Y and thereafter injected into athymic nude mice. The presence of the Y chromosome in the cancer cell line led to suppression of tumor growth in the nude mice, suggesting that a gene(s) in the Y chromosome has protective function <sup>159</sup>. However, loss of Y chromosome is described in other cancer forms including head and neck tumors (69% loss), hepatocellular cancer (90% loss), male breast cancer (63% loss) and esophageal squamous cell carcinoma (100% loss) <sup>160-163</sup>.

As previously mentioned in this thesis, Turner females are mosaics for Y-specific sequences that confer risk for gonadoblastoma. Early investigations pointed towards a gonadoblastoma locus Y (GBY) located close to the centromere <sup>164,165</sup>. Although the responsible gene(s) has not yet been found, evidence suggests genes in the proximity of the centromere. *TSPY* at the p-arm prior to the centromere and *USP9Y*, *DDX3Y*, *UTY* at the q-arm are the strongest candidates <sup>166,167</sup>.

Loss of Y chromosome (LOY) in peripheral blood has been observed to occur in elderly males, and a study by Forsberg *et al.* found that LOY was significantly associated with higher risk of nonhematological cancers and shorter cancer survival in aged men <sup>16,168</sup>.

The suggested gene candidate *TSPY* has been frequently found in other cancer types such as liver cancer <sup>169</sup>, melanoma <sup>170</sup> and prostate cancer <sup>171</sup>. In the case of prostate cancer, *TSPY* expression was found more frequently expressed in clinical prostate cancer specimens (78%) as compared to latent prostate cancer (57%) and non-cancer prostate tissues (50%) <sup>171</sup>. Another Y-

encoded gene implicated in cancer, namely *RBMY*, is involved in splicing of testis-specific transcripts and it has been shown to be involved in failure of meiosis upon deletion<sup>172</sup>. This gene presented abnormal expression levels in 36% of male cases with hepatocellular carcinoma (HCC)<sup>173</sup>.

Finally, and somewhat intriguing, it seems that fetal cell microchimerism (FCM) for Y-specific sequences has been associated with disease. FCM is defined as presence of fetal cells in maternal organs, and it has been reported to occur in both rodents and human<sup>174,175</sup>. Studies surveying the FCM from male fetuses have found that the presence of Y-specific sequences have a protective role in Alzheimer's disease (AD) and papillary thyroid cancer<sup>13,176</sup>. For other diseases such as autoimmune disease, there is still diverging data on whether FCM has a detrimental effect on the phenotype or not<sup>177</sup>.

## Psychiatric diseases

Besides a role in male biased diseases and several cancer forms, the Y chromosome has also been implicated in several psychiatric diseases. Most of the evidence regarding contribution of the Y chromosome indicates genes located in the PAR regions<sup>178</sup>. Most information available is found in case descriptions included in medical reports or investigation of cases with aberrations in sex chromosome ratios.

## Chromosome abnormalities

Overrepresentation of autism spectrum disorder (ASD) in males has indicated that sex chromosomes may be involved in the disease. Indeed, the prevalence of ASD in males is 4 times higher than in women<sup>179</sup>. Studies of monozygotic and dizygotic twins found that 60% of the MZ twins were concordant for autism while the DZ twins showed no concordance at all. These early results clearly suggested that there is a main genetic factor contributing to development of autism<sup>180</sup>. Also, early observations of duplications of Y chromosomes or presence of isodicentric Y chromosomes were found in autistic children<sup>181,182</sup>. Autism has also been described in males with 47, XYY syndrome and XYY Klinefelter's, but the prevalence of the disorder is more common in XYY than in XXY carriers (50% vs 12%)<sup>121</sup>. This suggests that an extra copy of the Y chromosome may have higher impact on the probability to develop autism.

## Gene abnormalities

In an attempt to associate a Y chromosome haplogroup with autism, Jamain et al. investigated 111 autism cases and 140 controls in males from France, Sweden and Norway. None of the 12 defined haplogroups could be significantly associated to the disease<sup>183</sup>. Another attempt of haplotype association

was done in 2009 by Serajee et al. using 126 autistic patients and 102 controls. This time, two haplotypes were overrepresented in autistic males. In total, 9 haplotypes could be assigned in this study after genotyping 12 SNPs on the MSY region. There were no significant differences in frequencies of individual SNPs between controls and patients, but haplotypes 3 and 4 were significantly overrepresented in the autistic males<sup>184</sup>. Unfortunately, the haplotypes have not been converted to haplogroups as defined by the Y chromosome consortium<sup>185</sup>. A hint of which MSY genes might be involved in autism came from association of mutations in *NLNG3* and *NLNG4X*<sup>186</sup>. These findings led to focus on the *NLGN4Y* and one missense variant in the gene (p.I679V) was detected in one case out of 290 autistic males. Interestingly, the father of the proband carried the same mutation and exhibited learning disabilities<sup>187</sup>. Nevertheless, the limitation with this finding is that the observed mutation is extremely rare and its possible contribution to autism is very limited. Another approach to study the possible influence of *NLGN4Y* on the autism-related behaviors was done by measuring expression levels of *NLGN4Y* and *RPS4Y* in XYY males and XY controls. The autism-related phenotype correlated with increased expression levels of *NLGN4Y* but not *RPS4Y*, suggesting that further investigations of *NLGN4Y* should be done to establish whether it is a risk gene in XYY males<sup>108</sup>.

# The Aims of the studies

## Paper I

### **Abundance of female-biased and paucity of male-biased somatically expressed genes on the mouse X-chromosome.**

- To identify X chromosome genes sex-biased in somatic mouse tissues.
- To investigate whether the X chromosome is masculinized or feminized in terms of X-linked gene expression in somatic tissues.

## Paper II

### **Microarray Analysis of Copy Number Variants on the Human Y Chromosome Reveals Novel and Frequent Duplications Overrepresented in Specific Haplogroups**

- To investigate whether CNVs on the human Y chromosome can be efficiently detected by the Genome-wide human SNP array 6.0 (Affymetrix).
- To test if any of the detected CNVs are associated in males diagnosed with schizophrenia or bipolar disorder.

## Manuscript III

### **Cellular sexual dimorphism of X and Y homolog gene expression in human central nervous system during early male development**

- To study spatial expression of PCDH11X/Y and NLGN4X/Y homologs in human developing central nervous system.

## Manuscript IV

### **Sex differences in expression of X and Y genes during development of the human and chimpanzee central nervous system**

- To study sex-biased expression in male and female developing CNS
- To investigate not previously described transcriptionally active regions in MSY

# Results and discussion

## Paper I

### **Abundance of female-biased and paucity of male-biased somatically expressed genes on the mouse X-chromosome.**

#### *Setting*

The mammalian sex chromosomes spend different amounts of time in males and females. The X chromosome evolved during 2/3 of the evolutionary time in females while only 1/3 in males. It's therefore expected that female-beneficial dominant genes would be enriched on the X chromosome<sup>188</sup>. Simultaneously, since the X is present in a single copy in male cells, any recessive alleles that result in reproductive advantage for males would be directly available for positive selection<sup>189</sup>. Since the presence of two copies of X exists in female cells, both beneficial and deleterious recessive alleles would be masked from sex-specific selection. The considerations explained above generated a hypothesis stating that the X chromosome should be enriched with male-biased genes<sup>190</sup>. Observations in *Drosophila melanogaster* indicated opposite results, with underrepresentation of male-biased genes on the X chromosome<sup>191</sup>. Also, studies in mice were performed on reproductive tissues and suggested both masculinization and feminization of the X chromosome<sup>192</sup>. Previously, several studies concentrated on expression of X-linked genes in reproductive tissues of males and females, but no equally extensive study of sex-biased expression in somatic tissue had been done when we initiated our investigations. Since sex-biased physiology and sex-biased gene expression occur also outside of gonadal tissues, we decided to investigate X-chromosome-wide sex-biased gene expression in non-reproductive tissues of the *Mus musculus*.

#### *Main results and significance*

The main outcome of our investigation was that female-biased genes were abundant on the X chromosome while the male-biased genes were underrepresented on the X chromosome. De-masculinization or in other words, feminization of the X chromosome has occurred during evolution in non-reproductive tissues. Among the six tissues included, kidney and liver were most enriched for sex-biased X-linked genes. We also observed a vast amount of genes with tissue specific sex bias. This could mean that we found

many novel tissue specific escapee genes or it could be interpreted as the sex-biased X-gene expression is a product of female-skewed gene-regulatory mechanisms. In situ hybridization experiments confirmed one novel escape candidate, *Tmem29* in mouse fibroblast cells. The bi-allelic rate of expression for *Tmem29* was found to be 7%, and confirmed the high sensitivity of our microarray analysis. Finally, we found that novel female-biased non-coding transcripts were located in a cluster close to the well-known escapee gene *Kdm5c*.

### *Study description and discussion*

Many of the differentially expressed genes between male and female mice showed small fold differences (generally below 1.2 fold)<sup>193</sup>. To increase the statistical power and therefore allow for the detection of sex-biased genes with small fold changes, the amount of samples studied must increase substantially. Therefore, we initiated collaboration with a group at the University of Tennessee, USA. This partnership provided us with microarray hybridization data obtained from 728 animals in which six different tissues were studied: kidney, liver, lung, striatum, eye and hippocampus. Sex-biased genes were identified in all of the tissues (Table 1, Paper I). The amount of sex-biased genes between the tissues varied for autosomal and X-linked genes while the Y-linked genes were biased at a constant rate. The two most dimorphic tissues were kidney and liver but also striatum showed a large number of sex-biased genes (Table 1, Paper I) which confirmed previous observations<sup>194</sup>. Besides the expected X-linked gene *Xist*, many of the differentially expressed X genes showed small fold changes compared with the sexually dimorphic autosomal genes (Figure 1, Paper I). This difference might be explained by an evolutionary constraint on sexual expression bias of X-linked genes or by evolutionary restriction on mechanisms that regulate X-linked expression differences.

To be able to scrutinize the hypothesis of both feminization and masculinization of the X chromosome, we compared the frequencies of sex-biased genes between X chromosome and the autosomes. In this comparison, we found the most interesting observation of this study. Namely, male biased genes exhibited a significant underrepresentation on chromosome X (Figure 3, Paper I). Female-biased genes were, on the contrary, significantly overrepresented on the chromosome X in comparison with autosomes. These results might be interpreted as follows: a process of de-masculinization and opposing feminization of the X chromosome took place during evolution within non-reproductive tissues. But a valid question of to what extent were these results affected by the known X escapee genes, needs to be asked?

In order to understand the effect of known X inactivation escapee genes on the observed results, we subtracted those from the observed female-biased genes. Despite the subtraction, the overall amount of female-biased genes did not change dramatically, which indicates that many novel genes

were present in our list (Figure 3a and b, Paper I) or that an unknown sex-skewed gene-regulatory mechanism orchestrated the remaining female-skewed set of genes on the chromosome X.

We also wanted to know whether the sex-biased genes showed any clustering tendency on the X chromosome. Therefore, we studied distribution of differentially expressed genes (Figure 2, Paper I). As the figure indicates, we could not find any clustering that diverged from the overall gene density on X. On the other hand, we found that four non-coding RNAs were expressed in the downstream proximity of the known escapee gene *Kdm5c*. To verify that the sex-biased genes actually escape from inactivation mechanism, we performed RNA in-situ hybridization experiments in mouse fibroblast cell cultures derived from female embryonic skin. Our analysis confirmed that one of these genes, *Tmem29* exhibits expression from both X chromosomes at a frequency of 7% in the cultured fibroblasts. This modest proportion of bi-allelic expression for *Tmem29* indicates that many escapee genes might operate under rather subtle expression levels as compared to the known ones. Unfortunately, none of the four non-coding RNAs detected in the *Kdm5c* cluster were confirmed by *in situ* hybridization. However, it is still possible that these genes, in fact escape inactivation in other cell types.

Several female-biased genes on the X chromosome showed significant expression throughout the six tissues. Interestingly, none of male-biased genes exhibited a significant differential expression across all tissue types, indicating perhaps a tissue specialized/directed male-sex bias of expression (Table 2, Paper I).

## Paper II

### **Microarray Analysis of Copy Number Variants on the Human Y Chromosome Reveals Novel and Frequent Duplications Overrepresented in Specific Haplogroups**

#### *Setting*

Most frequently, the Y chromosome has been studied in order to elucidate its effect on male infertility. In other studies, the PAR regions of the Y chromosome have been investigated with the aim to find loci responsible for short stature and mental retardation<sup>113,195</sup>. Structural variants in MSY have also been applied in forensics in order to tie and untie suspects to crime. The field of anthropology and population genetics has used the possibilities of sequence classifications that the non-recombining MSY has to offer in terms of accumulated mutations and variants. Several different deletion and duplication patterns in the AZF regions have been described and some have been linked to reduced male infertility. Mostly due to technical limitations, many studies focused on certain regions of the MSY leaving the majority of the

chromosome out of the analysis. To circumvent this, we decided to investigate whether the Affymetrix 6.0 array was solid enough for reliable CNV detection within the MSY. Another aim of our investigation was to test whether any detected pattern could be associated to differences in brain morphology. As far as we are informed, no other studies have yet tried to investigate a connection between CNVs in the MSY and possible effects on the male brain.

### *Main results and significance*

By taking into account intensity values from both male and female data we could conclude that 8170 probes out of 8179 on the Genome-wide human SNP array 6.0 (Affymetrix) were reliable for both duplication and deletion detection. By locating known sequence-tagged site (STS) marker positions, or previously described palindromic start and stop locations, we could determine which probes on the array represent ampliconic sequences. Further on, detected patterns could be classified according to the descriptions by Repping et al. 2006<sup>196</sup>, and part of them were confirmed by molecular methods. Most surprisingly, our results showed that duplications are more common than deletions at a rate of 2:1. Also, in total, we could classify 25 variation patterns of which 10 were significantly overrepresented in one or more haplogroups. Aside of confirming previous observations of b2/b3 deletion (c35) being fixed in haplogroup N, and gr/gr deletion (c8) being overrepresented in males belonging to haplogroup D, we found that palindromes 4 and 5 were subjected to both deletions and duplications while palindrome 6 deviated only in terms of duplication. Finally, we observed a novel b2/b3 derived pattern which we denoted as “Blue-grey duplication like” based on its close resemblance to previously described Blue-grey duplication (c449)<sup>196</sup>. Our study demonstrates the importance of controlling the male subjects for haplogroup identity in future genome-wide investigations of Y chromosomal effects, in order to avoid stratification bias. Also, the results can work as a guide as to which CNV patterns are expected to be found in specific haplogroups.

### *Study description and discussion*

Until recently, most of the studies investigating Y chromosomal CNVs focused mainly on AZF regions and they applied classical PCR amplification methods utilizing at best two dozens of STS markers to investigate the 22.5 Mb region of MSY<sup>27</sup>. To circumvent the limitations of such approach, we decided to use a Genome-wide human SNP array 6.0 (Affymetrix) to investigate CNVs at high resolution throughout the entire Y chromosome. This approach allowed us to survey the MSY region at 8279 positions with an average distance of 3.2kb, assessing deletions and duplications simultaneously. Our study was initiated by an analysis of 271 Norwegian males which were participating in a study surveying brain morphology in patients with

schizophrenia and bipolar disorder<sup>197</sup>. The initial results in our study indicated presence of both deletions and duplications within two regions on MSY, namely AZFc and palindrome 6. These observations were confirmed by PCR amplification of STS markers residing within the deleted regions and qPCR for STS markers residing within palindrome 6. Based on STS marker amplification results, and visual patterns displayed by the probe intensity values, we could conclude that distinctive types of deletions were observed, namely gr/gr deletion (c8), and b2/b3 deletion (c35) (supplementary Figure 1, Paper II). By carefully mapping known STS marker positions to the genomic regions on the array, we could identify and define which probes represented specific ampliconic sequences (Figure 2, Paper II). At that point, we could correlate observed probe patterns in the AZFc region to the known and predicted rearrangements described by Repping et al. 2006, supplementary figure 2<sup>196</sup>. Unfortunately, not all ampliconic sequences in the MSY are represented by probes. For amplicons b1, b2, g1, r1-r3 and g2, there are no probes on the array, and their status must be inferred from the response of the probes representing the same type of amplicon. The manufacturer claims that each probe within the amplicons is specific for that ampliconic region only, but if so, then why do we always detect the same effects on the probes representing regions Y1.1 and Y1.2? These two amplicons exhibit always the same variation pattern clearly suggesting that the probes in these regions can't discriminate between Y1 and Y2. Because of these array properties it is difficult to tell which of these two amplicons is actually affected in certain CNV patterns, unless qPCR method is used.

The most common CNV pattern in the Norwegian cohort was surprisingly the b2/b3 deletion (c35), which according to other sources is generally found at low frequency in comparison to gr/gr deletion (c8)<sup>198</sup>. This initial contradicting observation made me suspicious that we in fact might have a population bias in the Norwegian cohort. After all, b2/b3 deletion is known to be fixed in haplogroup N<sup>198</sup>. Could the Norwegian cohort in fact represent a bias in Y chromosomes belonging to haplogroup N? Since the array beside the CN probes also contain 288 SNPs we could, together with our collaborators, use the AMY-tree algorithm to classify the Y chromosome sequences into various haplogroups. The analysis assigned haplogroups at subclade resolution level, but since only 91 out of 246 SNPs that we could genotype were informative; we decided to merge the subclades into the main haplogroups. In this way, we could be more confident when we observed overrepresentation of certain CNVs in specific haplogroups.

At that stage, in order to increase the potential of this developed methodological approach, I expanded the amount of samples by acquiring data from publicly available datasets. This extended the study from 271 to 1718 males improving not only the statistical power but also including several distinct populations. The expansion of the samples resulted in discovery of addition-

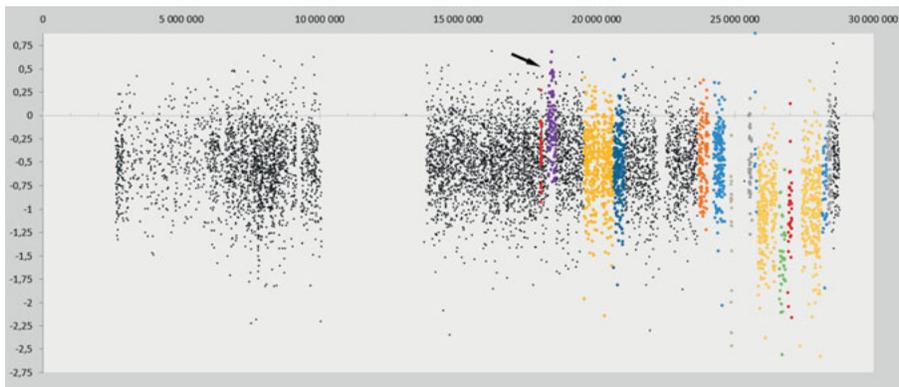
al patterns that were not observed in the Norwegian cohort. For example, the very interesting duplication of regions “prior to P4 post P5”, palindrome 4 duplications, inverted repeat 2 duplications and finally the “Blue-grey like” duplications (Supplementary Figure 1f, h, i, j and p, Paper II).

Most important, the observed overrepresentation of duplications vs. deletions in our data was somewhat surprising. Other studies have reported contrary results. This discrepancy might be explained by the fact that only limited genomic regions were surveyed, or that other array platforms were used which only investigate exon specific probes<sup>57,199</sup>. What might be the reason that we discovered more duplications than deletions? Could it be that duplications of MSY genes might have less severe impact on fitness than deletions? That might actually be the case, since the overrepresentation ratio is 2.1:1. As always, there might be some exceptions. The fact that we only observed palindrome 6 duplications and no deletions among 1718 Y chromosomes, suggests that this variant might be deleterious to the carrier. It should also be noted that this variant was only observed in haplogroups NO-M214(xM175) and R-M207 suggesting that it arose later during human evolution.

Unfortunately, we can't say what effect the observed CNVs have on the reproduction fitness since we don't have access to the fertility status for the included men. My hope is that data accumulated by this study will work as an indication of which variants and frequencies are to be expected in certain haplogroups and that these observations might help in the design of future infertility studies.

Interestingly, recent reports of Chinese Han population point towards the possibility that duplications and not deletions are the main risk factor for male infertility in Asian men<sup>95,109</sup>. In our data the AZFc spanning duplications (b2/b4 deletion duplication (c6), b2/b4 duplication (c21), b2/b4 duplication (c56) and gr/gr duplication (c9) did not show any overrepresentation in any haplogroup, suggesting that these variants are present throughout the entire spectra of the Y chromosome haplogroups. It should be noted that the gr/gr duplication (c9) is the most frequent of the AZFc duplications, reaching a frequency of 2.52%. Interestingly, gr/gr duplication + distal duplication which include duplication of a region after last gr amplicon occur at significant level only within haplogroup J. This could be interpreted as the Y chromosomes in the HG J lineage possess a specific sequence which acts as a substrate for this duplication and that the observed gr/gr duplication + distal duplication might be a potential risk factor for infertility in men from haplogroup J. A re-analysis of our data in which the gr/gr duplications (c9) and gr/gr duplication + distal duplications are merged should be performed in order to assess if the gr/gr duplications will be still overrepresented in haplogroup J.

Another novel observation in our study was that certain CNVs can occur in simultaneously. For example, some Y chromosomes exhibited both duplication of palindrome 6 and deletion of AZFc (Figure 2). This observation might be useful in studies that will further investigate certain CNVs and their contribution to infertility. This especially, as it has been shown that not all men with gr/gr deletion (c8) are infertile, so the duplications, deletions or other regions in the MSY might act as an additional factor that determines the outcome of the gr/gr deletion (c8)<sup>96</sup>. It should also be noted that this array contains probes on the X chromosome which should be investigated in the light of MSY detected variants.



**Figure 2. Duplication and deletion CNVs in MSY.** X axis represents genomic position on the Y chromosome, Y axis represents the Log<sub>2</sub> intensity values. Each dot represents one CN or SNP probe. PAR1, centromere and PAR2 are not represented by probes. Palindrome 6 duplication is indicated by the arrow, probes at the distal end exhibit a b2/b3 deletion (c35) pattern in a Norwegian male.

Finally, I would like to mention our initial aim for this study, namely to investigate whether MSY specific CNVs have any potential effect on male brain morphology. In the statistical analysis of the CNV carriers versus non-CNV males (both cases and controls) we did not detect any significant differences in the 282 brain regions measured. Neither did we find any significant differences when CNV males with diagnosis were compared against non-CNV males with diagnosis. Nevertheless, for regions of Sub Cortical Left Ventral DC and Cortical Area ctx rhentorhinal the p-values were approaching 0.001 even after correction for multiple testing. We know from our results that approximately 11.4% of the Norwegian cohort males carry a CNV on their Y chromosome. By increasing a study size to above 877 samples of males with access to their brain morphology data, one could expect to find approximately 100 males with various CNV patterns. This would dramatically improve the power of the statistical analysis and be hopefully enough for detection of significant differences. Since brain morphology has

been shown to be significantly affected in both schizophrenia and bipolar disorder, it would be necessary to perform such study in control males only. One could argue that AZFc encoded genes are not expressed in the brain and variation within the AZFc region can therefore not have any effect on brain morphology. On the other hand, AZFc variants reduce or overexpress the AZFc encoded genes in the male testes, which might have indirect hormonal effects on the brain. After all, Sertoli cell-only syndrome has been linked to reduced levels of testosterone as well as microdeletions in Yq11 region including AZFc<sup>200,201</sup>.

## Manuscript III

### **Cellular sexual dimorphism of X and Y homolog gene expression in human central nervous system during early male development**

#### *Setting*

Recovering after millennia of gene loss and decades of being mostly ignored by the scientific community, the Y chromosome slowly regains attention in sex difference research and in the neurobiology field. Two studies have shown the contribution of the Y chromosome in early development of male fetal brain<sup>202</sup>. Beside these, spatial and temporal investigations of Y-linked genes have been described in human CNS, ranging from second trimester to adult stages<sup>65</sup>, but yet nothing could be said about the differential expression of X and Y-linked homologues at a cell specific level. The main reason for this, is the extremely high sequence identity between X and Y-linked genes which can reach up to 99.1% between selected isoforms, rendering it difficult to design probes capable of discriminating between X and Y transcripts. Early investigations of Y-linked genes highlighted two genes, namely *PCDH11Y* and *NLGN4Y*, which were believed to be involved in cerebral asymmetry, autism and speech delay respectively<sup>37,187,203</sup>. By applying modern technology based on *in situ* Padlock Probing hybridization, we were able to map the expression of *PCDH11X/Y* and *NLGN4X/Y* in embryonic tissues from Medulla Oblongata and spinal cord. Special care was taken to obtain CNS tissues of males and females from early developmental stages to minimize hormonal influences which might blurry the genetic contribution of gene expression to pre-gonadal sex differences of the human brain.

#### *Experimental considerations*

Experiments in this paper were performed on human tissue from aborted embryos that were collected at the Uppsala University Hospital after written consent was obtained from the donors. Ethical permission for this study was granted by the Regional Ethics Committee in Uppsala (number 2011/329).

To distinguish transcripts with sequence identity as high as of 99.1%, we had to abandon the traditional *in situ* detection approach, which requires dissimilarities in X and Y transcripts of at least 150-200 base pairs, and we therefore looked for novel detection methods. The *in situ* Padlock Probing strategy is designed to discriminate RNA molecules based on a single base pair difference<sup>204</sup>. Since several variations between exons of *PCDH11X/Y* and *NLGN4X/Y* were to be found, we could choose which exons to detect by designing specific cocktails of probes. Also, to confirm the *in situ* padlock probing detection results by an alternative method, we sequenced total RNA and polyA enriched RNAs from both medulla oblongata and midbrain of male and female embryos at equivalent ages.

### *Main results and significance*

Our data showed that all four genes (*PCDH11X/Y* and *NLGN4X/Y*) were expressed in the gray matter of spinal cord and Medulla Oblongata at an early stage of development (8-11 weeks post gestation). It appeared that there were three categories of cells expressing these homologs. We found cells expressing only the X-linked homolog, only the Y-linked homolog and cells expressing both of them. Also, the statistical analysis showed that “cells expressing the X-linked or Y-linked homolog tended to cluster together to a further extent compared to cells expressing both”. Finally, there was a significant gradient for *PCDH11X* expression in the spinal cord. Indeed, despite the fact that there were less cell bodies in the ventral region compared to the dorsal, we found a significant overrepresentation of *PCDH11X* expressing cells in the ventral region.

### *Discussion*

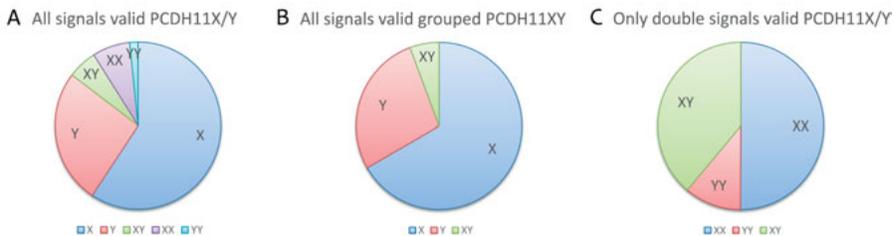
In this section I will concentrate on the discussion of a main statement in the abstract of the manuscript: “The most striking result was that the Y encoded genes are expressed in specific and heterogeneous cellular neural subpopulations that rarely express the X homologs”. First of all, the term rarely is too vague and here I will present the actual numbers for X and Y containing cells given from two different perspectives that I wish to discuss.

In the first scenario, the signal detection system we used is robust and all the existing transcripts within the cells are detected. If this is the case, the data under such assumption is represented in (Figure 3a and b). When all signal containing cells are counted, the percentages of cells exhibiting both X and Y-linked signals for *PCDH11Y* simultaneously reaches a frequency of 5.7%, which corresponds to the proportion described as “rare” in the manuscript.

In the second scenario, limitations of the *in situ* padlock probing technology do not allow for a complete detection of all existing transcripts due to low sensitivity of the method. Indeed, it has been estimated that no more than 30% of transcripts are detected in fixed tissue<sup>205</sup>. Under such condi-

tions, many cells that exhibit only one X- or Y-linked signal might in reality contain a second transcript of either the same or the opposite homolog. In other words, cells with singular X or Y-linked expression are actually cells expressing either XX, YY or XY homologs of *PCDH11*. We should here bear in mind that *PCDH11X/Y* is expressed at low levels compared to many other Y-linked genes (data not shown). In such case, if the analysis would be performed on cells exhibiting only two or more signals, the confidence of the observations would increase and represent to a better extent the actual expression profiles. To do this, a much larger population of tissue samples than the one we used in the manuscript should be included.

Data from the analysis using the second assumption is presented in (Figure 3c) where the XY expressing cells comprise 38.8% of all cells with signals. Such approach of data analysis does not reject the previously cited conclusions, but decreases the amount of cells that might form a male specific network in the brain. To get past the raised concern, a series of experiments in tissues of later developmental stages should be used, since it is known that *PCDH11X/Y* expression is elevated during fetal brain development and slowly drops throughout adult life<sup>65</sup>. Or in other words, signal detection in cells that produce a high number of transcripts should remove any doubts that might be raised in the current data set. Also, it would be of interest to include other X and Y-linked genes that are expressed at higher levels, to investigate whether the trend of sex chromosome specific expression applies for all X and Y homologous genes or if it is only utilized by the two neuro-specific genes on the chromosome Y, *PCDH11Y* and *NLGN4Y*.

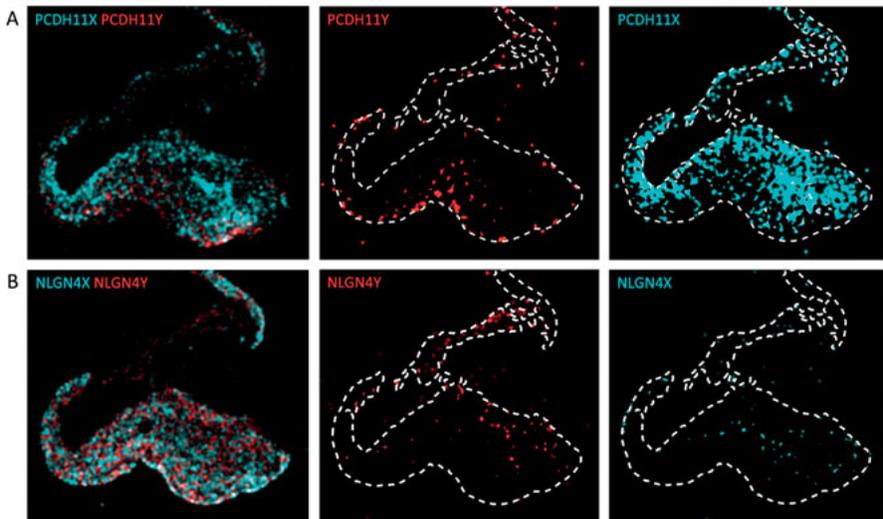


**Figure 3. Proportions of cells exhibiting X, Y, XY, XX, or YY signals.** Class of XX, YY and XY might contain two or more signals. **A** All cells with detected transcripts are presented in separate classes. **B** Cells with XX and YY transcripts are merged into the X and Y class. **C** Only cells exhibiting two or more transcripts are included. All single X and Y cells are excluded.

Regardless of which approach is used for data interpretation, the observation of existing CNS cells expressing only X or Y-linked transcripts of *PCDH11X/Y* holds true. This novel observation is followed by differences in spatial expression of *PCDH11X/Y* in the spinal cord. Indeed, in spite a higher density of cell nuclei in the dorsal horns, the amount of cells expressing *PCDH11X* is significantly higher in the ventral part of the spinal cord (Fig-

ure 2 g, h, i and figure 6 a, b, e, f, Manuscript III,). It's difficult to distinguish whether this enrichment in the ventral part of the gray matter of spinal cord is caused by specific developmental organization processes in the gray matter, or if it is a result of enrichment in X-specific networks. On the other hand, *PCDH11Y* presents an even distribution of transcripts. These observations opens up to a question as to whether *PCDH11X* might be higher expressed in interneurons that wire the efferent system (somatic motor system), while the *PCDH11Y* is biased towards the afferent system (somatic sensory and visceral sensory). Even if not yet quantified, the immunostaining experiments done by our group show that both *PCDH11X/Y* are expressed in some Islet-1 motorneuron progenitor cells.

Interestingly, signal randomization tests showed that *NLGN4X* is confined to certain regions in Medulla Oblongata, while *NLGN4Y* is, as the case with *PCDH11Y* more homogeneously expressed throughout the tissue (Figure 7 a-c, Manuscript III,). No graphical representation of the expression patterns in midbrain tissue was included in the manuscript, but visual inspection of midbrain samples indicated a gradient in *PCDH11X*, being more dominant in the inner part of the tissue while *PCDH11Y* is predominantly detected in the outer parts of the midbrain (Figure 4a).



**Figure 4. Spatial distribution of *PCDH11X/Y* and *NLGN4X/Y* signals in male midbrain.** Figure shows overview of signal distribution in male midbrain tissue. A. Combined expression of *PCDH11X/Y* signals and separate enhanced expression of *PCDH11Y* and *PCDH11X*. B. Combined expression of *NLGN4X/Y* and separate enhanced expression of *NLGN4Y* and *NLGN4X* in consecutive section of midbrain.

The hypothesis of male specific networks including expression of *PCDH11Y* can be questioned. For example, it is not clear whether *PCDH11Y* cells would have a different phenotype than *PCDH11X* containing cells, since the

X and Y homologs display high sequence similarity. Despite the high sequence identity, *PCDH11Y* has accumulated 7 non-synonymous amino acid changes in the ectodomain and 10 in the cyto-domain compared to *PCDH11X*<sup>36</sup>. Beside the non-synonymous substitutions, there is a difference between the homologs of *PCDH11* at the exon composition level. According to the exon definition by Ahn et al, 2009, exon 2 is *PCDH11Y* specific while exons 8, 9 and 10 are only to be found in *PCDH11X*. This setup of different exon usage might have different effects between the transcripts even if those exons were not defined as functional by the investigators. However, the different exons of *PCDH11X/Y* transcripts isoforms might have effects on human male and female brains, especially as some transcripts appear to dominate certain CNS regions. For example, *PCDH11X* isoform 3 appears predominantly in adult brain while *PCDH11Y* isoform 1 transcripts were abundant in fetal brain<sup>206</sup>. Corpus callosum, a region which is known for its contribution to sexual dimorphism of the brain, was found to be enriched for *PCDH11X* isoform 5<sup>207</sup>. It should also be noted that the 13 bp deletion in exon 5 of *PCDH11Y* causes a shift in translation start site leading to N-terminal region with different number of extracellular cadherin domains<sup>206</sup>.

Finally, one observation that may decrease the importance of the function of *PCDH11Y* should be mentioned: 4 males with complete deletion of *PCDH11Y* have been previously described and these males did not display any phenotypic deviations, suggesting that the presence of *PCDH11Y* lacks functional significance<sup>208</sup>. Nevertheless, it should be noted that the mentioned study did not assess the psychological traits of the males with deletions, nor their CNS morphology.

Current findings and technology applied in this study open up for future investigations even at isoform discriminating level of expression of *PCDH11X/Y* in CNS. Since null mutant studies of *PCDH11Y* are not biologically possible in humans (except for rare case descriptions of deletion or duplication events) future functional studies should be performed in human male cells of neuronal origin<sup>203</sup>.

## Manuscript IV

### **Sex differences in expression of X and Y genes during development of the human and chimpanzee central nervous system**

#### *Setting*

In mouse, the neural tube closure occurs at embryonic day 9.5 while sex hormones start to be produced at around gestation day 11.5<sup>209,210</sup>. In humans, the respective embryonic stages are about week 4 for the closure of the neural tube and weeks 8-10 for the initiation of synthesis of oestrogens or androgens and the maturation of sex organs<sup>211,210</sup>. These observations imply

that, in mammals, there is a window of very active CNS development that is not affected by gonadal hormones. During this sex-hormone-independent timeframe, any gender bias in development that is genetically controlled should be the result of the action of genes encoded in the X and Y chromosomes<sup>211-214</sup>. Early studies demonstrated that in mice, several genes encoded on the sex chromosomes are expressed in a sex-biased manner before sexual maturation of the gonads<sup>215-217</sup>. In a more recent study, we used a very large database of more than 700 microarray expression analysis in mouse brain, and we demonstrated that only about a dozen of genes on the mouse X chromosome, together with about a dozen genes in the Y chromosome present significant expression bias between the sexes<sup>194,218</sup>. To investigate the sex biased expression in human developing CNS we employed high-resolution RNA sequencing combined with a nested qPCR approach on embryonic tissues ranging between 8 to 10 weeks post gestation.

#### *Main results and significance*

It was surprising that our strategy only detected *XIST* among female-biased X-encoded genes. In a more detailed investigation of all 992 X-linked genes, only 7 presented significant differences between females and males, including *PAGE4*, *MAGEC3*, *CAPN6*, *APLN*, *ZFX*, *VGLL1* and *GYG2* (Supplementary Table 2, Manuscript IV,). Among the Y-encoded genes 17 were significantly higher expressed in males (Supplementary Table 3, Manuscript IV). Of 13 pairs of genes known to exist in gametolog pairs, including *KDM5D/C*, *DDX3X/Y*, *EIF1AX/Y*, *PRKX/Y*, *TXLNG/P2*, *NLGN4X/Y*, *RPS4X/Y1*, *TMSB4X/Y*, *ZFX/Y*, *USP9X/Y*, *UTX(KDM6A)/Y*, *TBLIX/Y* and *PCDH11X/Y*, only *ZFX* was detected as significantly biased in females. Also, we discovered five long RNAs encoded on the Y chromosome that are expressed in developing human CNS. Of these, five are expressed in new born Chimpanzee brain. None of the long RNAs found in our study was described in macaque cortex studies by He et al. suggesting that all of them are only conserved among higher primates<sup>219</sup>.

#### *Study description and discussion*

To study sex differences in gene expression prior to gonadal differentiation, we collected and RNA sequenced human embryonic CNS samples including medulla oblongata and midbrain. After initial statistical analyses of the differentially expressed genes, we decided to look for novel expression regions that were not previously annotated. Indeed, in close proximity to certain genes, the sequence tracks from males showed accumulation of sequence reads of various amounts. To remove sequences of poor quality and to avoid possibility that the reads map to more than one region, we only counted reads with quality score 30 or more. Thereafter, a design of qPCR primers specific for selected Y chromosome regions was conducted and primers with minimal mismatches to the chimpanzee Y chromosome were selected. For

confirmation of XY human homologs, we designed a strategy where the X primers were shown not to amplify Y specific sequences (Manuscript 4, Methods). Amplification data for the long RNA coding regions exhibited a clear pattern of expression in four out of five regions. Of the five investigated regions, *TTY15/intergenic* exhibited highest expression in both human and chimpanzee males. For all five regions, the male adult brain samples presented lower expression compared to the embryonic and fetal samples indicating that those long RNAs are transcriptionally more active during development. However, the amplification of *ARSEPI* primers resulted in clear amplification in the female chimp sample but not in human female sample. Upon blast of the primer sequences against the chimpanzee genome, we noted that these primers only have one mismatch each compared with a 143bp region located 20491bp at the 5' side of the Arylsulfatase D gene on the X chromosome, which most probably explains the cross-amplification observed.

Four of six of the newly found long RNAs are poly-adenylated suggesting that they may be precursors for micro RNAs (miR). If this is the case, this could change the current view that only the X chromosome and not the Y contain miR genes<sup>220</sup>. To further characterize these novel RNA expressing regions, another set of primers must be designed in order to localize the start and stop position of these transcriptionally active regions. Thereafter, investigations of sequence comparison to other known RNA coding sequences and secondary structure formation predictions must be performed before any categorization of these novel Y-encoded RNAs can be done.

In conclusion, the newly discovered long RNAs have several characteristics that suggest that they might be functional, including tissue specificity, temporal restriction, polyA adenylation and conservation of expression during primate evolution<sup>221</sup>. Functional studies will determine in the future their relevance for the male CNS during development.

# Conclusions

This thesis presents my studies concerning the human Y chromosome in two contexts: the structure in terms of copy number variations in several male populations and the expression of X and Y homologs in human embryos, at tissue- and cell-specific levels. The results of my studies provide descriptions of novel CNVs in the MSY and insight into specific patterns of expression from sex chromosomes, according to which not all CNS cells express X and Y homologs simultaneously.

In the first study, I showed that in somatic tissues of rodents, X chromosome expression is enriched for female-biased genes and at the same time, depleted from male-biased genes. In other words, the X chromosome has been feminized and de-masculinized throughout evolution.

In the second study, I showed that Affymetrix Genome-Wide Human SNP Array 6.0 can be successfully used for detection of CNVs on the Y chromosome. I also showed that duplications and not deletions are more frequent in the male population. Beside this, I described novel CNV patterns related to the b2/b3 deletion and Blue-grey duplication variants. This novel variation was denoted “Blue-grey duplication like” and is only detected in haplogroup NO-M214(xM175). I also showed that both deletion and duplication can occur in the same individual at two separate loci in the MSY.

One of the more important findings of my thesis was included in my third study, and brings a novel understanding of how the X and Y homologs are expressed at the cellular level in the CNS. Together with colleagues, I showed that *PCDH11X* and *PCDH11Y* are to large extent expressed in different cells throughout the human spinal cord and medulla oblongata. We showed that neurons (marked by NeuN antibody), oligodendrocyte precursor cells (*Sox10*) and motoneurons (*Islet-1*) express *PCDH11X/Y*. This is the first time that a description of *PCDH11Y* specific expression has been possible to perform, mainly due to the padlock probing and rolling circle amplification strategy applied. It should also be mentioned that the data presented in this study is acquired from prenatal CNS at a developmental stage where the gonadal hormones synthesis is just about to be initiated.

In the last study, I investigated the differential expression between male and female embryonal CNS. Beside few autosomal genes, the list of significantly sex-biased genes was dominated by the Y-encoded transcripts. I also showed that X encoded homologs in three out of 13 XY homologous gene pairs, exhibit escape from X inactivation. The most novel finding in this

study was the existence of novel non-coding RNAs located on the Y chromosome.

Future efforts should concentrate on long-read genome sequencing of the Y chromosome, which will allow for *de novo* assembly of AZF region diversities throughout a plethora of haplogroups. This approach will omit the limitation of variant discovery based on the currently used reference sequence approach, and it will allow for more detailed structure descriptions of the Y chromosomes in different haplogroups, especially the more ancient ones. To study the potential contribution of different CNVs to male infertility, future investigations should not only focus on the Y-linked genes variation in different CNV patterns, but also their X and autosomal homologs. It would also be of interest to investigate the epigenetic status of the Y chromosome in males with different fertility status.

Regarding Y-linked expression levels, it would be of interest to extend the expression studies described in this thesis to include CNS tissues from later developmental and postnatal stages. This with the aim to find out whether the observed patterns of spatial expression are affected by hormonal surges which occur during early fetal development and in adult life. It is time not only to distinguish *PCDH11Y* from *PCDH11X*, but also to discriminate between expression patterns of different isoforms of these genes. Studies of co-localization of other protocadherins with *PCDH11Y* or neuroligins with *NLGN4Y* in the human cortex will contribute to further understanding and mapping of the herein suggested male specific network in the male brain.

The results presented in my thesis suggest that, despite the limited size and gene content of the Y chromosome, a large amount of research remains to be done to further understand the role of Y chromosome genes in male fertility and in the connectome of the brain.

# Summary in Swedish

## Sammanfattning

Denna avhandling redogör för studier av hur Y kromosomen varierar i strukturform i hos män. Den beskriver även hur uttryck av hur X och Y homologer uttrycks på vävnads och cell specifika nivåer. Resultaten från mina studier presenterar nya beskrivningar av olika strukturvarianter och visar att X och Y specifika gener inte alltid uttrycks samtidigt i samma cell.

Inom ramen för första studien visar jag att X kromosomen i somatiska vävnader uppvisar en överrepresentation av gener med högre uttryck i honor. Med andra ord kan man säga att X kromosomen har blivit feminiserad eller avmaskuliniserad under evolutionens gång. Resultaten visar också att X kromosomala gener som uppvisar tydliga könsskillnader i uttrycksnivåer är evolutionärt sett gamla gener, med en ålder som ofta överstiger 100 miljoner år.

I den andra studien visar jag att Affymetrix Genome-Wide Human SNP Array 6.0 kan användas för detektering av strukturvarianter på Y kromosomen. Resultaten visar att duplikationer till skillnad från deletioner är främst förekommande hos män. Jag beskriver också en ny strukturvariant som är relaterad till b2/b3 deletionen (c35) och Blue-grey duplikationen (c499). Varianten benämns som "Blue-grey like duplication" och den påvisades endast i Y kromosomer tillhörande haplogrupp NO-M214(xM175) som är vanligen förekommande i Finland och norra Sibirien. Datan i studien visar också att två variationer kan uppstå oberoende av varandra längs Y kromosomen.

En av de viktigaste upptäckterna gjordes inom ramen för den tredje studien och visar för första gången hur två neurospecifika Y kromosom gener uttrycks på cellnivå. Tillsammans med samarbetsparteners beskriver vi för första gången genuttrycket för *PCDH11X* och *PCDH11Y* i human hjärnvävnad och påvisar samtidigt att generna till största delen uttrycks oberoende av varandra i olika celler. Vi lokaliserar uttrycket av dessa gener till neuroner, oligodendrocyter och motorneuroner. Studien genomfördes i embryonal hjärnvävnad från tidigt utvecklingsstadium för att därigenom kunna studera Y kromosomens bidrag till könsskillnader innan könshormon bildas och startar könsdifferentieringen av hjärnan.

I den sista studien undersöker jag könsskillnader i hjärnan med avseende på genuttryck mellan foster av manligt och kvinnligt kön. Vid sidan av ett

fåtal autosomala gener, domineras skillnaderna i uttryck av Y kromosomala gener. Jag visar också att endast tre av 13 XY homologa genpar uppvisar eskapism från X inaktivering för X varianten av genen. Den främst banbrytande upptäckten i denna studie är att Y kromosomen till skillnad från vad tidigare var känt, rymmer 5 ickekodande RNA molekyler vilka främst uttrycks i hjärnan hos män under fetal utveckling.

Framtida studier borde koncentreras på "long-read" genom sekvensering av Y kromosomen vilket kommer resultera i *de novo* sammansättning av Y kromosomens DNA sekvens. Genom detta tillvägagångssätt kommer man undkomma snedvridningen av kopiaantal detekteringen som uppkommer pga. att en och samma referenssekvens används vid de flesta av dagens undersökningar. Detta kommer leda till mer korrekta beskrivningar av olika variationsmönster främst i de evolutionärt äldre haplogrupper. Framtida studier av Y kromosomens inverkan på fertilitet borde inte bara fokusera på MSY regionen utan även inkludera variationer på X kromosomen samt de ancestrala generna varifrån dagens Y länkade gener härstammar. Det vore även av intresse att inkludera Y kromosomens epigenetiska status i framtida fertilitetsundersökningar.

När det gäller uttrycksstudier av X och Y homologer så vore det intressant att utöka forskningen till fler och bättre definierade hjärnvävnader som speglar utvecklingsstadier från den embryonala perioden till födseln. Detta för att observera om uppkomna könsskillnader i uttrycksmönster annulleras eller förstärks av den hormonproduktion som sker under olika faser av utvecklingen.

Det är äntligen dags att inte bara skilja uttrycket mellan *PCDH11Y* och *PCDH11X* på gennivå, utan även beskriva hur generna skiljer sig åt i uttryck på isoformnivå. För att vidare förstå rollen som *PCDH11Y* verkar ha på bildandet av för mannen specifika nätverk i hjärnan är det viktigt att studera genen tillsammans med andra gener i protocadherin familjen.

I denna avhandling presenterade resultat visar att trots kromosomens ringa storlek och knappa genantal så finns det mycket utrymme för vidare undersökningar av Y kromosomens bidragande roll i mannens fertilitet och utformningen av manspecifika neuronala nätverk i hjärnan.

# Acknowledgements

I would like to start by acknowledging all the tax payers in Sweden by which their financial contributions made these studies possible. First, big thanks to **Björn Reinius** for initiating me into research at BMC/EBC and patiently helping me out with technical and theoretical issues throughout my studies. Secondly, I would like to thank **Eir Löfgren** for help and enthusiasm during initial CNV investigations. **Carly Schott**, thank you for help with confirmation analyses. **Jon Jakobsson**, thank you for help with investigating the CN/SNP probe variations. **Simon Eckerström Liedholm**, thank you for help with the tedious work of PCR verification and dissection of the ampliconic sequences in AZFc. **Anneleen Van Geystelen** and **Maarten Lar-museau**, thank you joining the CNV project and haplogrouping the Y chromosomes. **Ingrid Agartz**, **Srdajan Djurovic**, and **Lavinia Athanasiu**, thank you for sharing and helping out with SNP data and DNA. **Bryn Farnsworth**, thank you for discussions regarding statistical analyses and introduction to SPSS. Thank you **Reza Mirzazadeh**, for initiating the Padlock probing experiments and transferring the knowledge of the method. **Elin Lundin**, thank you for a great cooperation and patience in stressed situations and tedious work of signal readout. **Mats Nilsson**, thank you for all your technical support and statistical discussions during the interpretations of the data. **Xiaoyan Qian**, thank you for the great heat maps and density plots of our experiments. **Elisabeth Darj**, thank you for your cooperation, without your contribution this thesis would not be possible. **Jonathan Halvardson**, thank you for the statistical analyses and figures from the RNAseq data. **Ammar Zaghlool** and **Mitra Etemadikhah**, thank you for help with RNA and tissue preparations. **Lars Feuk**, thank you for the collaboration and support regarding the RNAseq. **Katarzyna Radomska**, thank you for initial help with tissue collection and immuno experiments. **Sophie Sanchez**, thanks for the help with the EndNote. **Daniel Snitting**, thanks for interesting discussions about the cosmos. **Qingming Qu**, thanks for being such an inspiring researcher and colleague. **Helena Malmikumpu**, thank you for help with ordering and advices regarding tissue preparations. **Rose-Marie Löfberg**, thank you for help with practical and administrative issues. **Lina Emilsson**, thank you for being my secondary supervisor and showing your support. **Professor Per Ahlberg**, thank you for taking your time to discuss the history of the Scandinavian runestones. **Professor Reinald Fundele**, thank you for introducing me to the sport of archery. **Professor David**

**Skuse**, thank you coming all the way to Uppsala for my dissertation. **Professor Art Arnold**, thank you for making my stay in Mount Pinos unforgettable. Thank you my dear wife, **Marzieh Farzamfar**, for supporting, cheering me up and pushing me to work hard during those years. Thank you also for all help with initial Y chromosome plots. An important gratitude goes to my **Grandmother, Mother** and **Stepfather** who encouraged me to persuade the path of knowledge and cheered me to do the very best in every situation. Finally, and most importantly I would like to thank **Professor Elena Jazin**. Thank you for accepting me as your student and teaching me how to find the hidden details in the data. Also, thank you for all the support with scientific writing and manuscript preparation.

# References

- 1 Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488-493, doi:10.1038/nature13151 (2014).
- 2 Graves, J. A. The degenerate Y chromosome--can conversion save it? *Reproduction, fertility, and development* **16**, 527-534, doi:10.10371/rd03096 (2004).
- 3 Castillo ER, M. D., Bidau CJ. Sex- and neo-sex chromosomes in Orthoptera: a review. *Journal of Orthoptera Research.*, 213–231 (2010).
- 4 Bellott, D. W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494-499, doi:10.1038/nature13206 (2014).
- 5 Skuse, D. H. X-linked genes and mental functioning. *Human molecular genetics* **14 Spec No 1**, R27-32, doi:10.1093/hmg/ddi112 (2005).
- 6 Davies, W. & Wilkinson, L. S. It is not all hormones: alternative explanations for sexual differentiation of the brain. *Brain research* **1126**, 36-45, doi:10.1016/j.brainres.2006.09.105 (2006).
- 7 Nguyen, D. K. & Disteche, C. M. High expression of the mammalian X chromosome in brain. *Brain research* **1126**, 46-49, doi:10.1016/j.brainres.2006.08.053 (2006).
- 8 Hughes, J. F., Skaletsky, H., Koutseva, N., Pyntikova, T. & Page, D. C. Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. *Genome biology* **16**, 104, doi:10.1186/s13059-015-0667-4 (2015).
- 9 Hanley, N. A. *et al.* SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development. *Mechanisms of development* **91**, 403-407 (2000).
- 10 Krausz, C. & McElreavey, K. Y chromosome and male infertility. *Frontiers in bioscience : a journal and virtual library* **4**, E1-8 (1999).
- 11 Kopsida, E., Stergiakouli, E., Lynn, P. M., Wilkinson, L. S. & Davies, W. The Role of the Y Chromosome in Brain Function. *Open neuroendocrinology journal* **2**, 20-30, doi:10.2174/1876528900902010020 (2009).
- 12 Arnold, A. P. Mouse models for evaluating sex chromosome effects that cause sex differences in non-gonadal tissues. *Journal of neuroendocrinology* **21**, 377-386, doi:10.1111/j.1365-2826.2009.01831.x (2009).
- 13 Chan, W. F. *et al.* Male microchimerism in the human female brain. *PloS one* **7**, e45592, doi:10.1371/journal.pone.0045592 (2012).
- 14 Ahmadi Rastegar, D. *et al.* Isoform Level Gene Expression Profile of Human Y Chromosome Azoospermia Factor Genes and their X Paralogues in the Testicular Tissue of Non-obstructive Azoospermia Patients. *Journal of proteome research*, doi:10.1021/acs.jproteome.5b00520 (2015).
- 15 Massanyi, E. Z., Dicarlo, H. N., Migeon, C. J. & Gearhart, J. P. Review and management of 46,XY disorders of sex development. *Journal of pediatric urology* **9**, 368-379, doi:10.1016/j.jpuro.2012.12.002 (2013).

- 16 Forsberg, L. A. *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nature genetics* **46**, 624-628, doi:10.1038/ng.2966 (2014).
- 17 Navarro-Costa, P., Plancha, C. E. & Goncalves, J. Genetic dissection of the AZF regions of the human Y chromosome: thriller or filler for male (in)fertility? *Journal of biomedicine & biotechnology* **2010**, 936569, doi:10.1155/2010/936569 (2010).
- 18 Bloomer, L. D. *et al.* Coronary artery disease predisposing haplogroup I of the Y chromosome, aggression and sex steroids--genetic association analysis. *Atherosclerosis* **233**, 160-164, doi:10.1016/j.atherosclerosis.2013.12.012 (2014).
- 19 Toot, J., Dunphy, G., Turner, M. & Ely, D. The SHR Y-chromosome increases testosterone and aggression, but decreases serotonin as compared to the WKY Y-chromosome in the rat model. *Behavior genetics* **34**, 515-524, doi:10.1023/B:BEGE.0000038489.82589.6f (2004).
- 20 Warner, D. A. & Shine, R. The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* **451**, 566-568, doi:10.1038/nature06519 (2008).
- 21 Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964-967 (1999).
- 22 Graves, J. A. The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **350**, 305-311; discussion 311-302, doi:10.1098/rstb.1995.0166 (1995).
- 23 Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117-121, doi:10.1038/351117a0 (1991).
- 24 Morais da Silva, S. *et al.* Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. *Nature genetics* **14**, 62-68, doi:10.1038/ng0996-62 (1996).
- 25 Muller, H. J. THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. *Mutation research* **106**, 2-9 (1964).
- 26 Wallis, M. C., Waters, P. D. & Graves, J. A. Sex determination in mammals--before and after the evolution of SRY. *Cellular and molecular life sciences : CMLS* **65**, 3182-3195, doi:10.1007/s00018-008-8109-z (2008).
- 27 Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-837, doi:10.1038/nature01722 (2003).
- 28 Mueller, J. L. *et al.* Independent specialization of the human and mouse X chromosomes for the male germ line. *Nature genetics* **45**, 1083-1087, doi:10.1038/ng.2705 (2013).
- 29 Potrzebowski, L. *et al.* Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS biology* **6**, e80, doi:10.1371/journal.pbio.0060080 (2008).
- 30 Nguyen, D. K. & Disteche, C. M. Dosage compensation of the active X chromosome in mammals. *Nature genetics* **38**, 47-53, doi:10.1038/ng1705 (2006).
- 31 Lyon, M. F. X-chromosome inactivation. *Current biology : CB* **9**, R235-237 (1999).
- 32 Mittwoch, U. Sex determination in birds and mammals. *Nature* **231**, 432-434 (1971).

- 33 Smith, C. A., Katz, M. & Sinclair, A. H. DMRT1 is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos. *Biology of reproduction* **68**, 560-570 (2003).
- 34 Clepet, C. *et al.* The human SRY transcript. *Human molecular genetics* **2**, 2007-2012 (1993).
- 35 Mayer, A., Lahr, G., Swaab, D. F., Pilgrim, C. & Reisert, I. The Y-chromosomal genes SRY and ZFY are transcribed in adult human brain. *Neurogenetics* **1**, 281-288 (1998).
- 36 Williams, N. A., Close, J. P., Giouzezi, M. & Crow, T. J. Accelerated evolution of Protocadherin11X/Y: a candidate gene-pair for cerebral asymmetry and language. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **141B**, 623-633, doi:10.1002/ajmg.b.30357 (2006).
- 37 Priddle, T. H. & Crow, T. J. Protocadherin 11X/Y a human-specific gene pair: an immunohistochemical survey of fetal and adult brains. *Cerebral cortex* **23**, 1933-1941, doi:10.1093/cercor/bhs181 (2013).
- 38 Bottos, A., Rissone, A., Bussolino, F. & Arese, M. Neurexins and neuroligins: synapses look out of the nervous system. *Cellular and molecular life sciences : CMLS* **68**, 2655-2666, doi:10.1007/s00018-011-0664-z (2011).
- 39 Vogt, M. H. *et al.* UTY gene codes for an HLA-B60-restricted human male-specific minor histocompatibility antigen involved in stem cell graft rejection: characterization of the critical polymorphic amino acid residues for T-cell recognition. *Blood* **96**, 3126-3132 (2000).
- 40 Wang, H. *et al.* TALEN-mediated editing of the mouse Y chromosome. *Nature biotechnology* **31**, 530-532, doi:10.1038/nbt.2595 (2013).
- 41 Tyler-Smith, C., Taylor, L. & Muller, U. Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *Journal of molecular biology* **203**, 837-848 (1988).
- 42 Rappold, G. A. The pseudoautosomal regions of the human sex chromosomes. *Human genetics* **92**, 315-324 (1993).
- 43 Helena Mangs, A. & Morris, B. J. The Human Pseudoautosomal Region (PAR): Origin, Function and Future. *Current genomics* **8**, 129-136 (2007).
- 44 Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325-337, doi:10.1038/nature03440 (2005).
- 45 Rouyer, F. *et al.* A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**, 291-295, doi:10.1038/319291a0 (1986).
- 46 Flaquer, A., Rappold, G. A., Wienker, T. F. & Fischer, C. The human pseudoautosomal regions: a review for genetic epidemiologists. *European journal of human genetics : EJHG* **16**, 771-779, doi:10.1038/ejhg.2008.63 (2008).
- 47 Hinch, A. G., Altemose, N., Noor, N., Donnelly, P. & Myers, S. R. Recombination in the human Pseudoautosomal region PAR1. *PLoS genetics* **10**, e1004503, doi:10.1371/journal.pgen.1004503 (2014).
- 48 Li, L. & Hamer, D. H. Recombination and allelic association in the Xq/Yq homology region. *Human molecular genetics* **4**, 2013-2016 (1995).
- 49 Veerappa, A. M., Padakannaya, P. & Ramachandra, N. B. Copy number variation-based polymorphism in a new pseudoautosomal region 3 (PAR3) of a human X-chromosome-transposed region (XTR) in the Y chromosome. *Functional & integrative genomics* **13**, 285-293, doi:10.1007/s10142-013-0323-6 (2013).

- 50 Balaresque, P. *et al.* Gene conversion violates the stepwise mutation model for microsatellites in y-chromosomal palindromic repeats. *Human mutation* **35**, 609-617, doi:10.1002/humu.22542 (2014).
- 51 Gu, W., Zhang, F. & Lupski, J. R. Mechanisms for human genomic rearrangements. *PathoGenetics* **1**, 4, doi:10.1186/1755-8417-1-4 (2008).
- 52 Vissers, L. E. *et al.* Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Human molecular genetics* **18**, 3579-3593, doi:10.1093/hmg/ddp306 (2009).
- 53 Conrad, D. F. *et al.* Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nature genetics* **42**, 385-391, doi:10.1038/ng.564 (2010).
- 54 Stankiewicz, P. & Lupski, J. R. Genome architecture, rearrangements and genomic disorders. *Trends in genetics : TIG* **18**, 74-82 (2002).
- 55 Lieber, M. R. The mechanism of human nonhomologous DNA end joining. *The Journal of biological chemistry* **283**, 1-5, doi:10.1074/jbc.R700039200 (2008).
- 56 Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS genetics* **5**, e1000327, doi:10.1371/journal.pgen.1000327 (2009).
- 57 Turner, D. J. *et al.* Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nature genetics* **40**, 90-95, doi:10.1038/ng.2007.40 (2008).
- 58 Qureshi, I. A. & Mehler, M. F. Genetic and epigenetic underpinnings of sex differences in the brain and in neurological and psychiatric disease susceptibility. *Progress in brain research* **186**, 77-95, doi:10.1016/B978-0-444-53630-3.00006-3 (2010).
- 59 Garcia-Falgueras, A. & Swaab, D. F. Sexual hormones and the brain: an essential alliance for sexual identity and sexual orientation. *Endocrine development* **17**, 22-35, doi:10.1159/000262525 (2010).
- 60 Nugent, B. M. & McCarthy, M. M. Epigenetic underpinnings of developmental sex differences in the brain. *Neuroendocrinology* **93**, 150-158, doi:10.1159/000325264 (2011).
- 61 Sekido, R. The potential role of SRY in epigenetic gene regulation during brain sexual differentiation in mammals. *Advances in genetics* **86**, 135-165, doi:10.1016/B978-0-12-800222-3.00007-3 (2014).
- 62 Kashimada, K. & Koopman, P. Sry: the master switch in mammalian sex determination. *Development* **137**, 3921-3930, doi:10.1242/dev.048983 (2010).
- 63 Scott, W. J. & Holson, J. F. Weight differences in rat embryos prior to sexual differentiation. *Journal of embryology and experimental morphology* **40**, 259-263 (1977).
- 64 Beyer, C., Pilgrim, C. & Reisert, I. Dopamine content and metabolism in mesencephalic and diencephalic cell cultures: sex differences and effects of sex steroids. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **11**, 1325-1333 (1991).
- 65 Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489, doi:10.1038/nature10523 (2011).
- 66 Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11836-11841, doi:10.1073/pnas.1209647109 (2012).

- 67 Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 7285-7290, doi:10.1073/pnas.1507125112 (2015).
- 68 Hopps, C. V. *et al.* Detection of sperm in men with Y chromosome microdeletions of the AZFa, AZFb and AZFc regions. *Human reproduction* **18**, 1660-1665 (2003).
- 69 Okabe, M., Ikawa, M. & Ashkenas, J. Male infertility and the genetics of spermatogenesis. *American journal of human genetics* **62**, 1274-1281, doi:10.1086/301895 (1998).
- 70 Agarwal, A., Mulgund, A., Hamada, A. & Chyatte, M. R. A unique view on male infertility around the globe. *Reproductive Biology and Endocrinology : RB&E* **13**, 37, doi:10.1186/s12958-015-0032-1 (2015).
- 71 Sohrabvand, F., Jafari, M., Shariat, M., Haghollahi, F. & Lotfi, M. Frequency and epidemiologic aspects of male infertility. *Acta medica Iranica* **53**, 231-235 (2015).
- 72 Leibovitch, I. & Mor, Y. The vicious cycling: bicycling related urogenital disorders. *European urology* **47**, 277-286; discussion 286-277, doi:10.1016/j.eururo.2004.10.024 (2005).
- 73 Casella, R., Leibundgut, B., Lehmann, K. & Gasser, T. C. Mumps orchitis: report of a mini-epidemic. *The Journal of urology* **158**, 2158-2161 (1997).
- 74 Mukhopadhyay, D. *et al.* Semen quality and age-specific changes: a study between two decades on 3,729 male partners of couples with normal sperm count and attending an andrology laboratory for infertility-related problems in an Indian city. *Fertility and sterility* **93**, 2247-2254, doi:10.1016/j.fertnstert.2009.01.135 (2010).
- 75 Organization, W. H. WHO Laboratory Manual for the Examination and Processing of Human Semen. (2010).
- 76 Schultz, N., Hamra, F. K. & Garbers, D. L. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12201-12206, doi:10.1073/pnas.1635054100 (2003).
- 77 Tiepolo, L. & Zuffardi, O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Human genetics* **34**, 119-124 (1976).
- 78 Vogt, P. H. *et al.* Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human molecular genetics* **5**, 933-943 (1996).
- 79 Bosch, E. Duplications of the AZFa region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Human molecular genetics* **12**, 341-347, doi:10.1093/hmg/ddg031 (2003).
- 80 Ferlin, A., Moro, E., Rossi, A., Dallapiccola, B. & Foresta, C. The human Y chromosome's azoospermia factor b (AZFb) region: sequence, structure, and deletion analysis in infertile men. *Journal of medical genetics* **40**, 18-24 (2003).
- 81 Repping, S. *et al.* Recombination between Palindromes P5 and P1 on the Human Y Chromosome Causes Massive Deletions and Spermatogenic Failure. *American journal of human genetics* **71**, 906-922 (2002).

- 82 Krausz, C. & Degl'Innocenti, S. Y chromosome and male infertility: update, 2006. *Frontiers in bioscience : a journal and virtual library* **11**, 3049-3061 (2006).
- 83 Arruda, J. T., Silva, D. M., Silva, C. C., Moura, K. K. & da Cruz, A. D. Homologous recombination between HERVs causes duplications in the AZFa region of men accidentally exposed to cesium-137 in Goiania. *Genetics and molecular research : GMR* **7**, 1063-1069 (2008).
- 84 Vinci, G. *et al.* A deletion of a novel heat shock gene on the Y chromosome associated with azoospermia. *Molecular human reproduction* **11**, 295-298, doi:10.1093/molehr/gah153 (2005).
- 85 Jobling, M. A. Copy number variation on the human Y chromosome. *Cytogenetic and genome research* **123**, 253-262, doi:10.1159/000184715 (2008).
- 86 Johansson MM, V. G. A., Larmuseau MHD, Djurovic S, Andreassen OA, Agartz I, *et al.* (2015). Microarray Analysis of Copy Number Variants on the Human Y Chromosome Reveals Novel and Frequent Duplications Overrepresented in Specific Haplogroups. *PloS one*, doi:doi:10.1371/journal.pone.0137223 (2015).
- 87 Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature genetics* **29**, 279-286, doi:10.1038/ng757 (2001).
- 88 Vogt, P. H. Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. *Molecular human reproduction* **4**, 739-744 (1998).
- 89 Simoni, M. Molecular diagnosis of Y chromosome microdeletions in Europe: state-of-the-art and quality control. *Human reproduction* **16**, 402-409 (2001).
- 90 Navarro-Costa, P., Goncalves, J. & Plancha, C. E. The AZFc region of the Y chromosome: at the crossroads between genetic diversity and male infertility. *Human reproduction update* **16**, 525-542, doi:10.1093/humupd/dmq005 (2010).
- 91 de Carvalho, C. M. *et al.* Study of AZFc partial deletion gr/gr in fertile and infertile Japanese males. *Journal of human genetics* **51**, 794-799, doi:10.1007/s10038-006-0024-2 (2006).
- 92 Repping, S. *et al.* Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nature genetics* **35**, 247-251, doi:10.1038/ng1250 (2003).
- 93 Lahn, B. T. & Page, D. C. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature genetics* **21**, 429-433, doi:10.1038/7771 (1999).
- 94 Xu, E. Y., Moore, F. L. & Pera, R. A. A gene family required for human germ cell development evolved from an ancient meiotic gene conserved in metazoans. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7414-7419, doi:10.1073/pnas.131090498 (2001).
- 95 Lin, Y. W. *et al.* Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Human mutation* **28**, 486-494, doi:10.1002/humu.20473 (2007).
- 96 Giachini, C. *et al.* Partial AZFc deletions and duplications: clinical correlates in the Italian population. *Human genetics* **124**, 399-410, doi:10.1007/s00439-008-0561-1 (2008).
- 97 Repping, S. *et al.* A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. *Genomics* **83**, 1046-1052, doi:10.1016/j.ygeno.2003.12.018 (2004).

- 98 Fernandes, S. *et al.* A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. *American journal of human genetics* **74**, 180-187, doi:10.1086/381132 (2004).
- 99 Sathyanarayana, S. H. & Malini, S. S. N. Impact of Y chromosome AZFc subdeletion shows lower risk of fertility impairment in Siddi tribal men, Western Ghats, India. *Basic and Clinical Andrology* **25**, 1, doi:10.1186/s12610-014-0017-5 (2015).
- 100 Wu, B. *et al.* A frequent Y chromosome b2/b3 subdeletion shows strong association with male infertility in Han-Chinese population. *Human reproduction* **22**, 1107-1113, doi:10.1093/humrep/del499 (2007).
- 101 Eloualid, A. *et al.* Association of spermatogenic failure with the b2/b3 partial AZFc deletion. *PLoS one* **7**, e34902, doi:10.1371/journal.pone.0034902 (2012).
- 102 Lu, C. *et al.* The b2/b3 subdeletion shows higher risk of spermatogenic failure and higher frequency of complete AZFc deletion than the gr/gr subdeletion in a Chinese population. *Human molecular genetics* **18**, 1122-1130, doi:10.1093/hmg/ddn427 (2009).
- 103 Lu, C. *et al.* Gene copy number alterations in the azoospermia-associated AZFc region and their effect on spermatogenic impairment. *Molecular human reproduction* **20**, 836-843, doi:10.1093/molehr/gau043 (2014).
- 104 Xue, Y. & Tyler-Smith, C. An Exceptional Gene: Evolution of the TSPY Gene Family in Humans and Other Great Apes. *Genes* **2**, 36-47, doi:10.3390/genes2010036 (2011).
- 105 Shen, Y. *et al.* A significant effect of the TSPY1 copy number on spermatogenesis efficiency and the phenotypic expression of the gr/gr deletion. *Human molecular genetics* **22**, 1679-1695, doi:10.1093/hmg/ddt004 (2013).
- 106 Giachini, C. *et al.* TSPY1 copy number variation influences spermatogenesis and shows differences among Y lineages. *The Journal of clinical endocrinology and metabolism* **94**, 4016-4022, doi:10.1210/jc.2009-1029 (2009).
- 107 Sato, Y. *et al.* An association study of four candidate loci for human male fertility traits with male infertility. *Human reproduction* **30**, 1510-1514, doi:10.1093/humrep/dev088 (2015).
- 108 Ross, J. L., Tartaglia, N., Merry, D. E., Dalva, M. & Zinn, A. R. Behavioral phenotypes in males with XYY and possible role of increased NLGN4Y expression in autism features. *Genes, brain, and behavior* **14**, 137-144, doi:10.1111/gbb.12200 (2015).
- 109 Ye, J. J. *et al.* Partial AZFc duplications not deletions are associated with male infertility in the Yi population of Yunnan Province, China. *Journal of Zhejiang University. Science. B* **14**, 807-815, doi:10.1631/jzus.B1200301 (2013).
- 110 Case, L. K. *et al.* Copy number variation in Y chromosome multicopy genes is linked to a paternal parent-of-origin effect on CNS autoimmune disease in female offspring. *Genome biology* **16**, 28, doi:10.1186/s13059-015-0591-7 (2015).
- 111 Binder, G., Schwarze, C. P. & Ranke, M. B. Identification of short stature caused by SHOX defects and therapeutic effect of recombinant human growth hormone. *The Journal of clinical endocrinology and metabolism* **85**, 245-249, doi:10.1210/jcem.85.1.6375 (2000).
- 112 Seki, A. *et al.* Skeletal Deformity Associated with SHOX Deficiency. *Clinical pediatric endocrinology : case reports and clinical investigations : official journal of the Japanese Society for Pediatric Endocrinology* **23**, 65-72, doi:10.1297/cpe.23.65 (2014).

- 113 Fukami, M. *et al.* Rare pseudoautosomal copy-number variations involving SHOX and/or its flanking regions in individuals with and without short stature. *Journal of human genetics*, doi:10.1038/jhg.2015.53 (2015).
- 114 Suzuki, T. *et al.* Familial pulmonary alveolar proteinosis caused by mutations in CSF2RA. *The Journal of experimental medicine* **205**, 2703-2710, doi:10.1084/jem.20080990 (2008).
- 115 Chang, S. C., Pauls, D. L., Lange, C., Sasanfar, R. & Santangelo, S. L. Sex-specific association of a common variant of the XG gene with autism spectrum disorders. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **162b**, 742-750, doi:10.1002/ajmg.b.32165 (2013).
- 116 Flaquer, A. *et al.* A new susceptibility locus for bipolar affective disorder in PAR1 on Xp22.3/Yp11.3. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **153b**, 1110-1114, doi:10.1002/ajmg.b.31075 (2010).
- 117 Muller, D. J. *et al.* Association between a polymorphism in the pseudoautosomal X-linked gene SYBL1 and bipolar affective disorder. *American journal of medical genetics* **114**, 74-78 (2002).
- 118 Saito, T., Parsia, S., Papolos, D. F. & Lachman, H. M. Analysis of the pseudoautosomal X-linked gene SYBL1 in bipolar affective disorder: description of a new candidate allele for psychiatric disorders. *American journal of medical genetics* **96**, 317-323 (2000).
- 119 Tannour-Louet, M. *et al.* Increased gene copy number of VAMP7 disrupts human male urogenital development through altered estrogen action. *Nature medicine* **20**, 715-724, doi:10.1038/nm.3580 (2014).
- 120 Herlihy, A. S., Halliday, J. L., Cock, M. L. & McLachlan, R. I. The prevalence and diagnosis rates of Klinefelter syndrome: an Australian comparison. *The Medical journal of Australia* **194**, 24-28 (2011).
- 121 Ross, J. L. *et al.* Behavioral and social phenotypes in boys with 47,XXY syndrome or 47,XXY Klinefelter syndrome. *Pediatrics* **129**, 769-778, doi:10.1542/peds.2011-0719 (2012).
- 122 Glander, H. J. [Infertility in the Klinefelter syndrome]. *MMW Fortschritte der Medizin* **147**, 39-41 (2005).
- 123 Smyth, C. M. & Bremner, W. J. Klinefelter syndrome. *Archives of internal medicine* **158**, 1309-1314 (1998).
- 124 Visootsak, J. & Graham, J. M., Jr. Social function in multiple X and Y chromosome disorders: XXY, XYY, XXYY, XXXY. *Developmental disabilities research reviews* **15**, 328-332, doi:10.1002/ddr.76 (2009).
- 125 Milunsky, J. M. in *Genetic Disorders and the Fetus* 273-312 (Wiley-Blackwell, 2010).
- 126 Lalatta, F. *et al.* Early manifestations in a cohort of children prenatally diagnosed with 47,XXY. Role of multidisciplinary counseling for parental guidance and prevention of aggressive behavior. *Italian journal of pediatrics* **38**, 52, doi:10.1186/1824-7288-38-52 (2012).
- 127 Ottesen, A. M. *et al.* Increased number of sex chromosomes affects height in a nonlinear fashion: a study of 305 patients with sex chromosome aneuploidy. *American journal of medical genetics. Part A* **152a**, 1206-1212, doi:10.1002/ajmg.a.33334 (2010).
- 128 Lalatta, F. *et al.* Triple X syndrome: characteristics of 42 Italian girls and parental emotional response to prenatal diagnosis. *European journal of pediatrics* **169**, 1255-1261, doi:10.1007/s00431-010-1221-8 (2010).

- 129 Nguyen-Minh, S., Buhner, C., Hubner, C. & Kaindl, A. M. Is microcephaly a so-far unrecognized feature of XYY syndrome? *Meta gene* **2**, 160-163, doi:10.1016/j.mgene.2013.10.013 (2014).
- 130 Townes, P. L., Ziegler, N. A. & Lenhard, L. W. A Patient with 48 Chromosomes (Xyyy). *Lancet* **1**, 1041-& (1965).
- 131 Schoepflin, G. S. & Centerwall, W. R. 48,XYYY: a new syndrome? *Journal of medical genetics* **9**, 356-360 (1972).
- 132 James, C., Robson, L., Jackson, J. & Smith, A. 46,Xy/47,Xyy/48,Xyyy Karyotype in a 3-Year-Old Boy Ascertained Because of Radioulnar Synostosis. *American journal of medical genetics* **56**, 389-392, doi:DOI 10.1002/ajmg.1320560408 (1995).
- 133 Hori, N. *et al.* A Male-Subject with 3 Y-Chromosomes (48,Xyyy) - a Case-Report. *J Urology* **139**, 1059-1061 (1988).
- 134 Hunter, H. & Quaife, R. 48,Xyyy Male - Somatic and Psychiatric Description. *Journal of medical genetics* **10**, 80-82, doi:Doi 10.1136/Jmg.10.1.80 (1973).
- 135 Rodriguez, T. A. & Burgoyne, P. S. Spermatogenic failure in male mice with four sex chromosomes. *Chromosoma* **110**, 124-129 (2001).
- 136 Turner, H. H. A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology* **23**, 566-574 (1938).
- 137 Cockwell, A., MacKenzie, M., Youings, S. & Jacobs, P. A cytogenetic and molecular study of a series of 45,X fetuses and their parents. *Journal of medical genetics* **28**, 151-155 (1991).
- 138 Hook, E. B. & Warburton, D. The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Human genetics* **64**, 24-27 (1983).
- 139 Fayez, J. A. & Jonas, H. S. Virilization in Turner syndrome. *Obstetrics and gynecology* **52**, 490-492 (1978).
- 140 Binder, G., Koch, A., Wajs, E. & Ranke, M. B. Nested polymerase chain reaction study of 53 cases with Turner's syndrome: is cytogenetically undetected Y mosaicism common? *The Journal of clinical endocrinology and metabolism* **80**, 3532-3536, doi:10.1210/jcem.80.12.8530595 (1995).
- 141 Hsieh, Y.-Y. *et al.* Turner Syndrome with Pseudodicentric Y Chromosome Mosaicism. *Journal of assisted reproduction and genetics* **19**, 302-303, doi:10.1023/A:1015737515898 (2002).
- 142 Cortes-Gutierrez, E. I. *et al.* Molecular detection of cryptic Y-chromosomal material in patients with Turner syndrome. *Oncology reports* **28**, 1205-1210, doi:10.3892/or.2012.1916 (2012).
- 143 Lee, N. R. *et al.* Dosage effects of X and Y chromosomes on language and social functioning in children with supernumerary sex chromosome aneuploidies: implications for idiopathic language impairment and autism spectrum disorders. *Journal of child psychology and psychiatry, and allied disciplines* **53**, 1072-1081, doi:10.1111/j.1469-7610.2012.02573.x (2012).
- 144 Bishop, D. V. *et al.* Autism, language and communication in children with sex chromosome trisomies. *Archives of disease in childhood* **96**, 954-959, doi:10.1136/adc.2009.179747 (2011).
- 145 Cutler, J. A. *et al.* Trends in hypertension prevalence, awareness, treatment, and control rates in United States adults between 1988-1994 and 1999-2004. *Hypertension* **52**, 818-827, doi:10.1161/hypertensionaha.108.113357 (2008).

- 146 Stamler, J., Stamler, R., Riedlinger, W. F., Algera, G. & Roberts, R. H. Hypertension screening of 1 million Americans. Community Hypertension Evaluation Clinic (CHEC) program, 1973 through 1975. *Jama* **235**, 2299-2306 (1976).
- 147 Pencina, M. J., D'Agostino, R. B., Sr., Larson, M. G., Massaro, J. M. & Vasan, R. S. Predicting the 30-year risk of cardiovascular disease: the framingham heart study. *Circulation* **119**, 3078-3084, doi:10.1161/circulationaha.108.816694 (2009).
- 148 Tian, Y. *et al.* Y chromosome gene expression in the blood of male patients with ischemic stroke compared with male controls. *Gender medicine* **9**, 68-75 e63, doi:10.1016/j.genm.2012.01.005 (2012).
- 149 Ngo, S. T., Steyn, F. J. & McCombe, P. A. Gender differences in autoimmune disease. *Frontiers in neuroendocrinology* **35**, 347-369, doi:10.1016/j.yfrne.2014.04.004 (2014).
- 150 Reveille, J. D. Epidemiology of spondyloarthritis in North America. *The American journal of the medical sciences* **341**, 284-286, doi:10.1097/MAJ.0b013e31820f8c99 (2011).
- 151 Aggarwal, R. & Malaviya, A. N. Clinical characteristics of patients with ankylosing spondylitis in India. *Clinical rheumatology* **28**, 1199-1205, doi:10.1007/s10067-009-1227-7 (2009).
- 152 Asakura, K. *et al.* Prevalence of ulcerative colitis and Crohn's disease in Japan. *Journal of gastroenterology* **44**, 659-665, doi:10.1007/s00535-009-0057-3 (2009).
- 153 Moorthy, L. N., Peterson, M. G., Onel, K. B. & Lehman, T. J. Do children with lupus have fewer male siblings? *Lupus* **17**, 128-131, doi:10.1177/0961203307085111 (2008).
- 154 Aggarwal, R., Sestak, A. L., Chakravarty, E. F., Harley, J. B. & Scofield, R. H. Excess female siblings and male fetal loss in families with systemic lupus erythematosus. *The Journal of rheumatology* **40**, 430-434, doi:10.3899/jrheum.120643 (2013).
- 155 Cook, M. B., McGlynn, K. A., Devesa, S. S., Freedman, N. D. & Anderson, W. F. Sex disparities in cancer mortality and survival. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **20**, 1629-1637, doi:10.1158/1055-9965.EPI-11-0246 (2011).
- 156 Cannon-Albright, L. A. *et al.* Identification of specific Y chromosomes associated with increased prostate cancer risk. *The Prostate* **74**, 991-998, doi:10.1002/pros.22821 (2014).
- 157 Stahl, P. R. *et al.* Y chromosome losses are exceedingly rare in prostate cancer and unrelated to patient age. *The Prostate* **72**, 898-903, doi:10.1002/pros.21492 (2012).
- 158 Konig, J. J. *et al.* Loss and gain of chromosomes 1, 18, and Y in prostate cancer. *The Prostate* **25**, 281-291 (1994).
- 159 Vijayakumar, S. *et al.* The human Y chromosome suppresses the tumorigenicity of PC-3, a human prostate cancer cell line, in athymic nude mice. *Genes, chromosomes & cancer* **44**, 365-372, doi:10.1002/gcc.20250 (2005).
- 160 Jacobs, P. A. *et al.* Male breast cancer, age and sex chromosome aneuploidy. *British journal of cancer* **108**, 959-963, doi:10.1038/bjc.2012.577 (2013).
- 161 Kujawski, M. *et al.* Frequent chromosome Y loss in primary, second primary and metastatic squamous cell carcinomas of the head and neck region. *Cancer letters* **208**, 95-101, doi:10.1016/j.canlet.2003.11.006 (2004).

- 162 Park, S. J., Jeong, S. Y. & Kim, H. J. Y chromosome loss and other genomic alterations in hepatocellular carcinoma cell lines analyzed by CGH and CGH array. *Cancer genetics and cytogenetics* **166**, 56-64, doi:10.1016/j.cancergencyto.2005.08.022 (2006).
- 163 Yamaki, H. *et al.* Alteration of X and Y chromosomes in human esophageal squamous cell carcinoma. *Anticancer research* **21**, 985-990 (2001).
- 164 Page, D. C. Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. *Development* **101 Suppl**, 151-155 (1987).
- 165 Salo, P. *et al.* Molecular mapping of the putative gonadoblastoma locus on the Y chromosome. *Genes, chromosomes & cancer* **14**, 210-214 (1995).
- 166 Lau, Y. F., Li, Y. & Kido, T. Gonadoblastoma locus and the TSPY gene on the human Y chromosome. *Birth defects research. Part C, Embryo today : reviews* **87**, 114-122, doi:10.1002/bdrc.20144 (2009).
- 167 Knauer-Fischer, S. *et al.* Analyses of Gonadoblastoma Y (GBY)-locus and of Y centromere in Turner syndrome patients. *Experimental and clinical endocrinology & diabetes : official journal, German Society of Endocrinology [and] German Diabetes Association* **123**, 61-65, doi:10.1055/s-0034-1387734 (2015).
- 168 Wiktor, A. E., Van Dyke, D. L., Hodnefield, J. M., Eckel-Passow, J. & Hanson, C. A. The significance of isolated Y chromosome loss in bone marrow metaphase cells from males over age 50 years. *Leukemia research* **35**, 1297-1300, doi:10.1016/j.leukres.2011.05.002 (2011).
- 169 Yin, Y. H. *et al.* TSPY is a cancer testis antigen expressed in human hepatocellular carcinoma. *British journal of cancer* **93**, 458-463, doi:10.1038/sj.bjc.6602716 (2005).
- 170 Gallagher, W. M. *et al.* Multiple markers for melanoma progression regulated by DNA methylation: insights from transcriptomic studies. *Carcinogenesis* **26**, 1856-1867, doi:10.1093/carcin/bgi152 (2005).
- 171 Kido, T., Hatakeyama, S., Ohyama, C. & Lau, Y. F. Expression of the Y-Encoded TSPY is Associated with Progression of Prostate Cancer. *Genes* **1**, 283-293, doi:10.3390/genes1020283 (2010).
- 172 Elliott, D. J. *et al.* Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 3848-3853 (1997).
- 173 Tsuei, D. J. *et al.* RBMY, a male germ cell-specific RNA-binding protein, activated in human liver cancers and transforms rodent fibroblasts. *Oncogene* **23**, 5815-5822, doi:10.1038/sj.onc.1207773 (2004).
- 174 Jonsson, A. M., Uzunel, M., Gotherstrom, C., Papadogiannakis, N. & Westgren, M. Maternal microchimerism in human fetal tissues. *American journal of obstetrics and gynecology* **198**, 325.e321-326, doi:10.1016/j.ajog.2007.09.047 (2008).
- 175 Tan, K. H., Zeng, X. X., Sasajala, P., Yeo, A. & Udolph, G. Fetomaternal microchimerism: Some answers and many new questions. *Chimerism* **2**, 16-18, doi:10.4161/chim.2.1.14692 (2011).
- 176 Cirello, V. *et al.* Fetal cell microchimerism in papillary thyroid cancer: studies in peripheral blood and tissues. *International journal of cancer. Journal international du cancer* **126**, 2874-2878, doi:10.1002/ijc.24993 (2010).
- 177 Fugazzola, L., Cirello, V. & Beck-Peccoz, P. Fetal microchimerism as an explanation of disease. *Nature reviews. Endocrinology* **7**, 89-97, doi:10.1038/nrendo.2010.216 (2011).

- 178 d'Amato, T. *et al.* Pseudoautosomal region in schizophrenia: linkage analysis of seven loci by sib-pair and lod-score methods. *Psychiatry research* **52**, 135-147 (1994).
- 179 Skuse, D. H. Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatric research* **47**, 9-16 (2000).
- 180 Bailey, A. *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological medicine* **25**, 63-77 (1995).
- 181 Blackman, J. A., Selzer, S. C., Patil, S. & Van Dyke, D. C. Autistic disorder associated with an iso-dicentric Y chromosome. *Developmental medicine and child neurology* **33**, 162-166 (1991).
- 182 Kasparis, C. & Loffeld, A. Childhood acne in a boy with XYY syndrome. *BMJ case reports* **2014**, doi:10.1136/bcr-2013-201587 (2014).
- 183 Jamain, S. *et al.* Y chromosome haplogroups in autistic subjects. *Molecular psychiatry* **7**, 217-219, doi:10.1038/sj.mp.4000968 (2002).
- 184 Serajee, F. J. & Mahbulul Huq, A. H. Association of Y chromosome haplotypes with autism. *Journal of child neurology* **24**, 1258-1261, doi:10.1177/0883073809333530 (2009).
- 185 Consortium, T. Y. C. A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome research* **12**, 339-348, doi:10.1101/gr.217602 (2002).
- 186 Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature genetics* **34**, 27-29, doi:10.1038/ng1136 (2003).
- 187 Yan, J. *et al.* Analysis of the neuroligin 4Y gene in patients with autism. *Psychiatric genetics* **18**, 204-207, doi:10.1097/YPG.0b013e3282fb7fe6 (2008).
- 188 Bachtrog, D. A dynamic view of sex chromosome evolution. *Current opinion in genetics & development* **16**, 578-585, doi:10.1016/j.gde.2006.10.007 (2006).
- 189 Reinke, V. Sex and the genome. *Nature genetics* **36**, 548-549, doi:10.1038/ng0604-548 (2004).
- 190 Rice, W. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**, 735-742 (1984).
- 191 Parisi, M. *et al.* Paucity of genes on the Drosophila X chromosome showing male-biased expression. *Science* **299**, 697-700, doi:10.1126/science.1079190 (2003).
- 192 Khil, P. P., Smirnova, N. A., Romanienko, P. J. & Camerini-Otero, R. D. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nature genetics* **36**, 642-646, doi:10.1038/ng1368 (2004).
- 193 Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome research* **16**, 995-1004, doi:10.1101/gr.5217506 (2006).
- 194 Reinius, B. *et al.* Female-biased expression of long non-coding RNAs in domains that escape X-inactivation in mouse. *BMC Genomics* **11**, 614, doi:10.1186/1471-2164-11-614 (2010).
- 195 Jorgez, C. J. *et al.* Aberrations in pseudoautosomal regions (PARs) found in infertile men with Y-chromosome microdeletions. *The Journal of clinical endocrinology and metabolism* **96**, E674-679, doi:10.1210/jc.2010-2018 (2011).
- 196 Repping, S. *et al.* High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature genetics* **38**, 463-467, doi:10.1038/ng1754 (2006).

- 197 Rimol, L. M. *et al.* Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. *Biological psychiatry* **68**, 41-50, doi:10.1016/j.biopsych.2010.03.036 (2010).
- 198 Rozen, S. G. *et al.* AZFc deletions and spermatogenic failure: a population-based survey of 20,000 Y chromosomes. *American journal of human genetics* **91**, 890-896, doi:10.1016/j.ajhg.2012.09.003 (2012).
- 199 Wei, W. *et al.* Copy number variation in the human Y chromosome in the UK population. *Human genetics*, doi:10.1007/s00439-015-1562-5 (2015).
- 200 Ferras, C. *et al.* AZF and DAZ gene copy-specific deletion analysis in maturation arrest and Sertoli cell-only syndrome. *Molecular human reproduction* **10**, 755-761, doi:10.1093/molehr/gah104 (2004).
- 201 Tuttelmann, F. *et al.* Copy number variants in patients with severe oligozoospermia and Sertoli-cell-only syndrome. *PloS one* **6**, e19426, doi:10.1371/journal.pone.0019426 (2011).
- 202 Reinius, B. & Jazin, E. Prenatal sex differences in the human brain. *Molecular psychiatry* **14**, 987, 988-989, doi:10.1038/mp.2009.79 (2009).
- 203 Speevak, M. D. & Farrell, S. A. Non-syndromic language delay in a child with disruption in the Protocadherin11X/Y gene pair. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **156B**, 484-489, doi:10.1002/ajmg.b.31186 (2011).
- 204 Nilsson, M. *et al.* Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085-2088 (1994).
- 205 Grundberg, I. *et al.* In situ mutation detection and visualization of intratumor heterogeneity for cancer research and diagnostics. *Oncotarget* **4**, 2407-2418 (2013).
- 206 Ahn, K. *et al.* Quantitative analysis of alternative transcripts of human PCDH11X/Y genes. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **153B**, 736-744, doi:10.1002/ajmg.b.31041 (2010).
- 207 Witelson, S. F. Hand and sex differences in the isthmus and genu of the human corpus callosum. A postmortem morphological study. *Brain : a journal of neurology* **112 ( Pt 3)**, 799-835 (1989).
- 208 Jobling, M. A. *et al.* Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. *Human molecular genetics* **16**, 307-316, doi:10.1093/hmg/ddl465 (2007).
- 209 Lagercrantz, H. & Ringstedt, T. Organization of the neuronal circuits in the central nervous system during development. *Acta Paediatr* **90**, 707-715 (2001).
- 210 Paridaen, J. T. & Huttner, W. B. Neurogenesis during development of the vertebrate central nervous system. *EMBO Rep* **15**, 351-364, doi:10.1002/embr.201438447 (2014).
- 211 Arnold, A. P. Sex chromosomes and brain gender. *Nat Rev Neurosci* **5**, 701-708, doi:10.1038/nrn1494 (2004).
- 212 McCarthy, M. M. & Arnold, A. P. Reframing sexual differentiation of the brain. *Nature neuroscience* **14**, 677-683 (2011).
- 213 Jazin, E. & Cahill, L. Sex differences in molecular neuroscience: from fruit flies to humans. *Nature reviews Neuroscience* **11**, 9-17 (2010).
- 214 Cahill, L. Why sex matters for neuroscience. *Nature reviews Neuroscience* **7**, 477-484 (2006).
- 215 Xu, J., Burgoyne, P. S. & Arnold, A. P. Sex differences in sex chromosome gene expression in mouse brain. *Hum Mol Genet* **11**, 1409-1419 (2002).

- 216 Xu, J. & Disteché, C. M. Sex differences in brain expression of X- and Y-linked genes. *Brain Res* **1126**, 50-55, doi:10.1016/j.brainres.2006.08.049 (2006).
- 217 Dewing, P., Shi, T., Horvath, S. & Vilain, E. Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. *Brain Res Mol Brain Res* **118**, 82-90 (2003).
- 218 Reinius, B. *et al.* Abundance of female-biased and paucity of male-biased somatically expressed genes on the mouse X-chromosome. *BMC genomics* **13**, 607, doi:10.1186/1471-2164-13-607 (2012).
- 219 He, Z., Bammann, H., Han, D., Xie, G. & Khaitovich, P. Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. *RNA (New York, N.Y.)* **20**, 1103-1111, doi:10.1261/rna.043075.113 (2014).
- 220 Morgan, C. P. & Bale, T. L. Sex differences in microRNA regulation of gene expression: no smoke, just miRs. *Biology of sex differences* **3**, 22, doi:10.1186/2042-6410-3-22 (2012).
- 221 Palazzo, A. F. & Lee, E. S. Non-coding RNA: what is functional and what is junk? *Frontiers in genetics* **6**, 2, doi:10.3389/fgene.2015.00002 (2015).

# Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations  
from the Faculty of Science and Technology 1285*

Editor: The Dean of the Faculty of Science and Technology

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

Distribution: [publications.uu.se](http://publications.uu.se)  
urn:nbn:se:uu:diva-261789



ACTA  
UNIVERSITATIS  
UPSALIENSIS  
UPPSALA  
2015