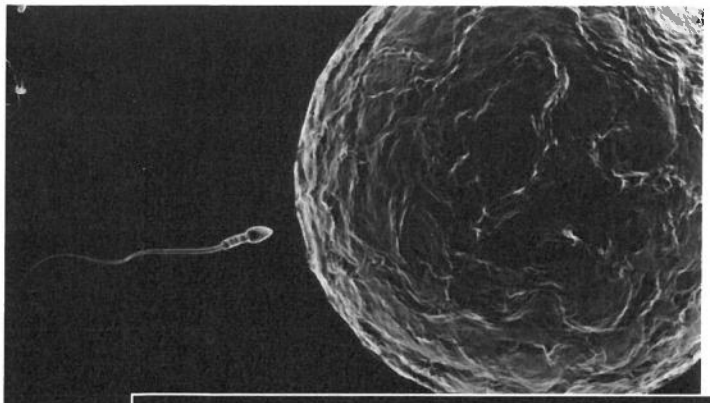


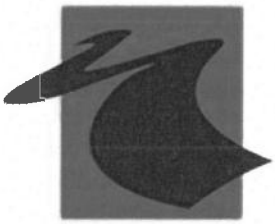
Prevalence of germ cell tumor susceptibility alleles in patients with disorders of sex development; thesis

Wendy de Klerk

Erasmus MC

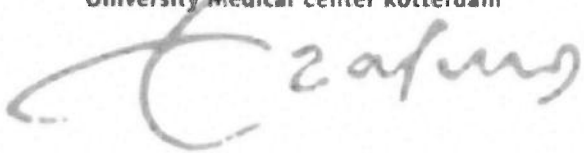
May 2016

Prevalence of germ cell tumor susceptibility alleles in patients with disorders of sex development; thesis



hogeschool Leiden

Erasmus MC
University Medical Center Rotterdam



Student:	Wendy de Klerk
Student number:	s1075211
Date:	13 May 2016
School:	Hogeschool Leiden
Address:	Zernikedreef 11, 2333 CK Leiden
Study:	Biology and medical laboratory research
Specialization:	Molecular Biology
Mentor:	Roel Schelland
Internship:	September 2015 – June 2016
Internship location:	Erasmus MC
Department:	Pathology, Laboratory for Experimental Patho-Oncology (LEPO)
Address:	Wytemaweg 80, 3015 CN Rotterdam
Supervisors:	Dr. R. Hersmus and Prof dr. L.H.J. Looijenga

Abstract

Disorders of Sex Development (DSD)-patients are patients with congenital conditions in which the development of chromosomal, gonadal or anatomical sex is atypical. Specific subgroups of patients with DSD have a higher risk to develop a malignant germ cell tumor (GCT). In genome wide association studies (GWAS)-studies it has been found that 17 single nucleotide polymorphisms (SNPs) are associated with testicular germ cell tumors (TGCT). These SNPs are all located in or near genes that influence embryogenesis, sex - , germ cell development and survival as well as DNA damage response. This study investigated the possible presence of a link between these SNPs and DSD-patients who developed a (T)GCT. This might lead to a more accurate prediction of tumor risk per patient. To answer the question, a cohort of 154 DSD patients was studied for all 17 SNPs using specific Taqman genotyping assays. For 16 SNPs the odds ratio (OR) was known from the GWAS studies. The OR for the 17th SNP was determined, using a TGCT group and a control group, resulting in an OR of 2.26. Using all 17 ORs a relative risk (RR) per DSD patient was calculated. To distinguish a difference between DSD-patients who did or did not develop a GCT statistical analysis was performed, showing a significant difference between the groups, although the groups overlapped. To analyze the distribution of the individual genotypes, the chi-square test was performed per SNP. This showed that there were several SNPs with an inverse distribution of the risk alleles when compared with the GWAS results. Excluding these SNPs from the RR analysis, increased significance was reached, although an overlap between the groups remained. A subset of DSD patients diagnosed with AIS were analyzed as a separate group, showing a significant difference between the non-affected and affected patients, although again overlap between the affected and non-affected groups was observed. By excluding four SNPS showing an inverse correlation when genotype distribution was compared with the GWAS studies, the separation between the two groups, and significance increased, resulting in absence of overlap. In conclusion, although analysis of the RR in DSD patients with and without a GCT shows a significant difference, the clinical relevance is debatable based on overlap between the groups. However, the same analysis in a sub-cohort of AIS patients shows a significant difference, after excluding four less relevant SNPs, there was a clear separation between the non-affected and affected groups. These 13 SNPs could be an additional risk-factor in the clinic.

Samenvatting

Disorders of Sex Development (DSD)-patiënten zijn patiënten met een congenitale conditie waarbij de chromosomale, gonadale of de anatomische geslachtsontwikkeling atypisch is. Bepaalde groepen patiënten gediagnostiseerd met DSD hebben een verhoogd risico op het ontwikkelen van een kwaadaardige kiemceltumor (KCT). In Genome Wide Association Studies (GWAS)-studies zijn 17 single nucleotide polymorfisme (SNP) gevonden die geassocieerd worden met een kwaadaardige testiculaire kiemceltumor (TKCT). Deze SNPs liggen in of nabij genen die invloed hebben op de embryogenese, geslachtsontwikkeling, kiemcelontwikkeling en overleving als ook een rol spelen in DNA herstel. In dit onderzoek wordt bestudeerd of er een link is tussen de gevonden SNPs en DSD-patiënten wie een (T)KCT hebben ontwikkeld bestudeerd. Dit kan resulteren in een nauwkeurigere risicobepaling voor het ontwikkelen van een KCT. Om deze link te vinden wordt er in een cohort van DSD-patiënten de genotypen van deze SNPs bepaald met behulp van genotyping-assays van Life Technologies. Van 16 van de 17 SNPs is de odds ratio (OR) al bepaald in de GWAS- studies. Voor de 17e SNP wordt de OR bepaald door een groep met TKCT te vergelijken met een groep gezonde mannen. Dit resulteert in een OR van 2.26. Met behulp van de 17 ORs wordt een Relative Risk (RR) per patiënt berekend in de DSD-groep. Per SNP werd de chi-kwadraat test uitgevoerd om te analyseren of er een verschil is in de verdeling van de genotypes. Hieruit blijkt dat een aantal SNPs een inverse verdeling lieten zien in de verdeling van de risico-allelen wanneer dit wordt vergeleken met de gepubliceerde GWAS resultaten. Wanneer deze SNPs worden geëxcludeerd in de RR analyse is er sprake van een meer significant verschil. Ondanks een stijgende significantie overlappen de groepen nog steeds. Verder is er gekeken naar een subpopulatie in de DSD populatie, de AIS patiënten. Deze groep is apart geanalyseerd en laat een significant verschil zien tussen de niet aangedane en de aangedane patiënten maar ook hier is er sprake van overlap tussen de groepen. Wanneer SNPs die een inverse correlatie, in vergelijking met de GWAS resultaten, laten zien worden geëxcludeerd in de RR analyse stijgt de significantie. Daarnaast wordt er ook een betere scheiding gevonden tussen de groepen. Dit alles houdt in dat de analyse van de RR in DSD patiënten met en zonder KCT een significant verschil laat zien, echter is het te betwisten of er sprake is van klinische relevantie omdat de groepen elkaar overlappen. Wanneer dezelfde analyse wordt uitgevoerd op een subgroep met AIS patiënten is er geen sprake van een significant verschil maar na het excluderen van vier SNPs, ontstaat er een goede scheiding tussen de niet aangedane en de aangedane patiënten. Deze 13 SNPs zouden een extra risicofactor kunnen zijn in de kliniek.

Table of Contents

List of abbreviations

AIS	androgen insensitivity syndrome
AMH	anti-Müllerian hormone
CAIS	complete androgen insensitivity syndrome
CI	confidence interval
DSD	disorders of sex development
EC	embryonal carcinoma
FFPE	formalin fixed paraffin embedded
GWAS	genome-wide association studies
GCNIS	germ cell neoplasia in situ
(T)GCT	(testicular) germ cell tumor
GB	gonadoblastoma
INSL3	insulin-like factor 3
ITNS	intratubular non-seminoma
ITSE	intratubular seminoma
LD	Linkage disequilibrium
LEPO	Laboratorium voor Experimentele Patho-Oncologie
NS	non-seminoma
OR	odds ratio
PAIS	partial androgen insensitivity syndrome
pBl	peripheral blood
PCR	polymerase chain reaction
PGCs	primordial germ cells
RR	relative risk
SE	seminoma
SNPs	single-nucleotide polymorphisms
KITLG	stem cell factor, officially named KIT ligand
TGCT	testicular germ cell tumors
TKCT	testiculaire kiemcel tumor
T	testosterone

Introduction

Background Literature

Normal gonadal development

Early during embryogenesis the bi-potential gonad develops under the influence of amongst others WT1, DMRT1, GATA4 and FGF9 (1). At this point in embryogenesis, sex cannot be phenotypically determined. During embryogenesis both the Müllerian and Wolffian ducts initially develop. In weeks 6-8 the bi-potential gonad will develop either into a testis or an ovary. When the Y-chromosome is present, a testis will develop under the influence of SRY, SOX9 and FGF9 (2). If the Y-chromosome is absent the bi-potential gonad will, under influence of WNT4, FOXL2 and RSPO1, develop into an ovary (2). When a testis develops, the Leydig cells start producing testosterone (T) and insulin-like factor 3 (INSL3). These factors will maintain the Wolffian ducts. Besides T and INSL3, Sertoli cells produce anti-Müllerian hormone (AMH), which will lead to regression of the Müllerian ducts. When an ovary is formed T and AMH will not be produced, leading to regression of the Wolffian duct, whilst the Müllerian duct persists and will develop further. The Müllerian ducts will give rise to the female internal genitalia, the Wolffian ducts will form the male internal genitalia. A schematic overview of male and female development is given in Figure 1.

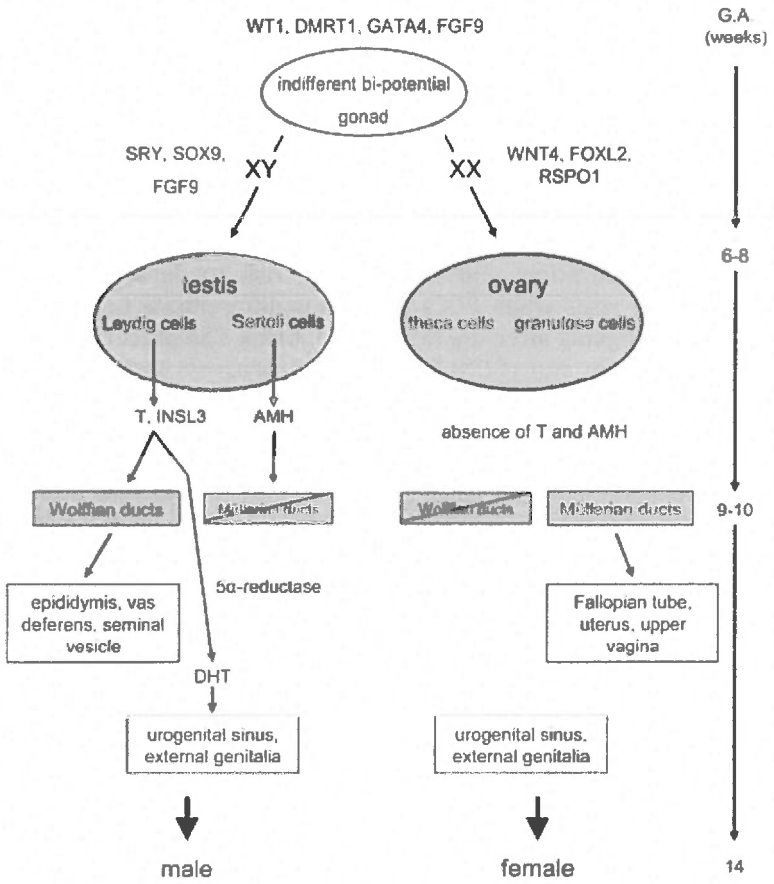


Figure 1: Schematic overview of male and female development. Male development: Under influence of SRY, located on the Y-chromosome, and SOX9 and FGF9 a testis will be formed. Leydig cells present in the testis will produce T and INSL3, and Sertoli cells will produce AMH. This maintains the Wolffian ducts and induces regression of the Müllerian ducts respectively. Female development: Under influence of WNT4, FOXL2 and RSPO1 the ovary will be formed. Because of the absence of T and AMH the Wolffian duct will regress and the Müllerian duct persists (3).

Germ cells

After a sperm cell fertilizes the ovum, these cells combine to form the zygote, which after one and a half days will start dividing. After two to three days this clump of cells is called a blastocyst. In the epiblast of the blastocyst the primordial germ cells (PGCs) form. The PGCs are the precursor cells of the gametes (ova and sperm as found at adult life). After approximately three weeks the PGCs will migrate through the hindgut, to the bi-potential gonads. Normally PGCs which do not reach the gonad will go into apoptosis. Stem cell factor, officially named KIT ligand (KITLG) and its receptor play an important role in the migration, proliferation and the survival of PGCs during migration and development into gametes (4). When the PGCs reach the bi-potential gonad they are called gonocytes. Depending on the presence or absence of the Y-chromosome the gonads will develop into testes or ovaries, respectively (see above). If a testis is formed the gonocytes will differentiate into pre-spermatogonia. When an ovary forms, the gonocytes will eventually differentiate into primary oocyte. A secondary oocyte will differentiate every month from puberty till menopause. Spermatogenesis (differentiation of a spermatogonium into spermatozoa) will start during puberty under influence of T.

Precursor lesions and invasive germ cell tumors

There are in total five different types of GCTs, but in this thesis only the type II GCTs will be discussed, as these are relevant in the context of DSD. A malignant type II germ cell tumor (GCT) arises from a PGC/gonocyte that does not properly mature, does not go into apoptosis and preserves its intrinsic pluripotent characteristics. The majority of GCTs will develop in the gonads. The tumor displays characteristics which PGCs/gonocytes have and would normally lose upon their differentiation into pre-spermatogonia. Although they still can be present in the first year after birth (2). To investigate the maturation/malignant state of the PGCs/gonocytes the markers OCT3/4, KITLG and TSPY can be used. OCT3/4 and KITLG are embryonic markers, and normally expression of these markers is lost in the germ cells during the first year after birth. When expression of these markers is still present after the first year, it is an indication that the PGCs/gonocytes are arrested/delayed in their maturation, and might be at risk to develop into a GCT precursor. Expression of TSPY normally rises when PGCs/gonocytes differentiate towards pre-spermatognia, and stays present in spermatogonia after the first year of birth. The precursor lesion of a GCT is the so-called germ cell neoplasia in situ (GCNIS). GCNIS can progress into the invasive components seminoma (SE) or non-seminoma (NS). Sometimes both tumor types occur in one gonad. The term NS is used for embryonal-carcinoma (EC), and its differentiated components teratoma, yolk-sac tumor and choriocarcinoma. An overview is given in Figure 2 (5). Type II GCTs can also develop out of the precursor lesion gonadoblastoma (GB). A GB develops mostly into dysgerminoma. SE and dysgerminoma are indistinguishable from each other on DNA, mRNA, miRNA and marker expression (2). Both precursor lesions express the same embryonal markers (OCT3/4, C-KIT, KITLG and TSPY) (2). In DSD patients both GB and GCNIS can occur, depending on the degree of testicularization. See also DSD and germ cell tumors (below).

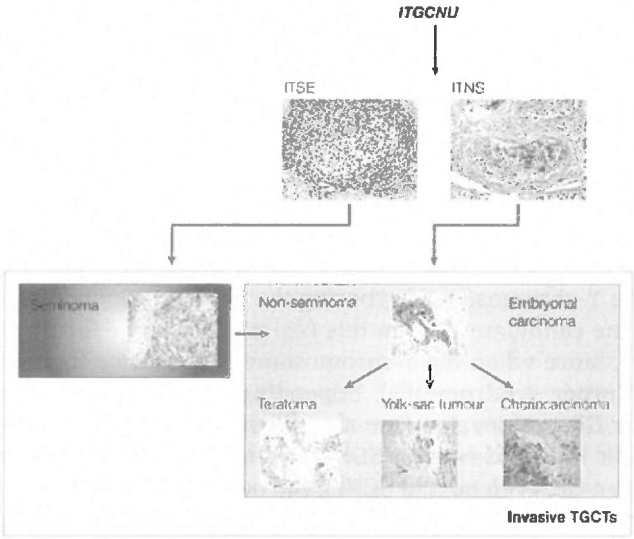


Figure 2: Developmental stages of GCT type II. The precursor lesion GCNIS (here referred as intratubular germ cell neoplasia unclassified (ITGCNU) can develop into intratubular seminoma (ITSE) and become invasive or into a intratubular non-seminoma (ITNS) and become an invasive embryonal carcinoma. Embryonal carcinoma can differentiate into teratoma, yolk-sac tumor or choriocarcinoma. (5)

GCTs are relatively rare, they make up 1% of the solid tumors in Caucasian men (6). However testicular germ cell tumors (TGCT) comprise 60% of all malignant tumors diagnosed in Caucasian men in the age of 20 to 40 years. In Europe every year around 12.000 patients will be diagnosed with a TGCT, and the incidence is still rising. In 2014 around 800 patients were diagnosed in the Netherlands. GCTs are, in most cases, well treatable by surgical removal of the tumor after which chemo- and/or radiotherapy can be given. However these treatments can lead to infertility, cardiovascular diseases, secondary tumors and cognitive impairment. The 5-year survival rate has been upwards of 92% since 1989 (7).

In genome-wide association studies (GWAS) (8, 9, 10, 11, 12) single nucleotide polymorphisms (SNPs) related to TGCT have been found. These SNPs occur more often in patients with a TGCT then in men without a TGCT. In this study the 17 different SNPs, found in the GWAS, are determined on DSD patients. In the future it might be possible to predict tumor risk in DSD patients based on the presence of these SNPs.

All 17 SNPs are located in or near genes that influence embryogenesis, sex - , germ cell development and survival as well as DNA damage response. For 16 out of the 17 SNPs an odds ratio (OR) has been determined. The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of the exposure (13). In this study the outcome is developing or not developing a precursor lesion/invasive tumor. If the OR is higher than 1, it is more likely to develop the disease if this patient scored positive for the risk factor. In this case, if a patient has a specific SNP he is more likely to develop a TGCT. The 17th SNP is located in the functional TP53-binding site of *KITLG* and influences transcription (9). The ability of TP53 to regulate transcription is crucial for tumor suppression, as TP53 conserves DNA stability by preventing the accumulation of possibly damaging mutations in the genome (14). This might indicate that polymorphisms in this binding site could influence cancer (9) but no OR has been determined.

Disorders of sex development (DSD)

There are various forms of disorders of sex development (DSD). DSD are defined as congenital conditions in which development of chromosomal, gonadal or anatomical sex is atypical. DSD can manifest itself in various forms: cryptorchidism (undescended testis), a male genotype with a female phenotype and vice versa, and ambiguous genitalia, a condition where an infant's external genitalia do not appear to be obviously male or female. The new DSD classification, in use since 2006, is based

on the chromosomal constitution, and consists of three main groups, namely: 46XY-DSD, 46XX-DSD and sex chromosome-DSD, an overview is given in Appendix 1. In this overview terms like intersex, pseudohermafroditism, hermaphroditism which were confusing for caregivers and confronting for patients were abolished (15).

DSD and germ cell tumors

Despite the low prevalence of type II GCTs in the whole population, in specific forms of DSD the occurrence is high (2). Risk factors for developing a type II GCT in DSD are: gonadal dysgenesis and presence of (a part of) the Y-chromosome, harboring the so called GonadoBlastoma on the Y (GBY) region, with *TSPY* being the candidate gene in this region (16). Some forms of DSD do not have an increased GCT risk, for instance when the Y-chromosome is absent or there are no PGC/gonocytes present, e.g. 45,X DSD (Turner syndrome)(3). Depending on the differentiation of the gonad, the precursor lesion GCNIS or GB can develop (see also above), and they can even occur in one gonad. The supporting Sertoli cells in GCNIS express SOX9. When testis formation (testicularization) of the gonad is not complete also a GB, even next to GCNIS, can occur with predominantly granulosa cells as supportive cells, indicated by expression of FOXL2. As described earlier both precursor lesions can develop into an invasive GCT. Figure 3 shows the embryogenesis and pathogenesis of DSD. The set of SNPs associated with TGCT discussed above could be an additional risk factor for developing a GCT in patients with DSD. This study is focused on finding a relation between these SNPs and DSD patients at risk of developing a type II GCT.

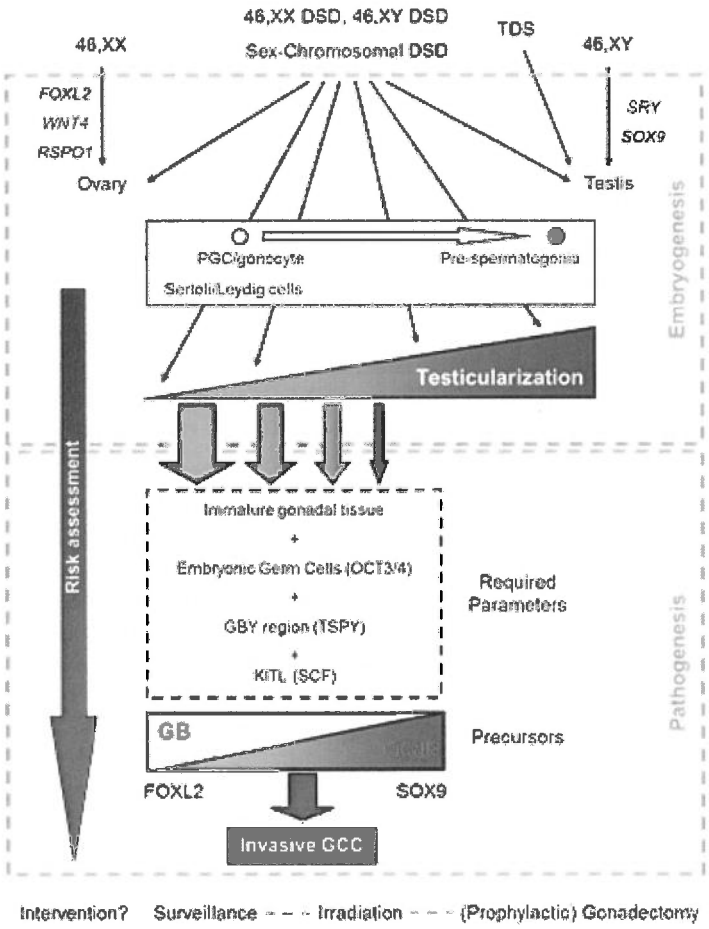


Figure 3: Embryogenesis and pathogenesis of DSD, on the left and right side normal female and male development is depicted and in between the various forms of DSD are shown. If testicularization is not complete, and the required parameters are present there is an elevated risk to develop a precursor lesion (either GCNIS or GB) and subsequently an invasive tumor. (2)

Aim of this study

The aim of this study is to get better insight in GCT risk in patients with DSD. To achieve this a defined set of 17 SNPs known to be associated with TGCT risk is investigated in a group of DSD patients with and without a GCT and/or precursor lesion. It is known that a subset of DSD patients have an increased risk to develop a GCT (17), and analysis of these SNPs, in addition to the other known risk factors (see Figure 3), might be an addition in (early) detection of GCT risk, possibly leading to a more personalized treatment. If patients have a low risk an orchidectomy is not needed and less screening might be necessary. If patients have a high risk, preventive measures such as orchidectomy can be taken, as this is the current treatment.

The hypothesis is that analysis of these 17 SNPs in DSD patients, can possibly distinguish between patients who will and who will not develop a TGCT.

Experimental design

The setup of this study is as follows: In this study 191 DSD patients are included, 28 patients were diagnosed with a form of 46,XX-DSD, and 163 patients had a form of 46,XY-DSD, patients with a chromosomal DSD were excluded. In the group of patients the genotype of the 17 SNPs shown to be associated with a higher TGCT risk will be determined. DNA from the patients is isolated from peripheral blood (pBL) or, when this is not available, from formalin fixated paraffin embedded (FFPE) or frozen tissue.

For 16 out of the 17 SNPs the OR has been determined in the GWAS studies, the RR these OR infer will be calculated using the method described by Kratz *et al.* (3). In brief: if a patient is homozygous for the non-risk allele of a SNP the OR will be multiplied with zero, is a patient heterozygous, the OR will be multiplied with one and if the patient is homozygous, the OR will be multiplied with two. This will be done for all 16 SNPs and the total sum of these multiplied ORs determines the RR. Determining the genotypes will be done using Taqman SNP genotyping assays from Life technologies.

For the SNP present in a TP53 binding site in KITLG an OR has not yet been determined. To do so the genotype of this SNP will be assessed in a group 698 men with a TGCT versus a group without a TGCT (control group). In this case the data for the control group comes from the 1000 genomes project (18). When determined the OR will be included in the RR calculations.

For statistics IBM SPSS statistics 21 is used. The Independent t-test, chi-square test and the OR is determined with SPSS.

Background techniques

DNA isolation with phenol/chloroform

DNA isolation using phenol/chloroform is based on the principle of polarity. The different phases originate because of the density-differences between water and phenol/chloroform. Phenol/chloroform is an apolar liquid while water is polar. DNA is polar and is soluble in water, in contrast to proteins, which are or will become apolair while shaking in this water and phenol/chloroform solution. After centrifugation DNA will be in the upper phase and the proteins and cell debris will be in the phenol/chloroform-phase (19, 20) . This process is shown in Figure 4.

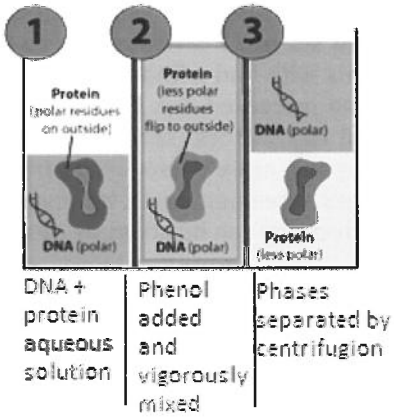


Figure 4: DNA isolation using phenol/chloroform to separate the DNA from the proteins and cell debris.(20)

Genotyping-SNP assays

TaqMan® 5'-nuclease assay chemistry provides a fast and simple way to get SNP genotyping results. Each predesigned TaqMan® SNP Genotyping Assay includes two allele-specific TaqMan® MGB (minor groove binder (21)) probes containing distinct fluorescent dyes (VIC and FAM, both having different emission wavelengths) and a Polymerase chain reaction (PCR) primer pair to detect specific SNP targets. These TaqMan® probe and primer sets (assays) uniquely align with the genome to provide unmatched specificity for the allele of interest (22). The advantage of using MGB probes is that they stabilize the hybridized probe and raises the melting temperature. This results in a short probe (23). In Figure 5 the technical process during this real time PCR is shown. If the probe binds to the DNA the fluorescent label (reporter, VIC/FAM) is near the non-fluorescent quencher which ensures that light cannot be emitted. During polymerization the reporter will be cleaved, resulting in a large enough distance between reporter and quencher so the report can emit light.

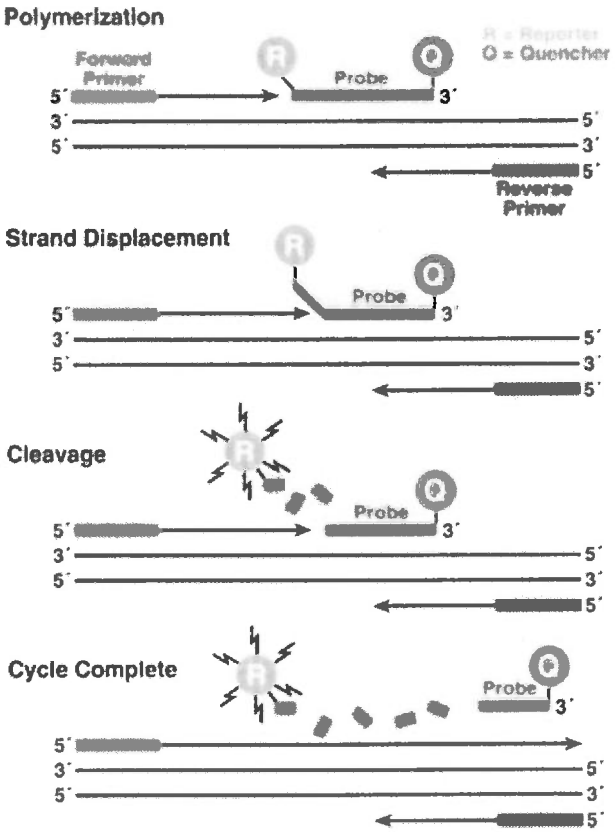


Figure 5: technical view of genotyping assay PCR (24)

Material and methods

Patient Cohort

In this study 191 DSD patients are included. 28 patients have 46,XX-DSD and 163 have 46,XY-DSD. These patients have different forms of DSD, see Appendix 1 for the subgroups in DSD. Chromosomal DSD variants were excluded in this study, this because the chromosomal changes most likely are the leading cause of the DSD.

DNA Isolation

Depending on the type of patient material available different DNA isolation protocols are used. Isolation of DNA from FFPE material was initially performed according to protocol 1, but the DNA isolated yielded sub-optimal results in the genotyping assays. Therefore the isolation protocol has been optimized from (25), see protocol 2. The different steps to determine this optimal isolation protocol are described in attachment 3. Isolation of pBL is performed with protocol 3, frozen tissue is performed using protocol 4.

Protocol 1 DNA isolation, FFPE tissue

In short, FFPE material will first be deparaffinized using xylene, then the tissue is lysed overnight at 52°C using a lysisbuffer with (10 mM Tris.Cl, 100mM NaCl, 5mM EDTA, 1% SDS, 1mM $CaCl_2$) and proteinase K (10 mg/ml). The next day a phenol/chloroform step is performed, and the DNA is precipitated. The DNA will be dissolved in $T_{10}E_1$ (10 mM Tris.Cl. 1 mM EDTA). Work protocol is attached in appendix 2.

Protocol 2 DNA isolation, FFPE tissue

Briefly, FFPE material will first be deparaffinized using xylene, then the tissue is lysed using a universal buffer (barbital, citric acid, KH_2PO_4, boric acid, all 28.6 mM and pH 10) for 30 min at 100 °C, cooled down to 52 °C and proteinase K is added. The samples will be lysed overnight at 52°C. The next day a phenol/chloroform step is performed. After this DNA will be precipitated, and the DNA will be dissolved in $T_{10}E_1$. Work protocol is attached in appendix 2.

Protocol 3 DNA isolation, pBl

To summarize, from total anticoagulated blood white blood cells will be isolated, using ammonium or bicarbonate to break down erythrocytes. After centrifugation the pellet containing white blood cells can be dissolved in nucleus-lysis buffer (Tris 1.21g; NaCl 23.4g; EDTA 0.74g in 1 L H_2O, pH 8.2). Sodium dodecyl sulfate (SDS) denatures all present proteins. By adding protease the denatured proteins will break down. Using saturated NaCl the protein compounds will be precipitated and removed with centrifugation. From this supernatant the DNA will be precipitated with ethanol. The DNA will be dissolved in $T_{10}E_1$. Work protocol is attached in appendix 2.

Protocol 4 DNA isolation, Frozen tissue

In short, from frozen tissue DNA will be extracted with use of the DNeasy kit (Qiagen, Venlo). This kit is designed for rapid purification of total DNA. The DNeasy membrane combines the binding properties of a silica-based membrane. DNA adsorbs to the DNeasy membrane in the presence of high concentrations of chaotropic salt, which remove water from hydrated molecules in solution. Buffer conditions in DNeasy Blood & Tissue procedures are designed to enable specific adsorption of DNA to the silica membrane and optimal removal of contaminants and enzyme inhibitors (26). Work protocol is attached in appendix 2.

Genotyping-SNP assays

All Taqman genotyping-SNP assays are designed and manufactured by Life technologies (Bleiswijk, the Netherlands) and have been validated. The SNPs used in this study are shown in appendix 5.

For these assays TaqMan® Universal PCR Master Mix, no AmpErase® UNG (ThermoFisher Scientific, Bleiswijk, the Netherlands) was used, to which dH2O and the genotyping SNP assays (ThermoFisher Scientific) will be added.

The mastermix will be pipetted in a 96-wells plate using 13 µl per well, then per well 2 µl DNA sample is added, with a concentration of 25 ng/µl.

For each SNP assay three positive controls are added: one heterozygote, one homozygote non-risk allele, one homozygote risk allele and one negative control. As an example, in Figure 6 the amplification curves for all controls of one genotyping assay are shown. For every plate the threshold is set on 0.02 arbitrarily. Below this threshold a sample cannot be interpreted. Determining the genotype for a patient is done by comparing with the positive controls, this is done by the software package, but all amplification curves are checked to prevent software mistakes. If the no template control comes above the set threshold none of the assayed samples can be interpreted

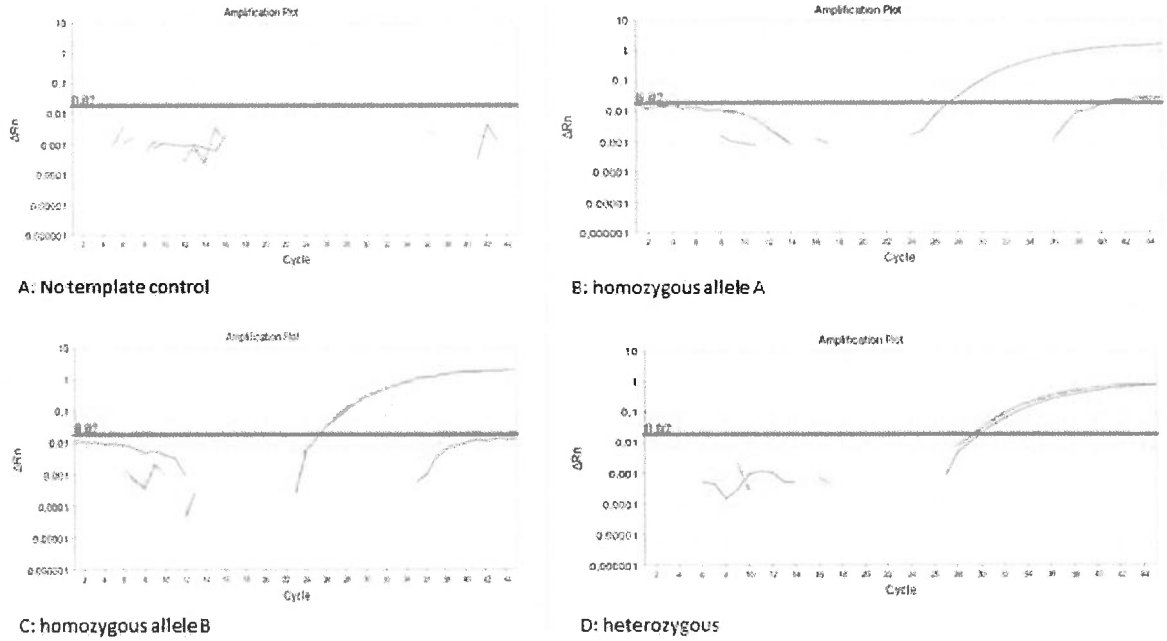


Figure 6: the four controls used for genotyping SNP assays, A: no template control, B: homozygous control for allele A, C: B: homozygous control for allele B, D: heterozygous

After pipetting a plate, the plate will be run on a AB 7500 Fast Real-Time PCR machine (ThermoFisher Scientific) according to the following program: 60.0°C for 1 min, 95.0°C for 10 min, 45 times: 95.0°C for 15 sec and 60.0°C for 1 min. following with 1 min 60.0°C.

The work protocol can be found in attachment 4

Odds ratio

For SNP rs4590952 no OR had been determined. This is determined using a group of men with a TGCT versus a control group without a TGCT. For all the men with a TGCT the genotype for this SNP is determined using the work protocol: genotyping-SNP assay in appendix 4. With these results the OR is calculated, using publicly available reference data from the 1000 genomes database.

Before it is possible to calculate a OR, the allele frequency is needed. Below in Table 1 an example is shown. Nine patients have the genotype AA, 15 patients with AG and 13 with GG. The allele frequency is then A: 33 (44.6%) and G: 41 (55.4%)

Table 1: Example, allele frequency

	No of patients	A allele	G allele
AA genotype	9	18	0
AG genotype	15	15	15
GG genotype	13	0	26
Total alleles		33	41

To calculate the OR a two-by-two frequency table is used. The outcome status is positive for patients with a TGCT and negative for the control group without a TGCT. The exposure status is positive for the risk allele G and negative for the A allele. In Table 2 below, a scheme is shown to calculate the OR. (13)

Table 2: how to calculate a OR (13)

		Outcome Status +	Outcome Status −
Exposure status	+	a	b
	−	c	d

With the data in Table 2 the following formula is used: $OR = \frac{ad}{bc}$

Statistics SPSS

In the overview below (Figure 7) a flowchart is shown, depicting which statistical test is needed for different kind of data. This flowchart can only be used for hypothesis tests.

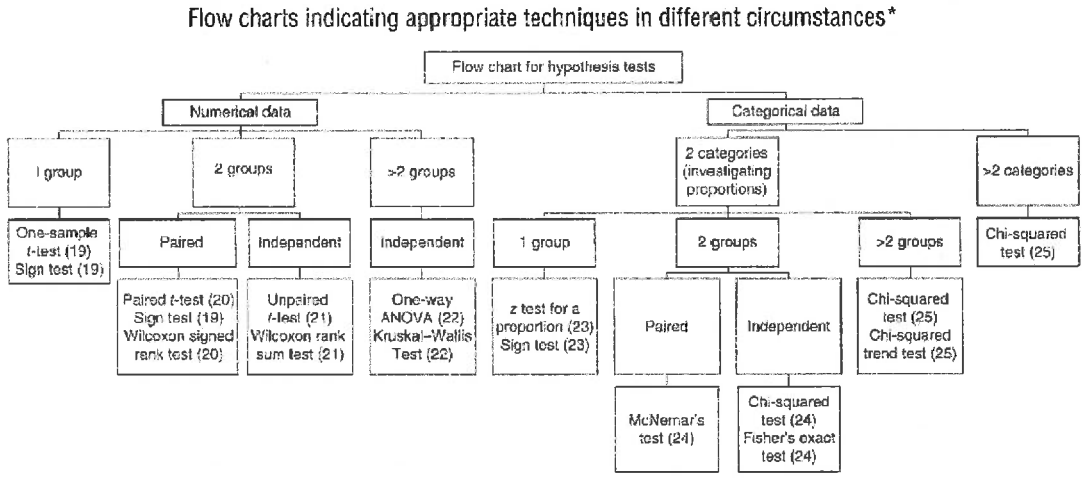


Figure 7: statistical analysis flowchart (27)

In this thesis the independent t-test and the Chi-square test are used for statistical analysis. Below these tests will be discussed in more detail.

Independent t-test

The t-test can only be used when numerical data is available and the variable is normally distributed in the groups. (27) This test is used when analyzing the RR or the OR per SNP. In this thesis the cohorts are divided into two groups, one without a GCT and the other group with a GCT or a precursor lesion. Because there a two different groups the independent t-test is chosen to do statistical analysis.

If these two groups can be compared, the Levene's test is performed in SPSS. When the outcome is not significant (P>0.05), the two groups are comparable (upper row, equal variances assumed). When there is a significant difference (P<0.05) between these groups, this data is included in the t-test and a different p-value is calculated (lower row, equal variances not assumed). An example of the SPSS output data is shown in Table 3. (28)

Table 3: example of the data gained with SPSS using the independent t-test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% CI	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	,002	,963	2,622	130	,010	2,75470	1,05068	,67607	4,83334
Equal variances not assumed			2,594	44,349	,013	2,75470	1,06189	,61508	4,89432

Chi-square test

This test is used when comparing categorical data, for example number of patients in two or more categories. Using this test with two groups, the following conditions need to be met: if any expected counts are less than 5, the Fischer exact test will be used (27, 29, 30). A p-value is given in SPSS.

When comparing more than 2 groups using the Chi-square test, a maximum 20% of the cells is allowed to have a count less than 5 (31). Because the Chi-square is mostly performed comparing >2 groups an example of this is shown below.

All measurements are divided per category and per category a total is calculated, see Table 4.

Table 4: measured values divided in categories.

Measured values (M)	No precursor/invasive tumor present	precursor/invasive tumor present	Total
Homozygous non-risk allele	31	20	**51**
Heterozygous	25	28	**53**
Homozygous risk allele	23	34	**57**
total	**79**	**82**	161

With these totals the expected values are calculated by multiplying the totals per category and dividing them by the total number of measurements. See Table 5

Table 5: expected values based on the measured values shown in Table 4

Expected values (E)	No precursor/invasive tumor present	precursor/invasive tumor present	Total
Homozygous non-risk allele	**(51x79)/161 = 25**	**26**	51
Heterozygous	**26**	**27**	53
Homozygous risk allele	**28**	**29**	57
total	79	82	161

In Table 5 none of the expected values are below 5, and the Chi-squared can be used in this case.

Results

Patient Cohort

This cohort started with 191 DSD patients (Dutch, Belgian and Czech patients), 19 were excluded because not enough material was available (absent or biopsy material). During the study for 18 not all 17 genotyping-SNP assays could be determined. This resulted in a total of 154 patients, see overview below, Figure 8. All (clinically) relevant patient information can be found in appendix 7.

Of these 191 samples, 134 samples were isolated from FFPE material, two from frozen tissues and 55 from pBl samples. One of the 18 samples which were incomplete due to technical issues was isolated from pBl , the rest was FFPE material.

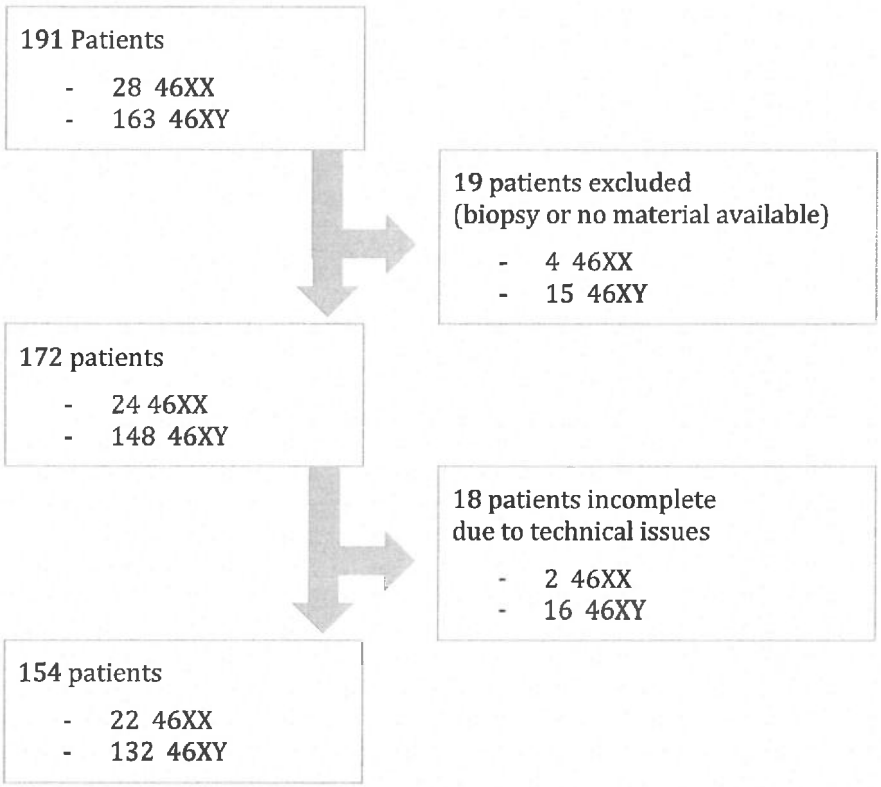


Figure 8: overview of patient cohort

This DSD-cohort is divided in two groups; a total cohort and a subgroup of DSD patients diagnosed with androgen insensitivity syndrome (AIS), the AIS cohort. For both cohorts, statistical analyses are performed.

As a first step the OR for SNP rs4590952 had to be determined, as this was needed for further statistical analysis of the RR. Hereafter the results of the genotype-SNP assays are shown, followed by the statistical analyses for the total and AIS cohort.

rs4590952; OR and allele frequencies in SE and NS

OR rs4590952

As there was no OR known for the SNP found in the p53 binding site of KITLG, rs4590952, which is needed for further analysis of the RR, this had to be determined. To this end, a group of 698 Caucasian patients with a TGCT were selected, from which 15 patients were excluded because of a lack of material. The remaining cases were analyzed using the specific Taqman genotyping assay. For 644 patients a genotype could be determined, the results are shown in Table 6. For 39 patients SNP could not be determined due to technical issues. For a control group data from the 1000 genomes project, phase 3 EUR (European) male population was used (summarized in Table 7), this because the TGCT population analyzed was predominantly of European descent.

Table 6: genotypes TGCT group

	Total	AA	AG	GG
Patients	644	7	113	524
Percentage	100	1.1	17.5	81.4

Table 7: genotypes 1000 genomes, phase 3 EUR population (18)

	Total	AA	AG	GG
Patients	237	14	66	157
Percentage	100	5.9	27.8	66.2

Based on these results the allele frequency for rs4590952 was calculated. In the TGCT group the allele frequency of the G allele is 90.1% and the frequency of the A allele is 9.9%. In the control-group the allele frequency for the G allele is 80.2% and for the A allele 19.8%. In attachment 6 the raw data of the allele frequencies is shown.

With this raw data SPSS calculates the OR, the 2x2 table to calculate the OR is shown in attachment 6. Rs4590952 has an OR of 2.26, 95% CI(1.69-3.02). This OR is included in the results of the DSD cohorts (total and AIS) shown below

Allele frequencies in SE and NS

To see if there is a significance difference in allele frequency for rs4590952 between SE and NS. The TGCT group of 644 patients is divided into these two groups. In total, 382 patients have a SE, 255 patients have a NS, and 36 patients have a combined tumor (CT), meaning that both SE and NS is present. Following the WHO classification of 2004 (32), when a tumor has a NS component, independent of a SE component, the lesion is classified as a NS. Therefore, the analysis is done by including the CT in the NS group and secondly excluding them entirely.

In this case the data is categorical, so the Chi-square test is used. The p-value of the Chi-square test is 0.144 (2-sided). Raw data is shown in appendix 6. In Table 8 the allele frequencies for NS and SE is shown, CT are included as NS.

Table 8: Allele frequencies NS and SE, CT included as NS

		allele	
		A	G
TGCT	NS	10%	90%
	SE	7.9%	92.1%

Because no significance was reached the CT are excluded in this analysis. In Table 9 the allele frequencies of SE and NS are shown, CT are excluded. The p-value of the Chi-square test is 0.792 (2-sided). Raw data is shown in appendix 6. So in both analyses there is no significant difference in allele frequencies between SE and NS.

Table 9: Allele frequencies NS and SE , CT excluded

		allele	
		A	G
TGCT	NS	7.5%	92.5%
	SE	7.9%	92.1%

Genotyping assays

In total on 172 patients the genotypes for the 17 SNPs were analyzed using the Taqman genotyping assays. A significant number of genotyping-SNP assays did not give interpretable results. The DNA used had been isolated with protocol 1, which yielded DNA with a sub-optimal quality, being (highly) degraded. To increase the quality of the DNA, protocol 2 was developed, which resulted in enhanced genotyping results, more samples gave interpretable results of the genotyping-SNP assays. However, still eight genotyping assays showed amplification problems, most likely due to the size of the amplicon (200bp or larger) in relation to the size of the isolated DNA. For these a custom made assay was ordered from Life Technologies (Bleiswijk, The Netherlands) with a maximum product length of 150 bp. This resulted in more amplifiable fragments. To further optimize the assays the amount of added DNA was reduced from 25 ng/µl, to 12.5 ng/µl and even further diluted to 6 ng/µl, thereby reducing the amount of inhibitors in the assay. With the help of these optimizing-steps it was possible to determine all 17 SNPs in 154 patients. The results of the genotyping assays is shown in attachment 8.

RR analysis of the total DSD cohort

The total DSD population consists of 154 patients, this population is divided in two groups. The first is the non-affected group (n=123), this includes DSD-patients without a precursor lesion (GCNIS) or an invasive tumor. The second is the affected group (n=31), being DSD-patients with GCNIS or an invasive tumor present. This population is analyzed using the RR of 17 SNPs. In Table 10 a p-value of 0.766 is shown for the Levene's test. This indicates for a normal distribution and the independent t-test can be performed. The t-test results in a p-value of 0.039. This means there is a significant difference between the groups affected and non-affected patients in this DSD-population.

Table 10: RR analysis, t-test results, groups are divided in affected (n=123) and non-affected (n=31).

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% CI	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RR of 17 SNP	Equal variances assumed	,017	,896	2,083	152	,039	2,10835	1,01204	,10887	4,10783
	Equal variances not assumed			1,998	44,069	,052	2,10835	1,05547	-,01871	4,23541

A boxplot of this cohort is shown in Figure 9, this boxplots represents the distribution of the RR between the two groups. Overlap between the groups is present.

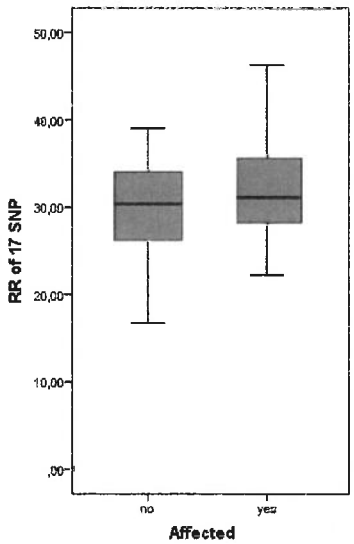


Figure 9: boxplot, RR of 17 SNPs, total cohort n=154 divided in affected and non-affected

Because 46-XX DSD patients seldom develop a TGCT, they are excluded for further analyses (n=22). This results in a total 46-XY DSD (n=132) population which is divided in two groups as before in a non-affected (n=102), and an affected group (n=30). Analyzing the RR on all 17 SNPS showed that there was a significant difference (p=0.014) between the non-affected and the affected group, although there was considerable overlap between the two groups, Table 11 and Figure 10 (green boxplot). Significance rises when excluding 46XX-DSD patients overlap between affected and non-affected is still present.

Per SNP the Chi-square test is performed, to investigate if individual SNPs differ significantly between the groups. All data of the Chi-square test is shown in appendix 9. The SNP rs210138 (nearby *BAK1*) shows a significant difference (p=0.011) but the distribution between the groups is inverse in comparison with the GWAS-studies; in the affected population the homozygous non-risk allele is more prevalent. To reach a higher significance and a better separation between the two groups this SNP is excluded from the RR analysis, see below.

Excluding rs210138 resulted in a higher significance (p=0.002) Table 11 and Figure 10 (blue boxplot). Analyzing all the Chi-square test results, three more SNPs showed an inverse distribution between the two groups: rs4699052, rs2736100 and rs7040024. Excluding all four SNPs (also taking rs210138 into account) increased significant difference further (p=0.001), Table 11 and Figure 10 (red boxplot). As the significance in all three groups for the Levene's test are above 0.05 equal variances are assumed.

Table 11: t-test results, groups are divided in affected and non-affected.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% CI	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RR of 17 SNPs	Equal variances assumed	,204	,652	2,496	130	,014	2,56506	1,02768	,53192	4,59820
	Equal variances not assumed			2,374	44,198	,022	2,56506	1,08036	,38801	4,74211
RR without BAK1	Equal variances assumed	,240	,625	3,118	130	,002	3,13506	1,00520	1,14440	5,12571
	Equal variances not assumed			2,949	43,913	,005	3,13506	1,06302	,99255	5,27756
RR without 4 SNPs	Equal variances assumed	,067	,795	3,525	130	,001	3,30767	,93831	1,45132	5,16401
	Equal variances not assumed			3,471	46,340	,001	3,30767	,95292	1,38993	5,22541

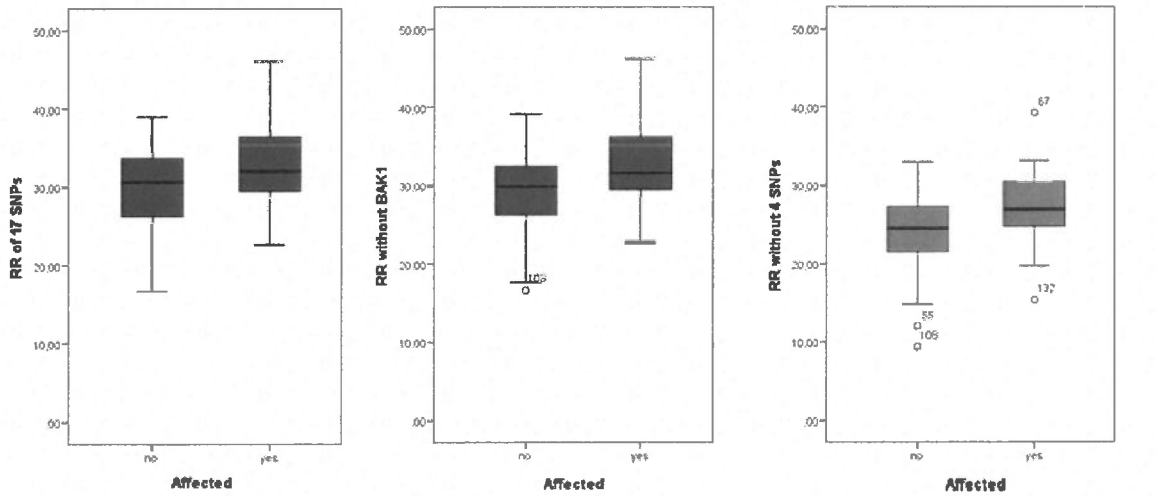


Figure 10: boxplots, RR for affected and non-affected patients. Green boxplots show the sum of 17 SNP resulting in a RR. The blue boxplots show the sum of 16 SNP were the SNP rs210138 (BAK1) is excluded. Outliers are present (dots). In the red boxplots the 4 SNPs are excluded which divide the two groups (affected and non-affected) poorly. Outliers are present (dots).

RR risk analysis in the AIS cohort

The AIS cohort includes 53 46-XY DSD patients, of which seven have a GCNIS. As done for the total cohort, per SNP genotype the Chi-square test was performed. All data is shown in Appendix 10.

When performing the Chi-square test, the results show that more than 20% of the expected values have a value below 5, so the test cannot be used. Instead the distribution of the genotypes within the two groups has been analyzed.

Analyzing the RR on all 17 SNPS showed that there was a significant difference (p=0.023) between the non-affected and the affected group, Table 12 and Figure 11 (green boxplot). As in the total 46XY-group, BAK1 (rs210138) shows an inverse distribution between the groups. This SNP is excluded from the RR analysis. Excluding rs210138 resulted a significance difference between the groups (p=0.015) Table 12 and Figure 11 (blue boxplot). Analyzing all the Chi-square test results, three more SNPs showed an inverse distribution between the two groups: rs4888262, rs1508959 and rs4590952. Excluding all four SNPs (also taking rs210138 into account) increased significant difference further (p=0.000), Table 12 and Figure 11 (red boxplot). As the significance in all three groups for the Levene's test are above 0.05 equal variances are assumed.

Table 12: t-test results, groups are divided in affected and non-affected.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% CI	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
RR 17 SNPs	Equal variances assumed	1,774	,189	2,344	51	,023	4,97062	2,12067	,71320	9,22804
	Equal variances not assumed			1,703	6,786	,134	4,97062	2,91838	-1,97456	11,91580
RR without BAK1	Equal variances assumed	2,811	,100	2,525	51	,015	5,21311	2,06498	1,06749	9,35872
	Equal variances not assumed			1,730	6,660	,129	5,21311	3,01294	-1,98566	12,41187
RR without 4 SNPs	Equal variances assumed	2,426	,126	5,193	51	,000	7,34702	1,41468	4,50694	10,18710
	Equal variances not assumed			3,835	6,824	,007	7,34702	1,91601	2,79256	11,90148

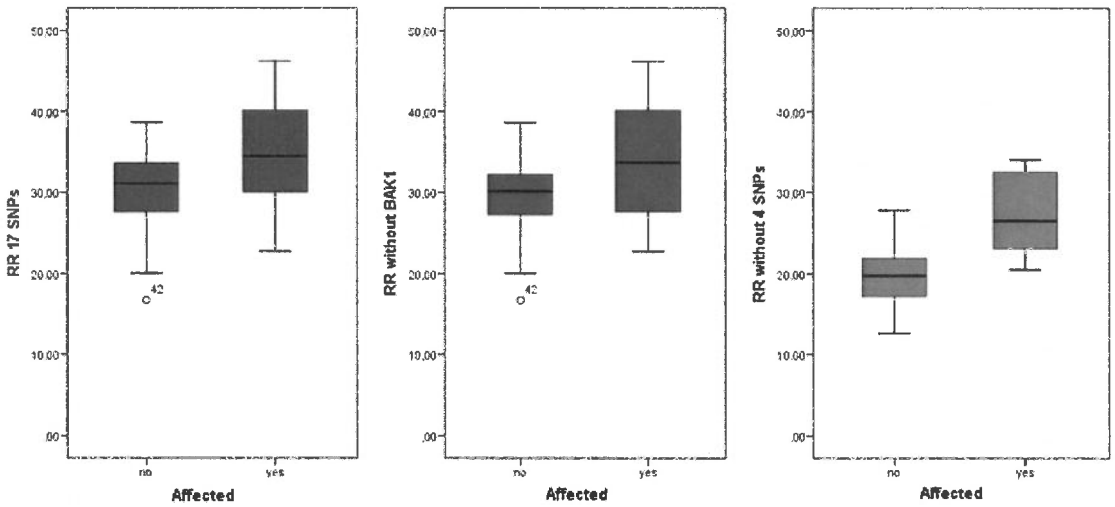


Figure 11: boxplots, RR for affected and non-affected patients. Green boxplots show the sum of 17 SNP resulting in a RR. Outliers are present (dots). The blue boxplots show the sum of 16 SNP were the SNP rs210138 (BAK1) is excluded. Outliers are present (dots). In the red boxplots the 4 SNPs are excluded which divide the two groups (affected and non-affected) poorly.

Summary t-test

In Table 13 a summary of the t-test is given. Per t-test there is shown which SNP is excluded.

Table 13: Overview of the SNPs and statistical analysis. An overview is given per cohort, For each performed t-test is shown which SNP is included (v) and which SNPs are excluded (x) in the RR.

All SNPs			t-test, total cohort			t-test, AIS cohort		
	SNP ID	Nearby gene	17 SNPs	-BAK1	- 4 SNPs	17 SNPs	-BAK1	- 4 SNPs
1	rs2900333	ATF7IP	v	v	v	v	v	v
2	rs4624820	SPRY 4	v	v	v	v	v	v
3	rs210138	BAK 1	v	x	x	v	x	x
4	rs4699052	CENPE	v	v	x	v	v	v
5	rs17021463	HPGDS	v	v	v	v	v	v
6	rs4888262	RFWD3	v	v	v	v	v	x
7	rs7221274	PPM1E	v	v	v	v	v	v
8	rs9905704	TEX 14	v	v	v	v	v	v
9	rs12699477	MAD1L1	v	v	v	v	v	v
10	rs4657482	UCK2	v	v	v	v	v	v
11	rs2736100	TERT	v	v	x	v	v	v
12	rs4635969	TERT/CLPTM1L	v	v	v	v	v	v
13	rs7040024	DMRT1	v	v	x	v	v	v
14	rs755383	DMRT1	v	v	v	v	v	v
15	rs995030	KITLG	v	v	v	v	v	v
16	rs1508595	KITLG	v	v	v	v	v	x
17	rs4590952	KITLG P53	v	v	v	v	v	x

Rs210138

This SNP shows an inverse distribution between the affected and non-affected groups . According to the GWAS study (8, 10) the nucleotide G is the risk-allele to develop a TGCT. The nucleotide A is the non-risk allele. In this study homozygous AA means no risk-alleles, heterozygous AG contains one risk-allele and patients homozygous for GG have two risk-alleles present.

In the total 46XY-DSD population 60% have no risk-alleles present, dividing per group the percentages are: affected 83.3% and in the non-affected group 51.0% are homozygous for the A allele. This means that in contrast to the results found in the GWAS the majority has the non-risk alleles. The distribution is shown in Table 14. Comparing the total DSD population with the EUR population the allele frequencies are similar. DSD population A:76.5% G: 23.5% vs the EUR population A:82% G:18% (33).

Table 14: distribution of amount of risk-alleles between affected and non-affected patient of SNP rs210138, in the total 46XY-cohort

		Affected		
	Genotype	no	yes	Total
rs210138	AA	55	25	80
	AG	37	5	42
	GG	10	0	10

In the AIS population 62.3% have no risk-alleles. Per group the percentages are: affected 85.7% and in the non-affected group 58.7% are homozygous for the G allele. This distribution is shown in Table 15. Comparing the total AIS population with the EUR population the allele frequencies are similar. AIS population A:75.5% G: 24.5% vs the EUR population A:82% G:18% (33).

Table 15 distribution of amount of risk-alleles between affected and non-affected patient of SNP rs210138, in the AIS-cohort

		Affected		
	Risk-allele	no	yes	Total
rs210138	0	27	6	33
	1	13	1	14
	2	6	0	6
Total		46	7	53

Discussion

The aim of this study is to get better insight in GCT risk in patients with DSD. The hypothesis is that analysis of 17 SNPs, found in GWAS studies to be associated with TGCT, in DSD patients, can possibly distinguish between patients who will and who will not develop a GCT. To do so the RR has to be determined using the OR for all 17 SNPs. As the OR of rs4590952 was not known, this is determined first. After this the RR analysis in the DSD- and AIS-populations is discussed.

rs4590952, KITLG p53 RE SNP

The SNP rs4590952 in KITLG which has been found in a p53 response element, is according to Zeron-Medina in linkage disequilibrium (LD) with another SNP found near KITLG, rs995030 (9) . Rs995030 has an OR of 2.38 (risk allele G) and is in high LD with rs1508595 according to Kratz (10), another SNP in KITLG, which has an OR of 2.56 (risk allele G) (8). The obtained OR in this study for rs4590952 is 2.26 (risk allele G). This OR is not completely in line with the results found for the other two KITLG SNPs. All three investigated SNPS are in LD with each other according to the literature. This OR of 2.26 is in line with the ORs of 2.38 and 2.56.

The G allele has an increased frequency in the Caucasian compared (80%) to the African population (24%). Looking at the allele frequency for rs4590952 in the Caucasian male population the frequency of the G allele is 82% (9). The G-allele frequency in the TGCT cohort is 90% an increase of 8% in comparison with the control group. This means that in the affected population the risk-allele is more present. Which influences, according to the literature (9) cancer risk. This can be correlated to the high OR of rs4590952. Because this SNP is in the functional p53-binding site, this can influences the regulation of transcription of the *KITLG* gene. The ability of p53 to regulate transcription is essential for tumor suppression. To see if this SNP within the TGCT cohort separated SE from NS, the allele frequency was calculated in both. This was done using the chi-square test giving p-values of 0.14 (with CT) and 0.79 (without CT), so no significant differences were found. This indicates that the genotypes between SE-patients and NS-patients do not differ for this specific SNP. So no separation can be based on genotyping of rs4590952.

With the OR determined, statistical analysis for all 17 SNPs could be performed on the DSD patients.

Statistical analysis on the total cohort

At first the independent t-test was performed using the RR of 17 SNP on the total cohort of 154 patients with 46XX-DSD and 46XY-DSD. These patients are divided in two groups, with and without precursor lesion/invasive tumor. There was a significant difference between these groups (p-value of 0.039). Although a significant difference is found, the 22 46XX-DSD patients are excluded in further analysis because 46XX-DSD patients seldom develop a GCT.

The t-test was performed using the RR in the total cohort of 132 46-XY DSD patients to see if patients with (n=102) and without precursor lesion/invasive tumor (n=30) could be distinguished. Using all 17 SNPs, this resulted in a significant difference between the two groups (p-value of 0.014). However, there is still an overlap between the two groups as shown in the boxplots (Figure 10). Because of the overlap these SNPs have no clinical relevance in predicting tumor risk.

The SNP nearby BAK1 (rs210138) showed an inverse correlation (Table 14) compared to the GWAS studies, as Chung *et al.* and Kratz *et al.* find the presence of the G-allele as a risk factor for TGCT. In the DSD-population analyzed here the A-allele is more present, In patients with a TGCT the frequency of the A allele is 91.7%. This difference might be explained because of the difference between DSD-patients with a GCT and men with a TGCT. In the DSD-population, during embryogenesis no complete testicularization is reached. Whereas patients with a TGCT, which are used in the GWAS-studies, no underlying problems in sex differentiation is present. This might be reflected in the genotypes of rs210138.

Excluding the SNP nearby *BAK1* (rs210138) from the RR, the significance increases (p-value of 0.002), but again here there still is overlap between the groups. Excluding three more SNP; rs4699052, rs2736100 and rs7040024 (near *CENPE*, *TERT* and *DMRT1* respectively), which showed a negative effect separating the two groups, the p-value drops down to 0.001. Although again there is an improved significant difference, the groups still overlap each other. This means that a SNP based screening as performed here cannot be used in clinical decision making. As these four SNP gave different results than those found in the GWAS studies, we wanted to see if the supposed associated genes fall in one or more specific pathways. This could give insight in the development of GCT in DSD compared to TGCT development in men without DSD. Looking at the 4 SNPs which are excluded in the RR, the genes CENPE, TERT, DMRT1 and BAK1 are on different chromosomes, protein function differs and none of the proteins are in the same pathway.

Statistical analysis of the AIS patient cohort

The t-test was also performed in the AIS cohort of 53 patients, to see if patients with (n=7) and without precursor lesion (n=46) could be distinguished. Using all 17 SNPs for RR analysis results in a p-value of 0.023. Excluding the SNP nearby BAK1, significance rises, with a p-value of 0.015. But the boxplots overlap. Excluding three more SNP; rs4888262, rs1508959 and rs4590952 (near *RFWD3*, *KITLG* and *KITLG* respectively), which show a negative effect on distinguishing these groups, the p-value drops down to 0.000. With this significance, looking at the boxplots the boxes itself don't overlap any more. But there is overlap with the whiskers. The RR analysis of the 13 SNPs could be an additional risk-parameter for developing a GCT.

Within this cohort all affected patients have a precursor lesion and none of the affected patients have an invasive GCT. Until recently the gonads in complete-AIS (CAIS) patients were removed in adolescence or directly at birth. These patients have female external genitals. Nowadays a more conservative approach is taken in which in many CAIS patients the gonads are retained until early adulthood. This to take advantage of endogenous hormone secretion in which excess androgens through aromatization are transformed in estrogens resulting in breast development, growth spurt and a build-up of bone mass. Around fifteen percent of these patients refuse subsequent orchidectomy (34). This leads to a lack of data regarding long-term TGCT risk in AIS patients. More factors might contribute to the low GCT risk in AIS patients. The testes have developed normally, and germ cells undergo apoptosis in early life. Partial-AIS (PAIS) patients are raised male. Phenotypically they have developed more male external genitals but are not completely male. An orchidectomy is avoided in these patients because during puberty T is needed to develop secondary sex characteristics.

Looking if the SNPs excluded from the RR calculations could give insight in the pathways which lead to pre-cursor/GCT development. As less invasive tumors are found in the AIS population there might be another pathway involved in the development of the precursor lesion, compared with men who, invariably, develop an invasive TGCT from the precursor GCNIS when left untreated. BAK1 and KITLG are in the same pathway (35). BAK1 induces apoptosis (36), within this cohort the affected group shows that they overall do not carry the risk-allele. It could be that in AIS patients the precursor lesions do not progress into an invasive tumor, because of induced apoptosis by BAK1 as expected in the normal situation. Having the risk-allele in BAK1 might prevent apoptosis which could result in survival of malignant germ cells and ultimately development of an invasive tumor. KITLG is in this pathway upstream of BAK1 and also regulates cell survival (37). One of the two SNPs (rs4590952) is in a p53 binding site. The ability of TP53 to regulate transcription is crucial for tumor suppression as TP53 conserves stability by preventing genome mutations (14). This might indicate that polymorphisms in this binding site could influence cancer development (9). As RFWD3 has an interaction with P53 and functions as an DNA damage checkpoint, this might indicate that p53 functions normally as a tumor suppressor gene. Taken together, analysis of this pathway leads to the hypothesis that although a precursor lesion can develop, having the non-risk-alleles for these four SNPs prohibits the development of an invasive tumor.

Conclusion

In both groups analyzed, total and AIS, for the SNPs shown to be associated with TGCT development, there is a statistical significant difference between the affected and non-affected groups. In the total group of 46-XY DSD patients, no clinical relevance is found, as the groups overlap.

Within the group of AIS patients, when four SNPs are excluded in the RR analysis, this results in a significant difference (p=0.000) between the non-affected and affected groups. The boxplots show a clear separation between patient groups. This RR analysis could be an additional risk-factor for GCT development in AIS patients. Because none of the affected patients have an invasive tumor longer follow-up is needed to gain more information about GCT development in AIS patients. In this study it is not possible to distinguish affected PAIS and CAIS patients, because only one of the seven patients is diagnosed with PAIS.

Because of the small number of patients analyzed, 154 patients in the total cohort and 53 patients in the AIS cohort, the chi-square test per SNP cannot be performed for every SNP, this because the expected values are below 5 in more than 20% of the expected values. Using larger patient groups a more robust statistical analysis can be performed and more reliable conclusions can be drawn. Looking at the AIS group it might be interesting to look further into the function of the genes associated with the four SNPs which do not show the risk-alleles within the affected groups, as this does not correlate with the results found in the GWAS-studies. This might give a clue why there is a low TGCT risk within the AIS patients. To investigate this further in mouse models/cell lines first the non-risk-alleles must be present in a bigger group of AIS patients. As making a solid conclusion based on the presence of non-risk-alleles in seven patients is not possible.

References

1. Wilhelm D, Koopman P. The makings of maleness: towards an integrated view of male sexual development. 2006;7:620-31.
2. Hersmus R. Disorders of Sex Development and Germ Cell Cancer: genetics and microenvironment. Erasmus University Rotterdam, 2012.
3. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90:7-24.
4. Yan W, Suominen J, Toppari J. Stem cell factor protects germ cells from apoptosis in vitro. J Cell Sci 2000;113 (Pt 1):161-8.
5. Oosterhuis JW, Looijenga LH. Testicular germ-cell tumours in a broader perspective. Nat Rev Cancer 2005;5:210-22.
6. cijfers over kanker. www.cijfersoverkanker.nl. Accessed: September 2015.
7. Fung C, Vaughn DJ. Complications associated with chemotherapy in testicular cancer management. Nat Rev Urol 2011;8:213-22.
8. Chung CC, Kanetsky PA, Wang Z, *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. Nat Genet 2013;45:680-5.
9. Zeron-Medina J, Wang X, Repapi E, *et al.* A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. Cell 2013;155:410-22.
10. Kratz CP, Greene MH, Bratslavsky G, Shi J. A stratified genetic risk assessment for testicular cancer. Int J Androl 2011;34:e98-102.
11. Kanetsky PA, Mitra N, Vardhanabhuti S, *et al.* Common variation in KITLG and at 5q31.3 predisposes to testicular germ cell cancer. Nat Genet 2009;41:811-5.
12. Kanetsky PA, Mitra N, Vardhanabhuti S, *et al.* A second independent locus within DMRT1 is associated with testicular germ cell tumor susceptibility. Hum Mol Genet 2011;20:3109-17.
13. Szumilas M. Explaining Odds Ratios. Journal of the Canadian Academy of Child and Adolescent Psychiatry 2010;19:227-9.
14. Read TSaAP. Human Molecular Genetics. John Wiley & Sons: 1999.
15. Hughes IA, Houk C, Ahmed SF, *et al.* Consensus statement on management of intersex disorders. Arch Dis Child 2006;91:554-63.
16. GBY Region. http://www.omim.org/entry/424500. Accessed: december 2015.
17. Pleskacova J, Hersmus R, Oosterhuis JW, *et al.* Tumor risk in disorders of sex development. Sex Dev 2010;4:259-69.
18. Ensemble. http://www.ensembl.org/Homo_sapiens/Variation/Sample?db=core;r=12:88559382-88560382;v=rs4590952;vdb=variation;vf=2777662#373520_tablePanel. 1000 genomes data rs 4590952. Accessed: 4 april 2016.
19. Zumbo P. Phenol-chloroform extraction. In: Weill cornell medical college Laboratory of christopher E. Masson PH.D, department op physiology & biophysics.
20. bitesizebio. http://bitesizebio.com/384/the-basics-how-phenol-extraction-works/. Accessed: 17 march 2016.
21. biolegio mp. http://www.biolegio.com/products-services/probes-for-real-time-pcr/xs-probes/?gclid=Cj0KEQjwmKG5BRDv4YaE5t6oqf0BEiQAwqDNfDBequO7QlfmBeVXBRY3mB3Yffs5-KcEW4oD4H2UQHcaAiKw8P8HAQ. Accessed: 03-05 2016.
22. Thermofischer. https://www.thermofisher.com/nl/en/home/life-science/pcr/real-time-pcr/real-time-pcr-assays/snp-genotyping-taqman-assays.html Accessed: 17 march 2016.

23. Thermofischer TMp. https://www.thermofisher.com/nl/en/home/products-and-services/product-types/primers-oligos-nucleotides/applied-biosystems-custom-primers-probes/custom-taqman-probes.html. Accessed: 03-05 2016.

24. all C-CYe. 5' Nuclease Assays for the loci CCR5- +/ 32, CCR2-V641, and SDFI-G80I A Related to Pathogenesis of AIDS. Clinical Chemistry 2000;46:24-30.

25. Shi SR, Cote RJ, Wu L, *et al.* DNA extraction from archival formalin-fixed, paraffin-embedded tissue sections based on the antigen retrieval principle: heating under the influence of pH. J Histochem Cytochem 2002;50:1005-11.

26. qiagen. https://www.qiagen.com/us/shop/sample-technologies/dna/dna-preparation/DNeasy-Blood-and-Tissue-Kit/?cmpid=QVen9GAdneasy_brand&kwid=qiagen%20dneasy%20tissue%20kit#productdetails. Accessed: 30 march 2016.

27. Sabin P. Medical statistics at a glance. 3th ed. Wiley-blackwell: 2009.

28. I. Olkin ea. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, In: Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press: Stanford University Press, Palo Alto; 1960. pp pp. 278-92.

29. WG C. The [chi-squared] test of goodness of fit. Annals of Mathematical Statistics 1952:315–45.

30. 10:417–451. CWSmfstcc-stB. Some methods for strengthening the common [chi-squared] tests. Biometrics 1954:417–51.

31. Yates D, Moore, Moore, D., McCabe, G. . The Practice of Statistics. 1 ed. 1999.

32. Woodward PJea. Pathology and Genetics of the Urinary System and Male Genital Organs. Vol 2004. IARC Press: Lyon. 217–78 pp.

33. allele frequency rs210138. http://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=6:33574261-33575261;v=rs210138;vdb=variation;vf=101552. Accessed: 10 may 2016.

34. Hughes IA, Davies JD, Bunch TI, *et al.* Androgen insensitivity syndrome. Lancet 2012;380:1419-28.

35. Nathanson EARaKL. Predispostition alleles for testicular germ cell tumour. Elsevier 2010;20:225-30.

36. Science WIo. BAK1 genecard. http://www.genecards.org/cgi-bin/carddisp.pl?gene=BAK1. 1996-2016. Accessed: 10 may 2016.

37. Science WIo. KITLG genecard. http://www.genecards.org/cgi-bin/carddisp.pl?gene=KITLG. 1996-2016 Accessed: 10 may 2016.

Appendices

Appendix 1: overview subgroups DSD

Sex chromosome DSD	46,XY-DSD	46,XX-DSD
A: 47,XXY (Klinefelter syndrome and variants)	A: Disorders of gonadal (testicular) development	A: Disorders of gonadal (ovarian) development
B: 45,X (Turner syndrome and variants)	1. Complete or partial gonadal dysgenesis	1. Gonadal dysgenesis
C: 45,X/46,XY (mixed gonadal dysgenesis)	2. Ovotesticular DSD	2. Ovotesticular DSD
D: 46,XX/46,XY (chimerism)	3. Gonadal regression	3. Testicular DSD
	B: Disorders in androgen synthesis or action	B: Androgen excess
	1. Disorders of androgen synthesis	1. Fetal
	a. LH receptor mutations	a. 3 β-hydroxysteroid dehydrogenase 2
	b. Smith-Lemli-Opitz syndrome	b. 21-hydroxylase
	c. Steroidogenic acute regulatory protein mutations	c. P450 oxidoreductase
	d. Cholesterol side-chain cleavage	d. 11 β-hydroxylase
	e. 3 β-hydroxysteroid dehydrogenase 2	e. Glucocorticoid receptor mutations
	f. 17 β-hydroxysteroid dehydrogenase	
	g. 5 α-reductase 2	2. Fetoplacental
		a. Aromatase deficiency
	2. Disorders of androgen action	b. Oxidoreductase deficiency
	a. Androgen insensitivity syndrome	
	b. Drugs and environmental modulators	3. Maternal
		a. Maternal virilizing tumours (e.g. luteomas)
		b. androgenic drugs
	C: Other	C: Other
	1. Syndromic associations of male genital development (e.g. cloacal anomalies, Robinow, Aarskog, Hand-Foot-Genital, popliteal pterygium)	1. Syndromic associations (e.g. cloacal anomalies)
	2. Persistent Müllerian duct syndrome	2. Müllerian agenesis/hypoplasia (e.g. MURCS)
	3. Vanishing testis syndrome	3. Uterine abnormalities (e.g. MODY5)
	4. Isolated hypospadias	4. Vaginal atresis (e.g. KcKusick-Kaufman)
	5. Congenital hypogonadotropic hypogonadism	5. Labial adhesions
	6. Cryptorchidism	
	7. Environmental influences	

Hughes IA, Houk C, Ahmed SF, Lee PA, Group LC, Group EC. Consensus statement on management of intersex disorders. Arch Dis Child 2006;91:554-63.

Appendix 2: DNA isolation protocols

Protocol 1: DNA isolation from formalin fixed paraffin embedded tissue (FFPE)

Solutions

- Xylene
- Etanol absolute ($EtOH_{abs}$)
- Lysis-buffer (10 mM Tris.Cl, 100 mM NaCl, 5mM EDTA, 1% SDS, 1 mM $CaCl_2$)
- Proteinase k (prot k) (10 mg/ml)
- Phenol/Chloroform
- Iso-propanol
- 2M NaAc
- Glycogen 20 mg/ml
- $T^{10}E^{1}$ (10 mM Tris.Cl, 1 mM EDTA)

Materials

- Tubes, eppendorf 1,5 and 2,0 ml
- Top-over
- Thermal-incubator 55°C

Procedure

- First cut 1 slice of 3 µm thickness for a HE-staining
- Cut 2-3 slices (depending on the size of the tissue) of 10 µm thickness and transfer it to a clean, sterile 2 ml eppendorf tube
- Add 1 ml xylene
- Shake for at least 30 min. in top-over
- Centrifuge for 5 min. at 14.000 rpm at Rt
- Discard supernatant
- Repeat xylene-step for another two times
- Add 1 ml $EtOH_{abs}$ to the pellet
- Mix well (approximately 15 min. in top-over)
- Centrifuge for 4 min. at 14.000 rpm at Rt
- Discard supernatant
- Repeat $EtOH_{abs}$ washstep for another two times
- Airdry pellet
- Add 800 µl lysisbuffer and mix well by vortexing
- Add 75 µl prot k
- Incubate at 52°C at 1200 rpm until tissue is completely lysed; lysis time varies depending on the type of tissue processed. Lysis should be done overnight in a shaking thermal-incubator.
- Add 900 µl phenol/chloroform
- Mix for at least 30 min. in top-over
- Centrifuge 5 min. at 14.000 rpm at Rt
- Pipet upper aqueous phase to a new clean 2,0 ml tube
- DNA is precipitated by adding 80 µl 2M NaAc and 600 µl iso-propanol in case of a biopt add also 1 µl glycogen
- Mix well by turning the tube several times
- Incubate minimal 1 hour at -20°C
- Centrifuge 10 min at 14.000 rpm at 4°C
- Discard supernatant
- Add 500 µl 70% ethanol and mix well
- Centrifuge 3 min at 14.000 rpm at Rt
- Discard supernatant

- Airdry pellet
- Dissolve pellet in about 30 µl $T^{10}E^{1}$ (amount depending on pelletsize)
- Load 1 µl DNA on a 0.7% TBE-agarose gel for DNA-quality control, and measure concentration on the Nanodrop 2000
- Store DNA at 4°C or at -20°C if stored for a long time

Protocol 2: DNA isolation from formalin fixed paraffin embedded tissue (FFPE)

Solutions:

- Xylene
- $EtOH_{abs}$
- Universal buffer (each chemical 28,6M: citrate acid, KH_2PO_4, H_3BO_3 and diethylbarbituric acid, pH 10)
- Prot k (10 mg/ml)
- Phenol/Chloroform
- Iso-propanol
- 2M NaAc
- Glycogen 20 mg/ml (Roche; 901393)
- $T^{10}E^1$ (10 mM Tris.Cl, 1 mM EDTA)

Materials:

- Tubes, eppendorf 1,5 and 2,0 ml
- Top-over
- Thermal-incubator 55/100°C

Protocol:

- First cut 1 slice of 3 µm thickness for a HE-staining
- Cut 2-3 slices (depending on the size of the tissue) of 10 µm thickness and transfer it to a clean, sterile 2 ml eppendorf tube
- Add 1 ml xylene
- Shake for at least 30 min. in top-over
- Centrifuge for 5 min. at 14.000 rpm at Rt
- Discard supernatant
- Repeat xylene-step for another two times, vortex after adding xylene
- Add 1 ml $EtOH_{abs}$ to the pellet
- Mix well (approximately 15 min. in top-over)
- Centrifuge for 4 min. at 14.000 rpm at Rt
- Discard supernatant
- Repeat EtOHabs washstep for another two times, vortex after adding $EtOH_{abs}$
- (Turn heatblock on 100°C and 52°C)
- Airdry pellet
- Add 500 µl universal buffer solution, mix well
- Heat at 100°C using an heatblock for 30 min at 1200 rpm, put special clips on the Eppendorf tubes to prevent opening as a result of the pressure on the tubes
- Allow the tube to cool for 5 min
- Add 47 µl proteinase k
- Incubate at 52°C at 1200 rpm until tissue is completely lysed; lysis time varies depending on the type of tissue processed. Lysis should be done overnight in a shaking thermal-incubator.
- Ad 550 µl phenol/chloroform (1:1)
- Mix for at least 30 min. in top-over
- Centrifuge 5 min. at 14.000 rpm at Rt
- Pipet upper aqueous phase to a new clean 2,0 ml tube
- Check volume
- DNA is precipitated by adding 55 µl 2M NaAc (0,1 volume) and 385 µl iso-propanol (0,7 volume) in case of a biopt add also 1 µl glycogen
- Mix well by turning the tube several times

- Incubate minimal 1 hour at -20°C (overnight is also possible)
- Centrifuge 10 min at 14.000 rpm at 4°C
- Discard supernatant
- Add 500 µl 70% ethanol and mix well
- Centrifuge 3 min at 14.000 rpm at Rt
- Discard supernatant
- Airdry pellet
- Dissolve pellet in about 30 µl T10E1 (amount depending on pelletsize)
- Load 1 µl DNA on a 0.7% TBE-agarose gel for DNA-quality control, and measure concentration on he Nanodrop 2000
- Store DNA at 4°C or at -20°C if stored for a long time

Protocol 3: DNA isolation from pBl using saturated NaCl

Solutions:

- Blood lysis buffer, NH_4Cl 165,8 g; $KHCO_3$ 22,4 g; EDTA 7,4 g. Dissolved in 800 ml H_2O, pH adjusted to 7,4
- PBS
- Nucleus Lysis buffer, Tris 1,21g; NaCl 23,4g; EDTA 0,74g. Dissolve in 800 ml H_2O, adjust pH to 8,2 with concentrated HCl and fill to 1L with H_2O
- Prot k, 10 mg/ml
- 20% SDS
- NaCl, 6M
- $EtOH_{abs}$
- $T_{10}E_1$ (10mM Tris.Cl and 1mM EDTA)

Material:

- 50 ml tubes
- 2ml Eppendorf tubes

Protocol

- 7 ml anticoagulated blood include in 25 ml Blood lysis buffer (BLB) in 50ml tubes, mix well
- 15' on ice, or till lysed o.n. 4 °C
- Centrifuge 12' 2000 rpm at 14 °C without brake
- Discard carefully the supernatant.
- Wash pellet carefully with PBS and clean the inside of the tube with a tissue.
- Suspend pellet in 5 ml nucleus lysis buffer, mix well
- Add 75 ml prot. k, mix well
- Add 250 µl 20% SDS, mix well
- Incubate o.n. at 37 °C
- Add 2 ml saturated NaCl
- Shake 15-20
- Centrifuge 15' 4500 rpm at RT without break
- Pipet upper layer (without foam) into a new 50 ml tube
- Add 12,5 ml $EtOH_{abs}$. mix until a precipitate arises
- Bring the DNA-precipitate into a clean 2 ml Eppendorf tube.
- Dissolve DNA in $T_{10}E_1$
- Store DNA at 4°C or at -20°C if stored for a long time