

Aus der Klinik für Kinder- und Jugendmedizin
der Universität zu Lübeck
Direktor: Prof. Dr. med. Egbert Herting

Identifikation eines neuen Pathomechanismus zur Entstehung einer 46,XY-Gonadendysgenese

Inauguraldissertation
zur Erlangung der Doktorwürde
der Universität zu Lübeck
- Aus der Sektion Medizin -

vorgelegt von
Jakob Arnold Meinel
aus Flensburg

Lübeck, 2021

1. Berichterstatter:

Prof. Dr. med. Olaf Hiort

2. Berichterstatter:

Prof. Dr. med. Malte Spielmann

Tag der mündlichen Prüfung:

24.02.2023

Zum Druck genehmigt. Lübeck, den

24.02.2023

Promotionskommission der Sektion Medizin

Inhaltsverzeichnis

0. Abkürzungsverzeichnis	V
1. Einleitung und Fragestellung	1
1.1 Grundlagen der menschlichen Geschlechtsentwicklung	1
1.2 Varianten der Geschlechtsentwicklung – Differences of sex development (DSD)	4
1.3 Molekulare Mechanismen der Geschlechtsdeterminierung mit besonderem Augenmerk auf die NROB1 abhängige SOX9 Regulierung	7
1.4 Einteilung molekulargenetischer Mutationen und deren Krankheitsrelevanz	10
1.4.1 Sequenzvariationen	11
1.4.2 Strukturelle Variationen	12
1.5 Fragestellung und Zielsetzung	13
2. Material und Methode	15
2.1 Patient:innen	15
2.2 Whole Genome Sequencing (WGS)	17
2.2.1 Prinzip	17
2.2.2 SNV-Analyse	21
2.2.3 CNV-Analyse	21
2.2.4 Integrative Genomics Viewer (IGV)	22
2.2.5 Bruchpunktbestimmung aus WGS Reads	22
2.3 Array-Based Comparative Genomic Hybridisation (aCGH)	23
2.4 Bruchpunktbestätigung	24
2.4.1 Primerdesign	24
2.4.2 Polymerase-Kettenreaktion (PCR)	24
2.4.3 Gelelektrophorese	25
2.4.4 DNA-Sequenzierung	25
2.5 Transkriptom-Sequenzierung	27
2.5.1 Prinzip	27
2.5.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)	27
2.5.3 Gene Set Variation Analysis (GSVA)	28

2.6	qPCR-Screening auf Xp21.2 Kopiezahlvariationen	29
2.6.1	Konzipierung der qPCR-Assays	29
2.6.2	Etablierung der qPCR-Assays	30
2.6.3	Validierung der qPCR-Assays	31
2.6.4	Auswertung	31
2.6.5	Bestimmung der DNA-Konzentration	31
2.6.6	Aktenrecherche	32
2.7	Konformationserfassung von Chromosomen durch Hi-C	33
3	Ergebnisse	34
3.1	Whole Genome Sequencing (WGS)	34
3.1.1	Patient:in 1	34
3.1.2	WGS-Analyse weiterer Patient:innen	36
3.2	Bestätigungen der Kopiezahlvariationen	38
3.3	Bruchpunktbestimmung	39
3.4	Transkriptom-Sequenzierung	43
3.5	Screening weiterer Patient:innen auf Xp21.2 Kopiezahlvariationen	46
3.6	Bestätigung und Charakterisierung der Triplikation von P3	48
3.7	Identifikation von Isolator-Bindungsstellen	52
3.8	Hi-C Analyse	54
3.9	Identifikation putativer Enhancer-Elemente	55
3.10	Potentieller Pathomechanismus der Kopiezahlvariationen am Xp21.2 Locus durch Disruption Topologisch Assozierender Domänen (TADs)	57
4	Diskussion	60
4.1	Whole Genome Sequencing	60
4.1.1	Möglichkeiten und Grenzen bei der Erkennung von strukturellen Variationen	60
4.1.2	Anwendung zum Ausschluss anderer Ursachen der 46,XY-GD	62
4.1.3	Anwendung zur Diagnosestellung bei DSD-Patient:innen	63
4.2	Beschränkungen der Erkennungsmöglichkeiten des qPCR-Screenings	64

4.3	RNA-Sequenzierung	65
4.3.1	Einfluss der Lagerungsdauer auf die FFPE-RNA-Sequenzierung	65
4.3.2	Expression gonadenspezifischer Gene	66
4.3.3	Expression des MYC Signalweges	67
4.3.4	Bedarf weiterer RNA-Sequenzierungs-Vergleichsdaten	68
4.4	ChIP-Seq	68
4.4.1	Bewertung als mögliche TAD-Grenze des identifizierten Isolators	68
4.4.2	Übertragbarkeit auf gonadales Gewebe	70
4.5	Hi-C Analyse	70
4.5.1	Einteilung in Konstitutive und Fakultative TADs	70
4.5.2	Schwankungsanfälligkeit von Hi-C-Daten und TAD-Bereichen	71
4.5.3	Übertragbarkeit auf Patient:innenproben	72
4.6	GeneHancer	72
4.6.1	Verlässlichkeit putativer Enhancer-Elemente	72
4.6.2	Enhancerwirkungen sind abhängig von vorliegenden Transkriptionsfaktoren	73
4.7	Schlussfolgerung	74
4.8	Ausblick	77
5	Zusammenfassung	79
6	Literaturverzeichnis	80
7	Anhang	92
8	Danksagungen	107
9	Lebenslauf	108

0. Abkürzungsverzeichnis

17 β HSD3	17 β -Hydroxysteroid-Dehydrogenase
46,XY-GD	46,XY-Gonadendysgenese
aCGH	Array-Based Comparative Genomic Hybridisation
AMH	Anti-Müller-Hormon
AR	Androgenrezeptor
BAM	Binary Alignment Map
BLAT	BLAST like alignment tool
Bp	Basenpaar
BWA	Burrows-Wheeler Aligner
cDNA	Komplementäre DNA
CMA	Chromosomal Microarray
Cq	Cycle quantification Value
CTCF	CCCTC-binding factor
CXorf21	Chromosome X Open Reading Frame 21
ddNTP	Didesoxyribonukleosid-Triphosphate
DMD	Dystrophin
DMRT1	Doublesex and mab-3 related transcription factor 1
DNA, DNS	Desoxyribonukleinsäure
dNTP	Desoxyribonukleosid-Triphosphate
DSD	Differences/Disorders of Sex Development
ESR2	Estrogen Receptor 2
FGFR2	Fibroblast Growth Factor Receptor 2
FOG2	Friend of GATA 2
FOXL2	Forkhead-Box-Protein L2
FRET	Förster-Resonanzenergietransfer
FSH	Follikelstimulierendes Hormon
FTHL17	Ferritin Heavy Chain Like 17
GATA4	GATA Binding Protein 4
GB	Gigabyte
GK	Glycerinkinase
GPR15	G Protein-Coupled Receptor 15

GSVA	Gene Set Variation Analysis
IGV	Integrative Genomics Viewer
IL1RAPL1	Interleukin 1 Receptor Accessory Protein Like 1
InDel	Insertion/Deletion
Kb	kilobase
LH	Luteinisierendes Hormon
MAGEB1-4	MAGE Family Member B1-4
MLPA	Multiplex Ligation-dependent Probe Amplification, Multiplexe ligationsabhängige Sondenamplifikation
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NROB1	Nuclear Receptor Subfamily 0 Group B Member 1
P1-3	Patient:in 1-3
PCR	Polymerase Chain Reaktion
POR	Cytochrome P450 Oxidoreductase
qPCR	quantitative PCR
RNA	Ribonukleinsäure
rSAP	Rekombinante Shrimp Alkalische Phosphatase
SAM	Sequence Alignment Map
SBS	Sequencing by Synthesis
SF1	Steroidogenic factor 1
SMC3	Structural maintenance of chromosomes protein 3
SNP	Single Nucleotide Polymorphism, Einzelnukleotid-Polymorphismus
SNV	Single nucleotide Variation
SOX9	SRY-box 9
ssDNA	single stranded DNA, Einzelstrang DNA
STAR	Steroidogenic Acute Regulatory Protein
SV	Strukturelle Variation
t-SNE	t-Distributed Stochastic Neighbor Embedding
TAB3	TGF-Beta Activated Kinase 1 (MAP3K7) Binding Protein 3
TAD	Topologically associating Domain, Topologisch assoziierte Domäne
Taq-Polymerase	DNA-Polymerase aus dem Bakterium Thermus-aquaticus

TBE	Tris-Borat-EDTA-Puffer
TPM	Transkript pro Millionen Transkripte
WGS	Whole Genome Sequencing
WNT4	Wnt Family Member 4
WT1	Wilms-Tumorsuppressorgen
ZFPM2	Zinc Finger Protein, FOG Family Member 2
ZNF80	Zinc Finger Protein 80

1. Einleitung und Fragestellung

1.1 Grundlagen der menschlichen Geschlechtsentwicklung

Die Entwicklung des menschlichen Geschlechts vollzieht sich auf verschiedenen Ebenen. Hierbei ist speziell die biologische, psychische und soziale Ebene zu nennen, welche sich in ihren Ausrichtungen nicht zwingend gegenseitig entsprechen müssen. In der vorliegenden Arbeit liegt der Fokus ausschließlich auf der Entwicklung des biologischen Geschlechts.

Die biologische Geschlechtsentwicklung lässt sich dabei unterteilen in Geschlechtsdeterminierung und Geschlechtsdifferenzierung.[1, 2] Grundlage dieser Prozesse ist das chromosomale Geschlecht, welches bei der Befruchtung der Eizelle durch das Spermium festgelegt wird. Das weibliche Geschlecht ist mit einem 46,XX-Chromosomensatz annotiert und das männliche Geschlecht durch 46,XY. Bis zur 6. Schwangerschaftswoche bleibt der Chromosomensatz der einzige Unterschied zwischen männlichen und weiblichen Embryos, da die übrigen Anlagen ansonsten noch nicht zu unterscheiden sind.[1, 2] Die Anlagen können sich zu dem Zeitpunkt noch omnipotent in männliche oder weibliche Richtung entwickeln.[3, 4]

Die Entwicklung der Gonaden und damit die Geschlechtsentwicklung beginnt mit der Bildung der Urogenitalfalte aus dem Mesonephros.[2, 4] Initiiert durch GATA4 (GATA Binding Protein 4) kommt es ab der 5. Schwangerschaftswoche im mittleren Drittel der Urogenitalfalte zur Proliferation von Coelomepithelzellen und in der Folge zu einer Verdickung, d.h. zur Bildung eines mehrschichtigen Epithels.[3] Aus den Coelomepithelzellen entwickeln sich dann die somatischen Zellen der späteren Gonaden. Im Gegensatz dazu entstammen die Keimzellen der Gonaden den ursprünglich extraembryonal im Dottersack gebildeten Urkeimzellen. Die Urkeimzellen beginnen in der 4. Woche ihre Wanderung aus dem Dottersack in die Urogenitalfalte, in der sie in der 5. Woche ankommen. Dabei entsteht eine zunächst indifferente Gonadenanlage. (Überblick bei [3])

Erst nach der 6. Woche beginnt die Geschlechtsdeterminierung. Dieser Prozess steht maßgeblich unter dem Einfluss verschiedener Transkriptionsfaktoren, und führt dazu, dass sich aus der indifferenten, paarigen Gonadenanlage die Testes beziehungsweise die Ovarien entwickeln.[1] In den Gonadenanlagen finden sich drei maßgebliche Zellarten (Stützzellen und

Zwischenzellen als somatische Zellen, sowie die Urkeimzellen), die sich weiter differenzieren und in der Folge später das testikuläre oder ovarielle Gewebe bestimmen.[5] Entscheidend für die Geschlechtsdeterminierung und die Differenzierung dieser drei Zelllinien ist die Expression des *SRY* (*Sex determining region of Y*)-Gens auf dem Y-Chromosom, welches den Testis-determinierenden Faktor (TDF) codiert.[6] TDF unterdrückt die Entwicklung des Ovars und aktiviert die spezifisch männliche Entwicklungssignalkaskade in der Gonade.[5] Dadurch entstehen aus den Stützzellen Sertoli-Zellen, aus den Zwischenzellen Leydig-Zellen und aus den Urkeimzellen Spermatogonien. Bei Abwesenheit von *SRY* setzt die Entwicklung zur Ausbildung weiblicher Keimdrüsen ein.[5, 7] Unter Einwirkung weiterer Genregionen entwickeln sich physiologisch nachfolgend die Stützzellen zu Granulosazellen, die Zwischenzellen zu Thekazellen und die Urkeimzellen zu Oogonien.[2, 5]

Abhängig von der Gonadenentwicklung kommt es nachfolgend zur Geschlechtsdifferenzierung, welche zur Ausbildung des phänotypischen Geschlechts führt. Das phänotypische Geschlecht ist sowohl durch die inneren, als auch die äußeren Geschlechtsorgane bestimmt, die aus verschiedenen Anlagen hervorgehen. Die inneren Genitalorgane leiten sich von den paarigen Müller- und Wolff-Gängen, sowie dem Sinus urogenitalis ab.[5, 8] (Abb. 1) Das äußere Genital entsteht aus dem Genitalhöcker, dem Genitalwulst und den Genitalfalten.[5, 8] Im Gegensatz zur von Transkriptionsfaktoren gesteuerten Geschlechtsdeterminierung steht die Geschlechtsdifferenzierung maßgeblich unter dem Einfluss von Hormonen und deren Wirkungsvermittlung durch Rezeptoren.[1]

Im XY-Embryo kommt es durch den Hoden zur Produktion von drei für die männliche Geschlechtsdifferenzierung wesentliche Hormone: Antimüllerhormon (AMH), Testosteron und 5α -Dihydrotestosteron. Das in den Sertoli-Zellen produzierte AMH führt zur Obliteration der Müllergänge, wohingegen das in den Leydig-Zellen produzierte Testosteron die Differenzierung und Stabilisierung des Wolff-Gangs stimuliert. Aus den Wolff-Gängen entstehen in der Folge Nebenhoden, Bläschendrüsen und Ductus deferens. Das 5α -Dihydrotestosteron wird in der Peripherie über die 5α -Reduktase aus dem in den Leydig-Zellen produzierten Testosteron gebildet und führt unter anderem zur Entwicklung der Prostata aus dem Sinus urogenitalis. Das äußere männliche Genital entsteht ebenfalls unter dem Einfluss von 5α -Dihydrotestosteron. Dabei geht aus dem Genitalhöcker der Penis, aus der Fusion der Genitalwülste das Scrotum und aus den Genitalfalten der Penisschaft hervor. (Überblick bei [5])

Im Gegensatz zu den Hoden sind die Ovarien pränatal hormonell inaktiv. Dementsprechend verläuft die weibliche Geschlechtsdifferenzierung unabhängig von Hormonen aus dem Ovar. Das heißt, sie erfolgt beispielsweise auch dann, wenn eine Agenesie oder Dysgenese der Gonaden vorliegt. In Ermangelung an Testosteron bilden sich im XX-Embryo die Wolff-Gänge zurück, während sich die Müller-Gänge, ohne Suppression durch AMH, in die Tuben, den Uterus und den oberen Teil der Vagina weiterentwickeln können. Der untere Teil der Vagina entwickelt sich aus dem Sinus urogenitalis, aus dem sich unter anderem auch die weiblichen Geschlechtsdrüsen bilden. Aus den drei indifferenten Anlagen des äußeren Genitals entwickeln sich bei der Frau die Klitoris (Genitalhöcker), die kleinen Schamlippen (Genitalfalten) und die großen Schamlippen (Genitalwülste). (Überblick bei [5, 8]).

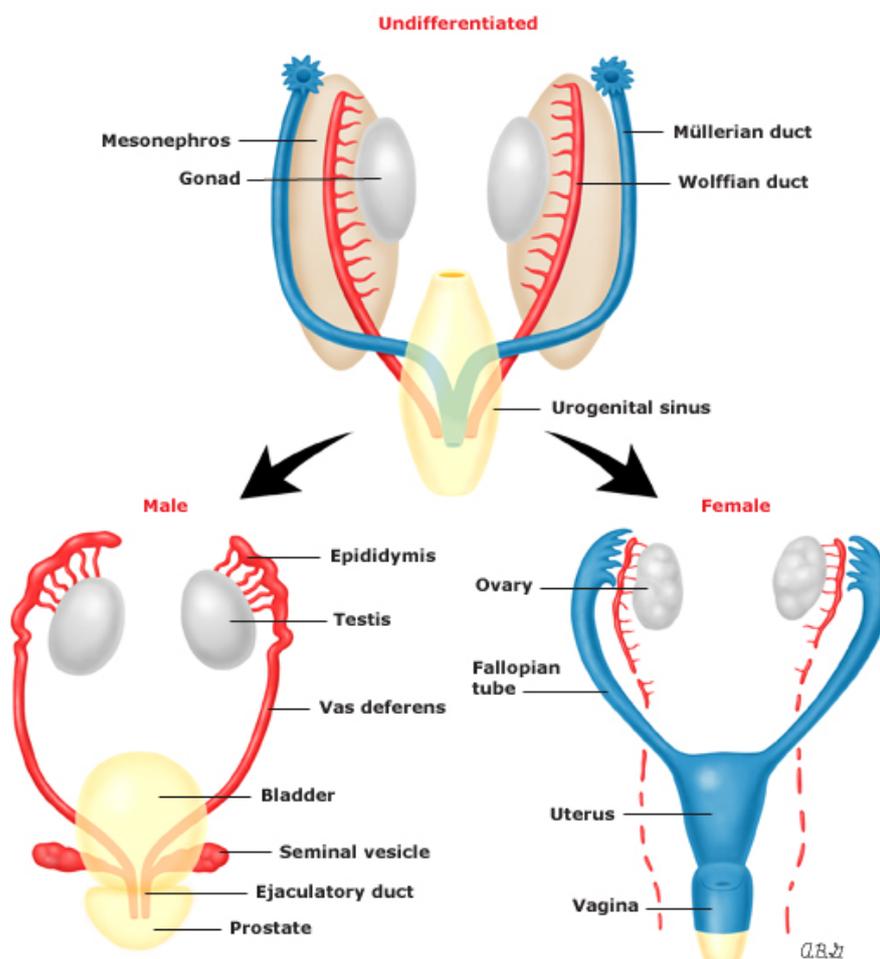


Abb. 1: Entwicklung der männlichen und weiblichen inneren Geschlechtsorgane aus gemeinsamer indifferenten Anlage.

Reproduziert von [9]. Mit freundlicher Genehmigung von UpToDate. Copyright © 2021 UpToDate.

1.2 Varianten der Geschlechtsentwicklung – Differences of sex development (DSD)

Wenn es durch genetische Veränderungen zu Abweichungen von der physiologischen Geschlechtsdeterminierung und/oder -differenzierung kommt, kann es zu einem atypischen chromosomalem, gonadalem oder phänotypischen Geschlecht kommen. Die klinische Präsentation ist dabei hoch variabel und reicht von einem relativ milden Erscheinungsbild, wie beispielsweise einer Hypospadie, bis hin zu einem komplett weiblichen oder männlichen inneren und äußeren Geschlecht, bei gleichzeitigem Vorliegen eines entgegengesetzten Chromosomensatzes. Diese von der physiologischen Geschlechtsentwicklung abweichenden Erscheinungsbilder, werden seit 2006 auf Grundlage des ‚Chicago Consensus Statement‘ unter dem gemeinsamen Begriff ‚disorders of sex development‘ (DSD) zusammengefasst.[10] Da der Begriff ‚disorders‘ (Störungen) zu Stigmatisierung führen und einen eventuell unangemessenen therapeutischen Imperativ implizieren kann, wird inzwischen zunehmend auch der Begriff ‚differences (Varianten/Unterschiede) of sex development‘ verwendet, etwa vom deutschen Ethikrat.[11]

Weil die möglichen Varianten vielfältig und heterogen sind, bedarf es nicht nur individueller Therapie und Betreuung, sondern auch gesellschaftlicher Rücksicht. Beispielsweise ist eine genaue Geschlechtszuordnung sowohl bei der Geburt, als auch später, nicht immer möglich. Dem wird seit 2018 vom Gesetzgeber dadurch Rechnung getragen, dass nach der Geburt bei der Eintragung ins Geburtenregister ‚divers‘ als drittes Geschlecht eingetragen werden kann.[12] In der vorliegenden Arbeit wird grundsätzlich geschlechtsneutral von Patient:innen gesprochen. Damit sind sowohl Kollektive mit sich männlich oder weiblich identifizierenden Personen gemeint, als auch Personen(gruppen), bei denen die Geschlechtszuordnung nicht eindeutig oder unbekannt ist.

Bis zu 1 von 300 Neugeborenen zeigt bei der Geburt eine Auffälligkeit des äußeren Genitals [13], jedoch nicht alle werden aufgrund ihrer Ausprägung auch dem Begriff DSD zugeordnet [14]. Die Inzidenz von Neugeborenen mit einem uneindeutigen Genital wird dementsprechend auf einen deutlich niedrigeren Wert von 1: 4 500 - 5 500 geschätzt.[15, 16] Genaue DSD-Inzidenzdaten sind jedoch nicht bekannt. Dieser Umstand ist zum einen auf die Seltenheit von DSD, zum anderen aber auch auf die Schwierigkeit zurückzuführen, eine klare klinische

Diagnose zu stellen.[17] Im Rahmen des ‚Chicago Consensus Statement‘ [10] wurde daher die Grundlage für eine einheitliche DSD-Klassifikation geschaffen, um eine allgemeine Nomenklatur zu etablieren, die als Gerüst bei der medizinische Diagnostik, Versorgung und Forschung dient. Die Klassifikation basiert in erster Linie auf dem Karyotyp und gliedert sich, je nach zytogenetischem Befund, in Untergruppen mit entsprechenden Erscheinungsbildern (Abb. 2).[10]

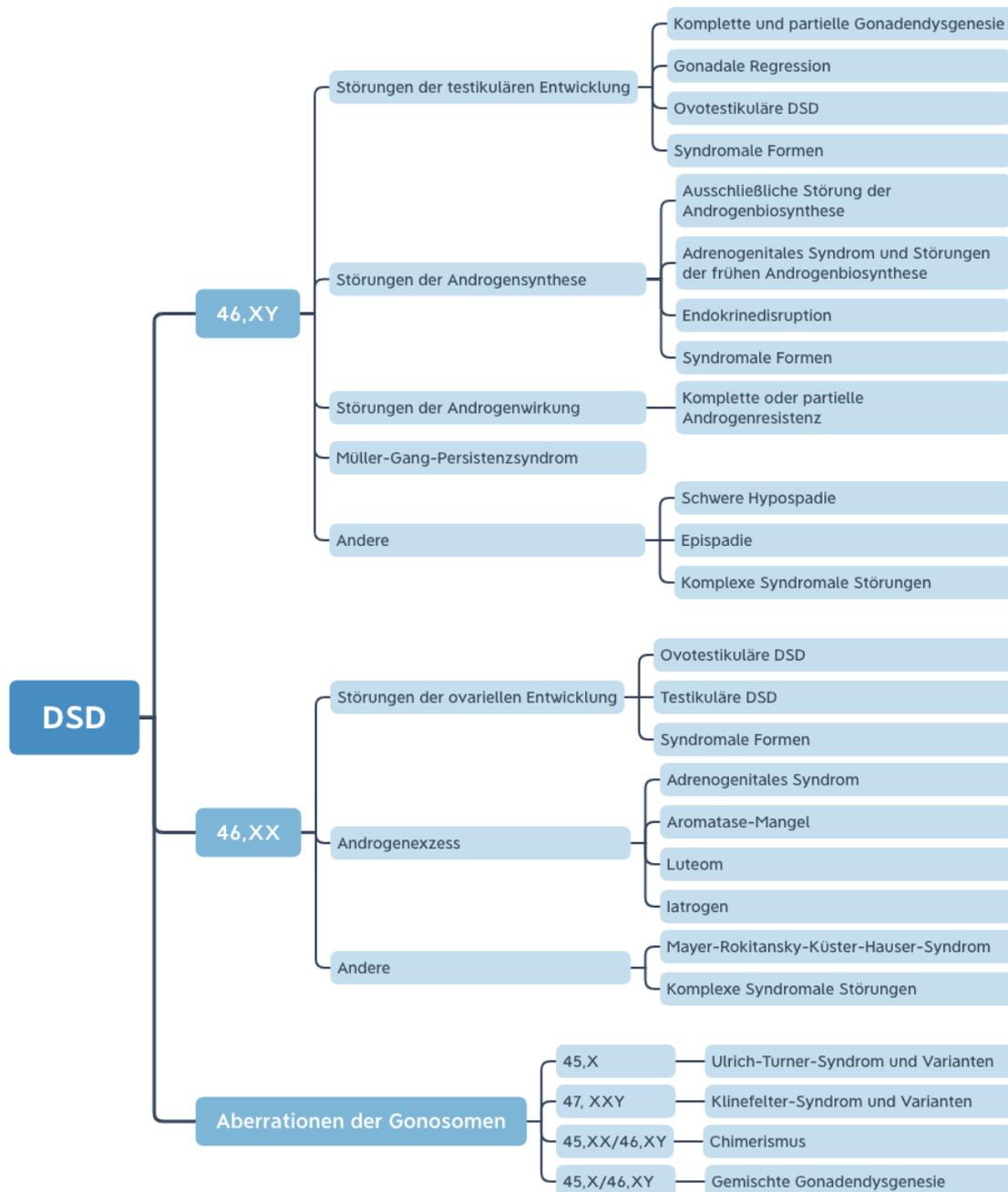


Abb. 2: DSD-Klassifikation. Nach [10] und erweitert durch [18, 19]

Bei Patient:innen mit einer 46,XY-DSD liegt beispielsweise eine Störung der testikulären Entwicklung oder eine unzureichende Maskulinisierung im Rahmen eines Hormonsynthese- oder Hormonwirkungsdefekt vor. Bei den ursächlichen Hormonen kann es sich um Androgene oder das AMH handeln. Unter den testikulären Entwicklungsstörungen sind die komplette und die partielle Gonadendysgenese (GD) hervorzuheben, auf die in der vorliegenden Arbeit ein besonderes Augenmerk gelegt werden soll.

Im Fall einer kompletten 46,XY-Gonadendysgenese liegt bei den Patient:innen ein typisch weibliches Genital und Müller-Derivate vor. Die Gonaden sind entweder undifferenziert oder nicht vorhanden. Bei einer partiellen 46,XY-Gonadendysgenese kommt es hingegen zu einer teilweisen Hodenentwicklung und in der Folge zu einer Mischung aus Wolff- und Müller-Derivaten sowie einem in unterschiedlichem Ausmaß maskulinisierten äußeren Genital. Trotz dieser Kategorisierung in komplette und partielle Gonadendysgenese zeigen sich die Phänotypen in der Realität auch als Zwischenformen mit unterschiedlichen Ausprägungen und sind somit keine klar trennbaren, unverbundenen Kategorien der Hodenfehlentwicklung.[17]

Die Gruppe der 46,XX-DSDs umfasst Varianten der Geschlechtsentwicklung, die bei 46,XX-Karyotyp mit einer vermehrten Virilisierung einhergehen. Die häufigste Ursache ist das Adrenogenitale Syndrom (AGS),[14, 17] welches in 95% der Fälle durch eine Mutation in *CYP21A2* hervorgerufen wird, dem Gen für die 21-Hydroxylase [20].

Die dritte Gruppe von Varianten der Geschlechtsentwicklung, sind die gonosomalen Aberrationen. Hierzu gehören unter anderem das Klinefelter-Syndrom, das Ullrich-Turner-Syndrom und die gemischte Gonadendysgenese, die nach dem AGS der zweithäufigste Grund für ein bei der Geburt uneindeutiges Genital ist.[14]

Bei allen Varianten der Geschlechtsentwicklung stellt die konkrete molekulargenetische Diagnosestellung ein entscheidendes Ziel im Rahmen der Betreuung einer/s Betroffenen und deren/dessen Familie dar. Erst dadurch ist es möglich, prognostische Aussagen und genetische Beratungen anzubieten, sowie individuelle Behandlungskonzepte zu erarbeiten.[19] Ein Beispiel für die Bedeutung der molekulargenetischen Differentialdiagnostik sind etwa Tumore der Gonaden, die bei DSD-Patient:innen gehäuft vorkommen. Inzwischen weiß man, dass diese besonders in Folge von Defekten der frühen Gonadenendeterminierung auftreten –

beispielweise bei Mutationen, die *SRY* oder *WT1* (*Wilms-Tumorsuppressorgen*) betreffen. Deutlich seltener sind Keimzelltumore bei ovotestikulärer DSD oder Störungen der Hormonsynthese oder Hormonwirkung.[19]

Trotz großer Fortschritte in der molekulargenetischen Diagnostik der letzten zwei Jahrzehnte lassen sich weiterhin nur bei ungefähr 40% der Patient:innen mit 46,XY-DSD die genetischen Ursachen aufklären.[16, 21, 22] Dieser Umstand ist auch Ausdruck dessen, dass die Regulation der Geschlechtsentwicklung noch nicht vollkommen verstanden ist. Somit stellt es eine Hauptaufgabe der DSD-Forschung dar, ein genaueres Verständnis dieser Regulation zu erlangen und dadurch eine bessere Verknüpfung von Genotyp und Phänotyp zu ermöglichen. Mit dem Auffinden weiterer pathogener Mutationen in der Zukunft sollte es möglich sein, die individuelle Betreuung verbessern zu können.

1.3 Molekulare Mechanismen der Geschlechtsdeterminierung mit besonderem Augenmerk auf die *NROB1* abhängige *SOX9* Regulierung

Erkenntnisse über die molekularen Mechanismen der Geschlechtsdeterminierung, also Gonadenentwicklung, beruhen in erster Linie auf Untersuchungen von Patient:innen mit angeborenen Varianten der Geschlechtsentwicklung sowie auf Mausmodellen. Weil die komplexe Abfolge der genetischen und molekularen Ereignisse der Geschlechtsdeterminierung unterschiedlich verändert sein kann, können sich in der Folge sehr vielgestaltige geschlechtliche Variationen ergeben. Der Phänotyp dieser Variationen lässt ggf. Rückschlüsse auf die Wirkung einzelner Faktoren der Geschlechtsentwicklung zu.

Abb. 3. vermittelt einen orientierenden Überblick über einige bisher bekannte, an der Geschlechtsdeterminierung beteiligte Gene und Transkriptionsfaktoren, sowie deren Interaktionen. In der vorliegenden Arbeit liegt das Augenmerk dabei insbesondere auf der Regulierung und Wirkung von *NROB1*. Im Sinne der Übersichtlichkeit wird daher im Folgenden nur auf die Beschreibung der *NROB1* abhängigen *SOX9* Regulierung eingegangen.

Ein Meilenstein im Verständnis der menschlichen Geschlechtsdeterminierung war die Entdeckung des *SRY*-Gens [6], welches allein und unabhängig die Entwicklung zum männlichen

Geschlecht initiierten kann. Das direkte Zielgen von *SRY* ist *SOX9* (*SRY-box 9*). Beide gehören zu einer Familie von Genen, die Transkriptionsfaktoren codieren und sich durch eine gemeinsame HMG (high mobility group)-box zur DNA-Bindung auszeichnen.

Das *NROB1*-Gen (*nuclear receptor subfamily 0, group B, member 1*; auch genannt *DAX1* oder *AHCH*) liegt auf dem X-Chromosom und codiert einen nukleären Orphan-Rezeptor, dessen Ligand man also zurzeit noch nicht entdeckt hat. Auch die genaue Wirkung von *NROB1* ist noch nicht umfassend geklärt. Untersuchungen deuten jedoch darauf hin, dass es sich bei *NROB1* um einen Repressor von Gentranskription handelt.[23] Ein weiterer wichtiger Waisen-Rezeptor in der Geschlechtsentwicklung ist *SF1* (*Steroidogenic Factor-1*; auch genannt *NR5A1*, *Nuclear receptor subfamily 5 group A member 1*), welcher wiederum eine aktivierende Wirkung auf die Gentranskription besitzt.[24]

SOX9 ist erforderlich für die Differenzierung von Sertoli-Zellen und die erfolgreiche Hodenentwicklung.[25] Die Aktivierung dieses Gens ist dementsprechend entscheidend, um eine Diskrepanz zwischen chromosomalem (46,XY) und morphologischem Geschlecht zu verhindern.[26] Nach dem auf Mausversuchen basierenden Modell von Ludbrook et al. (2012), verläuft die *Sox9*-Expression im 46,XY Embryo in drei Phasen (Abb.4).[27] Die Expression von *Sox9* beginnt *SRY* unabhängig durch *SF1* (Abb. 4. i), sowohl bei Vorliegen eines 46,XX als auch eines 46,XY-Chromosomensatzes. Nach vorausgegangener Initiierung folgt beim männlichen Embryo eine zusätzliche Heraufregulierung durch eine synergistische Wirkung von *SF1* mit *SRY* (Abb. 4. ii). In der Maus wird *Sry* nur transient für ungefähr zwei Tage exprimiert [28], weil die *Sry*-Expression von *SOX9* wieder unterdrückt wird [25]. Im Menschen hält die *SRY*-Expression hingegen länger an.[29] Die Aktivierung von *Sox9* wird in der Folge über eine Autoregulation durch *SOX9* aufrechterhalten, welches mit *SRY* das DNA-Bindungsmotiv teilt und ebenfalls mit *SF1* synergistisch zusammenwirkt. Dementsprechend ist *SRY* erforderlich, um die *Sox9*-Expression eine kritische Grenze übersteigen zu lassen. Im Sinne einer positiven Rückkopplungsschleife kann es dann zur Aufrechterhaltung der *Sox9*-Expression kommen.[30] Als *Sox9*-Zielgene sind unter anderem *Amh* (*anti-Müllerian Hormone*) [31], *Vnn1* (*Vanin 1*) [32], *Cbln4* (*cerebellin 4 precursor protein*) [33], *Ptgds* (*prostaglandin D2 synthase*) [34] und *Fgf9* (*fibroblast growth factor 9*) [35] bekannt. FGF9 und PTGDS scheinen, wie *SOX9* selbst auch, als positive Rückkopplung die *Sox9*-Expression aufrechtzuerhalten und führen zu einer Herunterregulierung von ovariellen Genen wie *Wnt4* (*Wnt Family Member 4*).[26, 36]

Wenn diese kritische Grenze der *Sox9*-Expression nicht erreicht wird, bleibt die ovarielle Entwicklungskaskade aktiv und die Stützzellen differenzieren sich, statt zu Sertoli-Zellen, zu Granulosazellen des Ovars.[37] NR0B1 stört die SF1-abhängige Aktivierung von Zielgenen, welches entweder durch Protein-Protein-Interaktionen [38] oder die Verdrängung von SF1 von dessen DNA-Bindungsstelle [27] geschieht. Entsprechend reduziert sich die Wirkung von SF1 bei steigender *NrOb1*-Expression. SRY allein ist aber nicht in der Lage, *Sox9* zu aktivieren.[39]

Die Hodenentwicklung hängt somit nicht nur vom bloßen Vorliegen bestimmter Hoden determinierender Faktoren ab, sondern wird maßgeblich auch von der transkriptionellen Dosis und dem korrekten zeitlichen Auftreten der Transkriptionsfaktoren gesteuert.

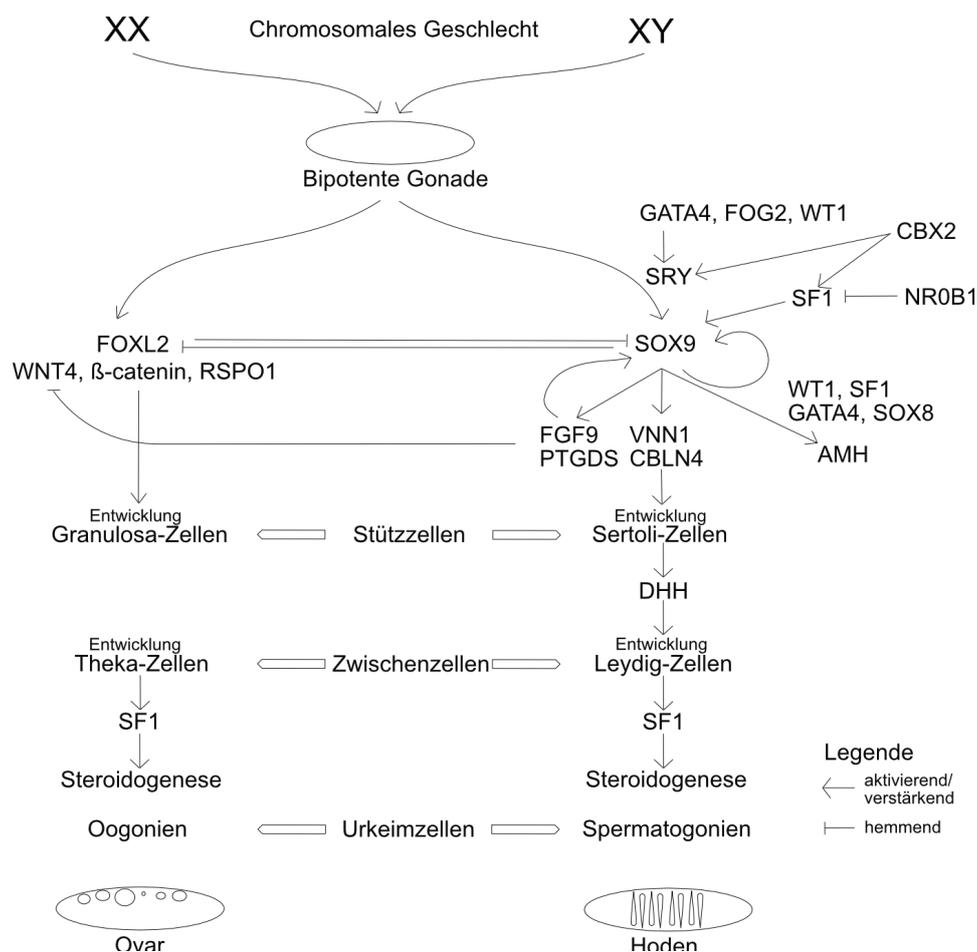


Abb. 3: Vereinfachte Darstellung der Geschlechtsdetermination im Menschen/Maus

Entwicklung der bipotenten Gonade zum Ovar (links) oder Hoden (rechts). FOG2 (Friend of Gata Family Member 2); CBX2 (Chromobox2); FOXL2 (forkhead box L2); RSP01 (R-Spondin 1); SOX8 (SRY-Box transcription Factor 8); DHH (Desert Hedgehog Signaling Molecule); Abbildung nach [2, 40-42].

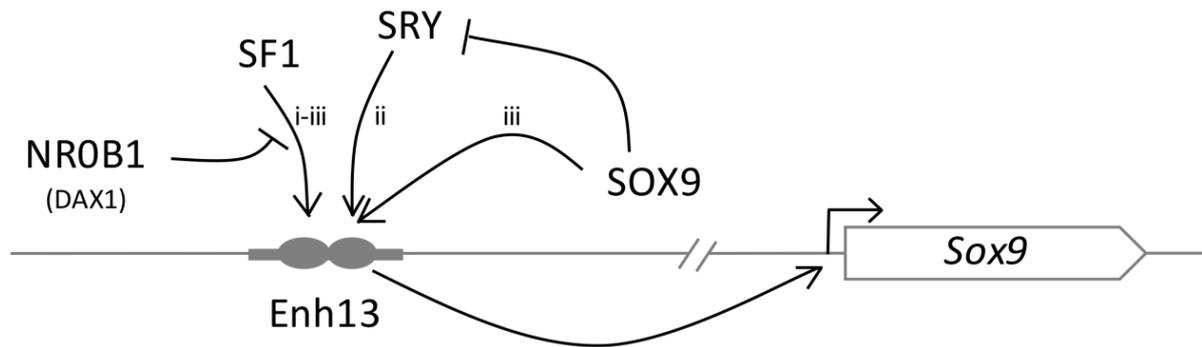


Abb. 4: Sox9-Aktivierung im Rahmen der Hodenentwicklung

Das *Sox9*-Gen ist schematisch mit seinem Enhancer Enh13 [37] (grauer Balken) und den darauf wirkenden Faktoren dargestellt. Die grauen Ovale stellen die SF1 und die gemeinsame SRY/SOX9-Bindungsstelle dar. Neben dem Enhancer Enh13 wurde zuvor in der Maus auch ein als TESCO bezeichneter Enhancer entdeckt.[43] Putativ könnten in den verschiedenen drei Phasen der *Sox9*-Regulierung ggf. unterschiedliche Enhancer aktiviert sein, was nicht Gegenstand dieser Abbildung ist.

1.4 Einteilung molekulargenetischer Mutationen und deren Krankheitsrelevanz

Zur Begriffsklärung, sowie um eine bessere Einschätzung der Häufigkeiten und deren Relevanz im Kontext von DSD zu erlauben, soll hier ein Überblick über gängige genetische Mutationsergebnisse gegeben werden. Mutationen werden in der Regel als Abweichung der untersuchten (Patient:innen-)DNA-Sequenz von der Referenzsequenz erkannt. Die Mutationen können entweder in Keimbahnzellen auftreten, in der Folge vererbt werden und alle Zellen des Nachkommens betreffen, oder in somatischen Zellen entstehen, somit nur eine Subpopulation an Zellen betreffen und keine Auswirkung auf zukünftige Generationen besitzen.[44] Die Mutationen können dabei vielfältige Formen annehmen und in erster Linie in zwei Kategorien eingeteilt werden: Sequenzvariationen und strukturelle Variationen (SV).[45]

Sequenzvariationen und strukturelle Variationen können sowohl die kodierenden (Exone), als auch die nicht-kodierenden Bereiche (Introne, intergenische Bereiche) des Genoms betreffen. Während Veränderungen von exonischen Bereichen in erste Linie Änderungen der Aminosäuresequenz hervorrufen und damit die Proteinfunktion beeinträchtigen können, können Mutationen in nicht-kodierenden Bereichen einen Einfluss auf die Genregulation nehmen, mit Auswirkungen auf den Zeitpunkt, das Gewebe und das Ausmaß einer Genexpression.[46] Regulative Elemente in nicht-kodierenden Bereichen sind beispielsweise Promotoren, Enhancer, Isolatoren, Operons oder Silencer.[46]

1.4.1 Sequenzvariationen

Bei den Sequenzvariationen lassen sich Einzelnukleotid-Variationen (Single Nucleotide Variations – SNVs) von kleinen Insertionen und Deletionen (Indels) unterscheiden. SNVs stellen die Veränderung eines einzelnen Basenpaars im Vergleich zur Referenzsequenz dar und sind mit einer Häufigkeit von 3-4 Millionen die zahlenmäßig verbreitetsten Mutationsereignisse im menschlichen Genom [47, 48]. SNVs sind dabei von SNPs (Single Nucleotide Polymorphism) zu unterscheiden. SNPs sind Einzelnukleotid-Variationen, die mit einer Frequenz von >1% in der Allgemeinbevölkerung vorkommen, wohingegen SNVs keine Häufigkeitsbeschränkung aufweisen.[49]

Der Großteil der SNVs im menschlichen Genom hat keinen funktionellen Einfluss und ist weder molekular noch phänotypisch bemerkbar. Die Unterscheidung zwischen pathogenen und funktionell neutralen SNVs ist dabei nicht immer einfach.[50] Als Orientierung können die exonischen SNVs unter anderem als ‚synonymous‘, ‚missense‘ und ‚nonsense‘ klassifiziert werden; abhängig von ihrer Auswirkung auf die Aminosäuresequenz.[45] ‚Synonymous‘ SNVs verändern zwar das Codon, haben aber aufgrund der Redundanz des genetischen Codes keine Auswirkung auf die in das Protein eingebaute Aminosäure. Mit sehr wenigen Ausnahmen [51] haben sie daher keinen pathogenen Effekt. ‚Missense‘ SNVs verändern hingegen nicht nur die Sequenz eines Codons, sondern auch die in das Protein eingebaute Aminosäure. Die daraus resultierende Auswirkung auf die Funktionalität des Proteins und damit auch der potentiell krankheitsauslösende Charakter der Mutation hängt maßgeblich von den miteinander ausgetauschten Aminosäuren und deren Position im Protein ab. Bei ‚nonsense‘ SNVs entsteht in der Folge der Mutation ein STOP-Codon, welches zum vorzeitigen Abbruch der Translation des Proteins führt. ‚Nonsense‘ SNVs bergen ein besonders hohes Risiko pathogen zu wirken. Jeder Mensch besitzt schätzungsweise ca. 100 dieser Mutationen [52], die teilweise durch das zweite Allel oder andere Gene kompensiert werden können [53].

InDels sind als Insertionen und Deletionen bis zu einer Größe von 50 Basen definiert [48] und treten durchschnittlich 400 000 bis 500 000 Mal im menschlichen Genom auf [47]. Wenn InDels in exonischen Bereichen liegen und ihre Größe nicht einem Vielfachen von 3bp entspricht, verschiebt sich in der Folge der Leserahmen bei der Translation (frameshift). Daraus resultiert zumeist eine starke pathologische Wirkung, sowohl durch die hinter dem Ereignis in ihrer Sequenz veränderte Aminosäurekette, als auch durch das häufige Generieren eines vorzeitigen STOP-Codons. SNVs und InDels in exonischen Bereichen stellen die Gruppe der

häufigsten bekannten krankheitsauslösenden Mutationen dar [53] und sind dementsprechend auch im Bereich von DSD sehr verbreitet.[17, 54]

In nicht-codierenden Bereichen sind neben Veränderungen der bereits oben erwähnten regulativen Elemente der Transkription noch Veränderungen von Spleißstellen durch SNVs und InDels besonders hervorzuheben. Spleißstellen-Mutationen können die posttranslationelle Regulation und damit korrekte Synthese von Proteinen stören. Im Kontext von 46,XY-DSD ist dies beispielsweise in den Genen *SRD5A2* (*3-oxo-5 alpha-steroid 4-dehydrogenase 2*) [55], *17βHSD3* (*hydroxysteroid 17-beta dehydrogenase 3*) [55] oder *AR* (*Androgen Rezeptor*) [56] beschrieben.

1.4.2 Strukturelle Variationen

Strukturelle Variationen (SVs) stellen eine sehr heterogene Gruppe von Genomvariationen dar, denen jedoch gemein ist, dass sie eine Größe von über 50 bp aufweisen.[47, 57, 58] Bei unbalancierten SVs kommt es zu einer Kopiezahlvermehrung (Duplikation, Triplikation etc.) oder Kopiezahlreduktion (Deletion) und diese Ereignisse werden dementsprechend auch Kopiezahlvariationen (copy number variations – CNVs) genannt. Sie könnten in einem Fünftel der DSD-Fälle die zugrundeliegende Ursache sein und treten besonders häufig im Kontext syndromaler Erkrankungen auf.[54] CNVs sind im Kontext der Varianten der Geschlechtsentwicklung an vielen verschiedenen Genorten beschrieben [59], unter anderem im Bereich von *SOX9* [60-62] oder *NROB1* [63-67]. Ihre Wirkung können sie über Genabbrüche oder Genfusionen erreichen, sowie über Auswirkungen auf die 3D-Struktur von DNA-Bereichen mit einer daraus folgenden Beeinflussung von regulativen Prozessen.[68, 69]

Zu den balancierten SVs gehören Inversionen und inter- oder intrachromosomale Translokationen. Hier kommt es zu keiner Veränderung der absoluten Kopiezahl eines DNA-Bereichs. Der betroffene Bereich ändert dabei aber lediglich seine Ausrichtung bzw. Position im Genom. Zusätzlich gibt es noch die Gruppen der high-copy repeats und mobilen genetischen Elemente, die auch zu den SVs gezählt werden können [47, 57]. Aber bisher wurden diese nur sehr selten als Ursache von genetischen Erkrankungen beschrieben.[70] Im Kontext der Varianten der Geschlechtsentwicklung können high-copy repeats bei der Androgenresistenz eine Rolle spielen.[71] Mischungen verschiedener SV-Arten werden als komplexe SVs bezeichnet.

Die Häufigkeit der pro Person identifizierten SVs hängt maßgeblich von der angewendeten Technik ab. Anfang des 21. Jahrhunderts ließen sich pro Person im Schnitt noch 12 SVs nachweisen.[72] Inzwischen lassen sich mit Methoden des ‚Long-read whole genome sequencings‘ bis zu 20 000 SVs nachweisen. Mit der in der vorliegenden Arbeit verwendeten Methode des ‚Short-read sequencing‘ können, je nach Auswertungsalgorithmus, bis zu 10 000 SVs entdeckt werden. Die Unterschiede beruhen jedoch in erster Linie auf der Erkennung von sehr kleinen SVs (<1kb) in repeat-reichen nicht-kodierenden Bereichen [73], vermutlich ohne pathogene Relevanz. In der Regel trägt jeder Mensch weniger als 50 SVs, welche exonische Bereiche umspannen, wie sie in dieser Arbeit beschrieben werden.[58, 74]

1.5 Fragestellung und Zielsetzung

Seit Mitte der 90er-Jahre ist bekannt, dass 46,XY-Gonadendysgenesien (GD) genetisch mit Duplikationen auf dem kleinen Arm des X-Chromosoms (Xp21.2) assoziiert sein können.[75] In der Folge wurde der Begriff ‚Dosage Sensitive Sex reversal‘ geprägt und die pathologische Wirkung der Duplikationen dem doppelten Vorliegen des in der Region lokalisierte *NROB1* zugeschrieben, und der daraus resultierenden doppelten *NROB1*-Gendosis.[65, 75]

Im Rahmen einer Kooperation mit der Universität Campinas, Sao Paulo / Brasilien, ist nun im Frühjahr 2018 eine Patient:in mit 46,XY-GD aufgefallen, die eine Duplikationen am Xp21.2-Locus aufweist, welche jedoch nicht *NROB1* beinhaltet und somit nicht dem gängigen Verständnis des Pathomechanismus der Duplikationen am *NROB1*-Locus entspricht.

Im Rahmen der vorliegenden Arbeit sollte unter Anwendung von Whole Genome Sequencing (WGS) die in Brasilien gefundene Duplikation in Bezug auf ihre exakte Größe und Anordnung genauer charakterisiert werden. Außerdem ging es darum ggf. andere bekannte Ursachen als molekulargenetischen Auslöser der 46,XY-GD auszuschließen.

Es bestätigte sich, dass die Duplikation der brasilianischen Patient:in stromaufwärts von *NROB1* beginnt und die wahrscheinlichste Ursache für die Gonadendysgenese darstellt.

Möglicherweise sind Duplikationen stromaufwärts von *NROB1* ein verkannter und übersehener Auslöser für 46,XY-GD, da sich zum einen die molekulargenetische Diagnostik bei DSD-

Patient:innen häufig auf das Sequenzieren einzelner Kandidatengene konzentriert oder – in der jüngeren Vergangenheit – Whole Exome Sequencing (WES) für die Diagnose nutzt, womit Duplikationen allerdings nicht sicher nachweisbar sind. Sofern Methoden zur Bestimmung von Kopiezahlveränderungen angewendet wurden, beschränkte sich die Auswertung in der Vergangenheit auf *NROB1*, so dass Mutationen in der weiteren Umgebung von *NROB1* bei der Analyse womöglich nicht auffallen konnten.

Vor diesem Hintergrund und in Anbetracht des Umstands, dass nur etwas über 40% der Patient:innen mit 46,XY-DSD eine mutmaßliche genetische Diagnose erhalten [21], stellte sich die Frage, ob es sich bei den Duplikationen stromaufwärts von *NROB1* um eine bisher unterdiagnostizierte Ursache der 46,XY-GD handelt. Um dies zu untersuchen und weitere Fälle zu identifizieren und so mögliche Hinweise auf den zugrundeliegenden Pathomechanismus zu erhalten, wurden in einem ersten Schritt die WGS-Daten von weiteren Patient:innen mit 46,XY-GD auf Duplikationen in diesem Bereich analysiert. In einem zweiten Schritt wurde ein adäquates Screening etabliert. Mit diesem Screening sollten bisher ähnliche ungeklärte Fälle neu aufgerollt werden und zukünftig die Diagnostik von Kopiezahlveränderungen sowohl in- als auch exklusive *NROB1* am Xp21.2-Lokus in der Klinik als einfache Option zu etablieren.

Die Zusammenschau mehrerer Fälle ermöglichte darüber hinaus ein besseres Verständnis der Wirkung von Kopiezahlveränderungen und die Untersuchung eines möglichen gemeinsamen Pathomechanismus von Kopiezahlveränderungen am Xp21.2-Locus.

2. Material und Methode

2.1 Patient:innen

In der vorliegenden Arbeit werden drei besondere Patient:innen klinisch beschrieben und detailliert genetisch analysiert. Die erste Patient:in stammt aus Brasilien und wurde im Rahmen einer Kooperation zwischen der Universität Campinas, Sao Paulo und der Universität zu Lübeck wissenschaftlich genauer untersucht. Die beiden anderen Patient:innen werden am Universitätsklinikum Schleswig-Holstein (UKSH) Campus Lübeck betreut. Alle drei Patient:innen bzw. deren Sorgeberechtigten haben in die Teilnahme an der Studie eingewilligt. Die Arbeit erfolgte mit den Zustimmungen der Ethikkommissionen der Universität zu Lübeck und der Universität Campinas, Sao Paulo (siehe Anhang, S. 92).

Patient:in 1 (P1)

Nach komplikationsloser Schwangerschaft und Geburt präsentierte das Neugeborene (Patient:in 1) ein auffälliges Genital. Im Alter von 5 Tagen wurde es daher in der DSD-Klinik der Universität Campinas, Sao Paulo vorgestellt. Bei der dortigen Untersuchung zeigte sich ein 0,5 cm langer Phallus, eine einzelne perineale Öffnung, teilweise fusionierte Labioskrotalfalten und nicht-tastbare Gonaden. Der Ultraschall ergab keine Auffälligkeiten. In der Genitographie stellte sich ein persistierender Sinus urogenitalis dar. Der Karyotyp wurde als 46,XY bestimmt und in einer Laparoskopie konnte kein Uterus und keine linke Gonade dargestellt werden. Laborchemisch zeigte sich im 1. Lebensmonat ein erhöhtes FSH (24 IU/l) und LH (10 IU/l) und niedriges Testosteron (0.2 ng/ml). Schließlich wurde die Diagnose ‚partielle Gonadendysgenese‘ gestellt. Im Rahmen der weiteren molekulargenetischen Abklärung mittels CMA wurde eine Duplikation von *CXorf21* (*Chromosome X Open Reading Frame 21*; auch genannt *TASL*, *TLR Adapter Interacting With SLC15A4 On The Lysosome*) am Xp.21.2-Locus gesehen, die sich mittels MLPA bestätigen ließ.

Patient:in 2 (P2)

Die Patient:in 2 besuchte erstmals im Alter von 17 Jahren die DSD-Sprechstunde des UKSH Lübeck. Zuvor wurde bei ihr bereits ein symptomatisches Dysgerminom (links) entfernt und zytogenetisch ein 46,XY-Karyotyp festgestellt. Ihre Pubertätsentwicklung begann verzögert mit 14 Jahren und präsentierte sich in erster Linie durch Schamhaarentwicklung. Es lag eine Amenorrhoe vor und erst kurz vor Diagnose des Dysgerminoms begann mit 16 Jahren das Brustwachstum. Äußerlich erschien das Genital weiblich und ohne Virilisierungszeichen,

Tanner-Stadium B4, P4-5. Müller'sche Derivate ließen sich nachweisen, insbesondere ein Uterus, welcher aber im Vergleich zum Alter der Patient:in klein und von weitgehend tubulärer Form war. Nach der Entfernung des Dysgerminoms zeigte sich labordiagnostisch funktionell ein Hypergonadotroper Hypogonadismus mit auffällig hohem Testosteron (123 ng/dl), was dem unteren Bereich männlicher Pubertätsentwicklung entsprach. Ferner wurde mit 0,71 µg/l ein deutlich zu niedriger Spiegel des Anti-Müller Hormons (AMH) gemessen, was in der Zusammenschau zur Diagnose einer ‚kompletten Gonadendysgenese‘ führte. Im Sprechstundengespräch wurde deutlich, dass sich die Patient:in mit dem weiblichen Geschlecht identifiziert. Im weiteren Verlauf der Behandlung wurde dann auch noch die vorhandene rechte Gonade entfernt und histologisch ein Gonadoblastom nachgewiesen. Eine molekulargenetische Untersuchung des Androgen-Rezeptors war unauffällig. Bei einem im Rahmen der vorgelegten Arbeit durchgeführten Whole Genome Sequencing ließ sich eine Duplikation am Xp21.2-Locus nachweisen.

Patient:in 3 (P3)

Die dritte Patient:in wurde ebenfalls nach einer komplikationslosen Schwangerschaft geboren und wies postnatal keine Auffälligkeiten des weiblichen Genitalbereichs auf. Auch eine spätere Familienanamnese zeigte sich unauffällig in Bezug auf Varianten der Geschlechtsentwicklung. Im Rahmen einer Vorsorgeuntersuchung fiel bei dem Kind jedoch eine Hypotonie auf, speziell der Extremitäten und des Gesichts. Die veranlasste molekulargenetische Untersuchung ergab als Zufallsbefund einen 46,XY-Chromosomensatz. Sonographisch offenbarte sich in der Folge ein präpubertärer Uterus und laborchemisch ein niedriger AMH-Spiegel (0,42 µg/l) sowie Inhibin B unter der Nachweisgrenze. Es wurde wie bei Patient:in 2 die Diagnose einer ‚kompletten Gonadendysgenese‘ gestellt und die Patient:in zur DSD-Sprechstunde des UKSH Lübeck überwiesen. Im Alter von 6 Jahren wurde dann eine beidseitige Gonadektomie durchgeführt. Die histologische Aufarbeitung zeigte rechts ein Gonadoblastom ohne Anzeichen für eine Transformation in einen malignen Keimzelltumor. Links präsentierte sich tumorfreies ovarielles Stroma. Bei der molekulargenetischen Untersuchung war das *SRY*-Gen unauffällig.

Im Kontext des umfangreichen Screenings von DSD-Patient:innen im Rahmen der vorliegenden Arbeit konnte bei der Patient:in 3, wie bei P1 und P2, eine Kopiezahlvariation am Xp21.2-Locus identifiziert und sie so im späteren Verlauf der Arbeit als Studienpatient:in eingeschlossen werden.

2.2 Whole Genome Sequencing (WGS)

2.2.1 Prinzip

Zur Sequenzierung des (annähernd) kompletten Genoms der einzelnen Patient:innen wurde im Rahmen einer Auftragssequenzierung die Illumina HiSeq Plattform (Illumina, Kalifornien, USA) benutzt, welche zu den Systemen der Hochdurchsatzsequenzierung zählt – auch ‚Next Generation Sequencing‘ (NGS) genannt. Die Methoden des NGS sind dadurch gekennzeichnet, dass Sequenzierungen vieler Millionen von DNA-Fragmenten gleichzeitig stattfinden können. Das Illumina WGS erfolgt dabei im Wesentlichen in vier Schritten. Die vier Schritte bis zur Datenaufbereitung erfolgten im Rahmen einer Auftragssequenzierung durch Novogene (Hong Kong, China).

Schritt 1: Erstellung einer DNA-Fragmentbibliothek (Library preparation)

Zur Erstellung einer DNA-Fragmentbibliothek kommt das Truseq Nano DNA HT Sample preparation Kit (Illumina, Kalifornien, USA) zur Anwendung, nachdem zuvor 1 µg isolierter DNA mittels eines Covaris cracker (Covaris, Massachusetts, USA) zufällig in ca. 200-500 bp große dsDNA Fragmente zerkleinert wurde. Überhänge am 3' und 5' Ende (Abb. 5A). werden durch 3'→5' Exonucleasen und 5'→3' Polymerasen in glatte Enden überführt (Abb. 5B), welche danach am 3'-Ende adenyliert werden (Abb. 5C). Wie bei der Methode der Library preparation üblich, ist durch die Adenylierung eine spezifische Ligation mit Adaptern möglich, die ihrerseits einen komplementären Thyminüberhang besitzen. (Abb. 5D). So wird eine Ligation zwischen zwei DNA-Fragmenten der genomischen DNA verhindert. Die Adapter bestehen aus einer Sequenz, die der Befestigung an der gläsernen Flusszelle (Flow-Cell) dienen, in der die Sequenzierungsreaktionen stattfinden. Außerdem verfügen die Adapter über eine 6bp lange Erkennungssequenz (Barcode) und eine Bindungsstelle für den Sequenzierungsprimer. Die Erkennungssequenz erlaubt es, die DNA verschiedener Proben in derselben Spur einer Flusszelle zu sequenzieren. So können später die abgelesenen Sequenzen anhand der Erkennungssequenz wieder der entsprechenden Ursprungsprobe zugeordnet werden. Nach dem erfolgreichen Verbinden der Fragmente mit den Adaptern erfolgt in einem nächsten Schritt die Amplifizierung mittels Polymerase-Kettenreaktion (polymerase chain reaction, PCR). Im Anschluss an die PCR wird die anfangs erstellte Bibliothek (Library) mittels AMPure XP (Kalifornien, USA) aufgereinigt. Die Qualität der Library, in Form der Größenverteilung der

Fragmente, lässt sich mittels eines Agilent 2100 Bioanalyser (Agilent Technologies, Kalifornien, USA) überprüfen. Die Quantitätsbestimmung erfolgt über qPCR.

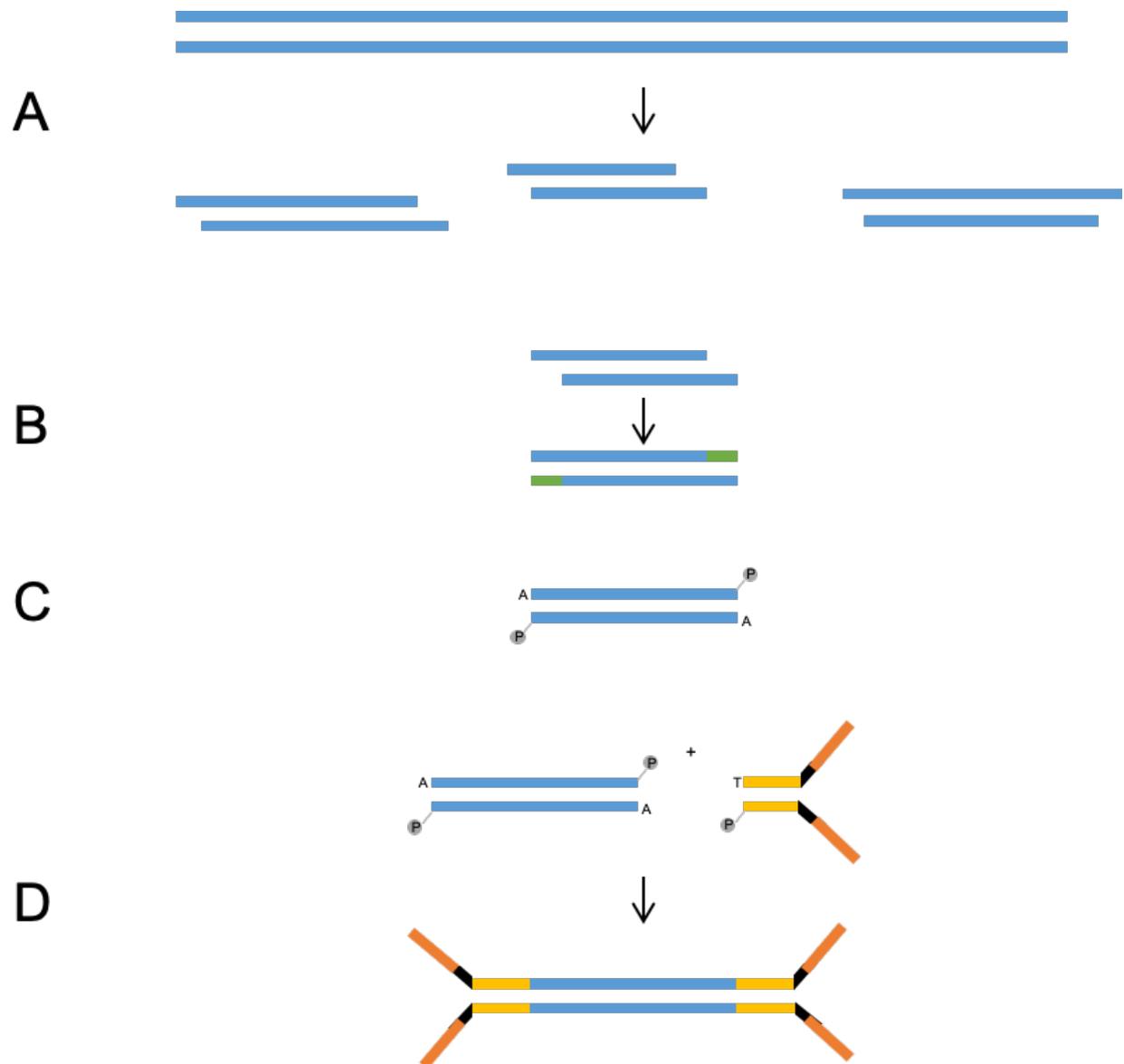


Abb. 5: Erstellung einer DNA-Fragmentbibliothek für das Whole Genome Sequencing.

A) Mechanische Scherung der genomischen DNA mittels Ultraschall in Fragmente. B) Erzeugen von glatten Enden (End Repair). C) Adenylierung der 3'OH-Enden (A-Tailing). D) Beidseitige Ligation der DNA-Fragmente mit Adaptern. Die Adapter bestehen aus Bindungsstellen für Sequenzierungsprimer (gelb), einer patient:innenspezifischen Erkennungssequenz (schwarz) und dem Bindungsbereich für die Flusszelle (orange). Abbildung modifiziert nach [76, 77].

Schritt 2: Gruppengenerierung (Clustering)

Die erstellte Fragmentbibliothek (s.o.) wird nun in die Flusszelle gegeben und kann als ssDNA über die Adapter an die auf der Oberfläche der Flusszelle befestigten Oligos binden (Abb. 6A). Mit dem Oligo als Ursprung lässt sich über eine Polymerase der komplementäre DNA-Strang synthetisieren (Abb. 6B). In einer folgenden DNA-Denaturierung löst sich das Originalfragment vom Komplementärstrang und wird ausgewaschen (Abb.6C). Danach beginnt die sogenannte Brückenamplifikation (Bridge Amplification). Dabei bindet das noch ungebundene Sequenzende mit dessen Adapter an einen anderen in der Flusszelle befestigten Oligotyp und bildet so eine Brücke zwischen zwei Oligos. Entlang der Brücke wird der komplementäre Doppelstrang synthetisiert (Abb. 6D & E). Nach einer erneuten Denaturierung wiederholt sich der Prozess vielfach (Abb. 6F & G), wodurch es zu einer klonalen Amplifikation in Gruppen mit je ca. 1 000 Kopien kommt (Abb. 6H). Nach Entfernung der Reverse-Stränge verbleiben nur noch identische Forward-Stränge in einer Gruppe. So entsteht ein stärkeres, einheitliches Signal bei der folgenden Sequenzierung.

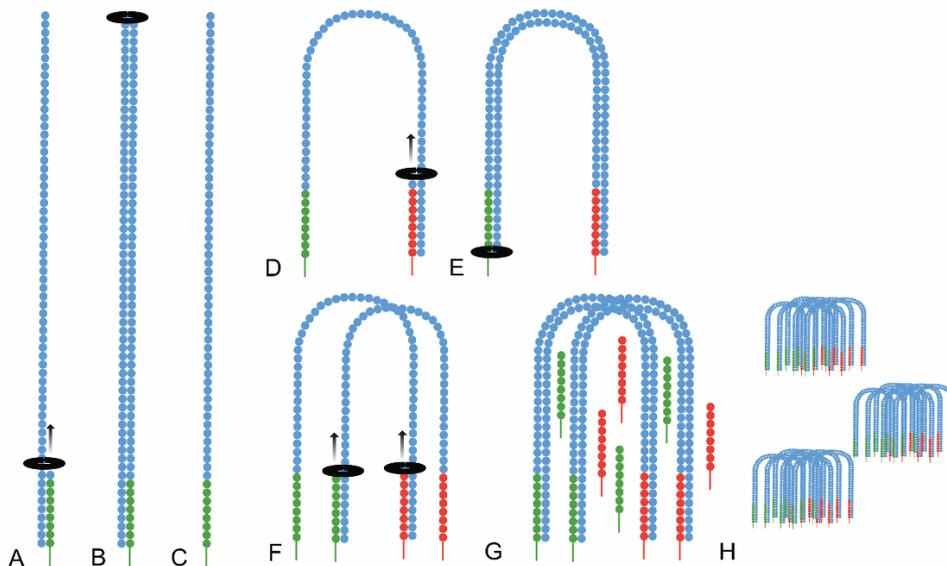


Abb. 6: Brückenamplifikation in der Flusszelle.

Blaue Ketten stellen die mit Adaptern versehenen zu sequenzierenden DNA-Fragmente dar. Grüne und rote Ketten sind die auf der Flusszelle verankerten Oligos, die der Fixierung der DNA-Fragmente dienen. Der schwarze Kreis repräsentiert die DNA-Polymerase, die komplementäre DNA-Stränge synthetisiert. Abbildung modifiziert nach [76, 77].

Schritt 3: Sequenzierung

Illumina nutzt das ‚sequencing by synthesis‘ Prinzip. Hierfür bindet ein Primer an die entsprechende Bindungsstelle des Adapters. An diesem Primer erfolgt dann die Neusynthese eines DNA-Stranges mittels fluoreszenzmarkierter Nukleotide. Da 4-Farb-Sequenzierung durchgeführt wird, sind die Nukleotide je nach Base farblich unterschiedlich markiert. Eine Gruppe erscheint demnach in der Farbe des an der entsprechenden Stelle in allen Strängen synchron eingebauten Nukleotids. Die Farbe wird optisch detektiert und daraus die Sequenz abgeleitet. Eine 3'-Schutzgruppe am Nukleotid führt zum reversiblen Kettenabbruch und gewährleistet damit, dass in jedem Zyklus nur ein Nukleotid eingebaut wird. Vor Beginn des nächsten Zyklus wird die Schutzgruppe mit dem Farbstoff abgespalten. Mit zunehmender Sequenzlänge akkumulieren Fehler, entweder durch den fehlenden oder mehrfachen Einbau einer Base in einem Zyklus. Dadurch kommt es zu einer Verschiebung der Leseposition und des entsprechenden Farbsignals. Dieses zunehmende Auseinanderdividieren des Signals einer Gruppe begrenzt die maximale Länge einer abgelesenen Sequenz (Read). Üblich ist daher eine Länge von 150bp, wie es auch in dieser Arbeit der Fall ist. Die Reaktion endet dementsprechend nach 150 Zyklen. Danach wird das DNA-Fragment nochmals aus der anderen Richtung sequenziert, wodurch aus den beiden in entgegengesetzte Richtung orientierten Reads ein Read-Paar entsteht. Diese beiden Partner-Reads können je nach Länge des DNA-Fragments teilweise überlappen oder auch gänzlich voneinander getrennte Sequenzen beinhalten. Die beiden Partner-Reads werden häufig mit /1 und /2 gekennzeichnet und bleiben miteinander verknüpft, da von ihnen bekannt ist, dass sie durch die Brückenamplifikation aus nahegelegenen komplementären DNA-Bereichen entstehen.

Schritt 4: Datenaufbereitung

Methodisch werden die Reads mittels der Erkennungssequenz (Barcode) im Adapter den entsprechenden Proben zugeordnet. In der vorliegenden Untersuchung entstanden pro Patient:in über 90 GB Datenmenge und es wurde eine durchschnittliche Sequenzierungstiefe von x30 erreicht. Das heißt, jeder Base im Genom wurden im Durchschnitt 30 Reads zugeordnet, die diese bestimmte Base sequenziert haben.

Zunächst werden Read-Paare mit anderen Read-Paaren zusammen geordnet, mit denen sie Sequenzbereiche teilen (Alignment). Dies erzeugt Gruppen von überlappenden Reads, aus denen sich dann längere DNA-Sequenzen ergeben. Solche Gruppen können dann anhand des menschlichen Referenzgenoms ausgerichtet und aneinandergereiht werden (Mapping). Im

Fall dieser Arbeit diene als menschliche Referenzgenomsequenz GRCh37/hg19, unter Verwendung des Burrows–Wheeler Aligner (BWA) [78].

Für die Auswertung ist wichtig: Wenn ein Read zu Teilen zwei entfernten DNA-Bereichen zugeordnet werden kann, dann handelt es sich um einen sogenannten ‚Split-Read‘ oder ‚chimären Read‘. Wenn dies mehrere Reads in einem DNA-Bereich betrifft, deutet dies auf eine Abweichung der Patient:innensequenz von der Referenzsequenz durch ein chromosomales Rearrangement (Inversion, Deletion, Duplikation) hin.

2.2.2 SNV-Analyse

Um bekannte Ursachen einer 46,XY-DSD bei den Patient:innen auszuschließen, wurden die durch das WGS aufgedeckten SNVs (Single Nucleotide Variants) und InDels (Inversions/Deletions) analysiert. Diese wurden mittels GATK Version 3.8.0 [79] aus BAM Files ermittelt und durch das Programm ANNOVAR [80] funktionell annotiert. Zunächst sind Mutationen mit einer Allelfrequenz < 1% in der Gesamtbevölkerung extrahiert worden. Als Grundlage für die Allelfrequenz fanden die Daten der ExAc-Datenbank Verwendung.[52] Dies lieferte bei den Patient:innen zwischen 5 000 und 7 000 Mutationen. Diese Mutationen wurden dann mit einer Liste von ca. 350 DSD-Kandidatengenen abgeglichen. Da es sich bei 46,XY-DSD um ein sehr seltenes Erkrankungsbild handelt, erfolgte zusätzlich ein Filtern nach Mutationen mit einer Häufigkeit < 0,01%. Sofern auffällige Mutationen vorlagen, wurde nochmals nach zusätzlichen, und sich so eventuell gegenseitig verstärkenden Mutationen in demselben Gen geschaut. Die Basenveränderungen wurden mit drei gängigen Programmen zur Bewertung von deren pathogenen Auswirkung auf die Proteinstruktur eingeschätzt – SIFT [81], PolyPhen2 [82] und Mutationtaster [83]. Um als potentiell krankheitsauslösend in Betracht gezogen zu werden, sollten zwei von drei Programmen diese als ‚deleterious‘ oder ‚possibly deleterious‘ einstufen. Ferner wurde überprüft, ob der Phänotyp der Patient:in zu dem mit dem Gen in der Literatur beschriebenen Phänotyp passte.

2.2.3 CNV-Analyse

Zur Identifikation von Kopiezahlvariationen fand das Programm CONTROL-FREEC [84] Anwendung. Dabei erfolgt die Einteilung grundsätzlich in ‚gain‘, ‚loss‘ und ‚loss of heterozygosity‘ mit den zugehörigen Chromosomenbereichen. Zusätzlich kam das Programm DELLY [85] zur

Erkennung von strukturellen Variationen (SVs) zum Einsatz. Neben dem Erkennen von Duplikationen und Deletionen markiert das Programm auch Inversionen und reziproke Translokationen. Von den Programmen bei den betrachteten Patient:innen identifizierte Kopiezahlvariationen und SVs wurden im ‚Integrative Genomics Viewer‘ [86] visualisiert, um das Vorliegen eines CNVs anhand einer veränderten Sequenzierungstiefe und erkennbaren chimären Reads an den Übergängen zu bestätigen.

2.2.4 Integrative Genomics Viewer (IGV)

Für die Speicherung der Position der ausgerichteten Reads und der sich daraus ergebenden Sequenz wird in der Regel entweder ein Textformat genutzt: das Sequence Alignment Map (SAM) [87] oder das komprimierte Binary Alignment Map (BAM). Im Integrative Genomics Viewer [86] lassen sich BAM-Files öffnen, und so die ausgerichteten Reads visualisieren. Die Software erlaubt einem schnellen Überblick über die Anzahl der Reads, die einem bestimmten ausgewählten Chromosomenbereich zugeordnet sind (Sequenzierungstiefe). Auch SNVs sind farblich hervorgehoben und die Anzahl der Reads erkennbar, die dieses beinhalten. Somit lassen sich u.a. auch heterozygote von homozygoten SNVs unterscheiden.

2.2.5 Bruchpunktbestimmung aus WGS Reads

Beim Mapping wurden die Reads anhand der menschlichen Referenzsequenz ausgerichtet und dabei an diese Vorlage angeglichen. In der Konsequenz heißt das, bei Patient:innen, die beispielsweise eine Duplikation aufweisen, werden die Reads dem Chromosomen-Bereich zugeordnet, aus der die Duplikation entstammt und nicht so angeordnet, wie die Duplikation im Patient:innengenom eingebaut ist. Erkennbar wird das an den Bruchpunkten von strukturellen Variationen, da hier chimäre Reads entstehen. In den Reads gehen Sequenzbereiche ineinander über, die normalerweise voneinander getrennt sind. Diese chimären Reads haben somit zwei Sequenzbereiche, die beim Mapping zwei verschiedenen Orten zugeordnet werden können. Dies wird teilweise als Fehler erkannt und als ein verkürzter und farblich markierter Read im IGV dargestellt und kann helfen, Bruchpunkte zu ermitteln.

Um die tatsächliche Anordnung der Kopiezahlvariationen bei den konkreten Patient:innen zu identifizieren, wurden sämtliche Reads inklusive ihrer Partner-Reads an den Bruchpunkten extrahiert. Dies erfolgte manuell im IGV über die ‚Copy Read Sequence‘-Funktion. Die

Ermittlung von Read-Paaren erleichterte das Verknüpfen der Übergänge von Bruchpunkten, die im IGV erkennbar waren. Wechsel vom Sense- zum Antisense-Stang innerhalb eines chimären Reads gaben Aufschluss über das Vorliegen von Inversionen. So konnten jeweils Modelle der Anordnung der strukturellen Variationen hergeleitet werden. Die extrahierten Sequenzen der Reads wurden zusätzlich über SeqEdit (DNASTAR, Wisconsin, USA) verwaltet und für ein eigenständiges Assembly ohne Vorlage einer Referenzsequenz im Programm SeqMan (DNASTAR, Wisconsin, USA) aufbereitet. Die dabei generierten Bruchpunktsequenzen bestätigten jeweils die zuvor entworfenen Modelle.

2.3 Array-Based Comparative Genomic Hybridisation (aCGH)

Erstmals beschrieben Kallioniemi et al. 1992 diese Methode zur Identifikation von Kopiezahlvariationen bei Tumorproben.[88] Die Patient:innen- und Kontrollproben wurden dabei mit unterschiedlichen Farbstoffen markiert und konkurrieren um Hybridisierungsstellen an fixierten komplementären Metaphasenchromosomenbereichen. Heutzutage werden Arrays von Oligonukleotiden zur Hybridisierung der Proben- und Kontroll-DNA verwendet. Das Verhältnis der Fluoreszenz der beiden Farbstoffe wird gemessen. Die relative Menge der gebundenen Patienten- und Kontrollproben DNA ist abhängig von der relativen Häufigkeit des Vorliegens dieses Genlocus in den Proben und entsprechend auch die gemessene Fluoreszenz. Wenn bei Patient:innen eine Kopiezahlvermehrung vorliegt, hybridisiert die Probe entsprechend häufiger und das Verhältnis zwischen der Fluoreszenz der Patient:innenprobe und der Kontrollprobe verschiebt sich in Richtung der Farbe der Patient:innenprobe. Bei einer Deletion kommt es umgekehrt zu einer Verschiebung des Fluoreszenzverhältnisses zugunsten der Kontrollprobe.

Die Proben der konkret untersuchten Patient:innen 2 und 3 sowie der Mutter von Patient:in 3 wurden durch das Institut für Humangenetik der Universität zu Lübeck auf einem Agilent 180K Array (Agilent Technologies, Kalifornien, USA) hybridisiert und mit einer gepoolten Kontrollprobe aus 10 normalen Männern (P1 und P2) bzw. 10 normalen Frauen (Mutter von P3) verglichen. Die Datenanalyse erfolgte mit der CytoGenomics Software version 4.0.2.21 (Agilent Technologies, Kalifornien, USA).

2.4 Bruchpunktbestätigung

2.4.1 Primerdesign

Um die aus den WGS-Daten abgeleiteten Bruchpunktsequenzen zu bestätigen, mussten diese spezifisch amplifiziert und im Anschluss sequenziert werden. Die hierfür nötigen Primerpaare konnten mit der Software PrimerSelect 14 (DNASTAR, Wisconsin, USA) berechnet werden. Hierzu wurden die Sequenzen der ‚forward‘ und ‚reverse‘ orientierten chimären Reads zusammen mit weiteren benachbarten, die jeweiligen chimären Read-Sequenzen überlappenden Reads aus dem IGV extrahiert. Mit Hilfe der Software Megalign14 (DNASTAR, Wisconsin, USA) wurden die Reads dann zu einer linear angeordneten Sequenz stromaufwärts und stromabwärts der Bruchpunktsequenz zusammengesetzt. Ausgewählte Primer erfüllten die Bedingung, mindestens 100 bp vom Bruchpunkt entfernt zu binden. Mittels BLAT-Suche [89] konnte die Spezifität der 19 bis 24 Nukleotid langen Primersequenzen geprüft werden. Primer wurden über die Firma metabion international AG (Planegg, Deutschland) bezogen. Eine Übersicht der Primersequenzen findet sich im Anhang auf S. 103. Mit Hilfe des Programms SeqBuilder 14 (DNASTAR, Wisconsin, USA) erfolgte die Dokumentation.

2.4.2 Polymerase-Kettenreaktion (PCR)

Die entworfenen Primer wurden mit thermostabiler (*Thermus aquaticus*) Taq-Polymerase zur Polymerase-Kettenreaktion [90] verwendet, um spezifisch die Bruchpunktsequenzen zu amplifizieren. Eine Gradienten-PCR diente der Ermittlung der optimalen Annealing-Temperatur im Temperaturbereich zwischen 50-70°C. Zum Nachweis des *SRY*-Gens ließen sich in der PCR-Diagnostik bereits etablierte Primer zur Amplifikation eines Segments in Exon 1 verwenden.

Für die PCR bzw. Gradienten-PCR befanden sich in den Ansätzen Primer mit je einer Konzentration von 0,2 pmol/µl sowie 0,025 IU/µl AmpliTaq (Applied Biosystems, ThermoFisher Scientific, Massachusetts, USA). Als Substrat für die DNA-Polymerase dienten Desoxynukleosidtriphosphate (Carl Roth, Karlsruhe, Deutschland) in einer Konzentration von 0,2 mM. Zusätzlich erfolgte die Zugabe eines zehnfach konzentrierten PCR-Puffers (Applied Biosystems, ThermoFisher Scientific, Massachusetts, USA) in der Menge von 0,1 µl pro µl Reaktionsansatz. Für die konventionelle PCR wurden 50 µl Ansätze und für die Gradienten-PCR 20 µl Ansätze verwendet. Zur Auffüllung fehlenden Volumens je Ansatz diente Aqua bidest. Die zugegebene

DNA-Menge lag zwischen 2,2 bis 2,9 ng/ μ l Reaktionsvolumen. In den Negativkontrollen wurde statt DNA zusätzliches Aqua Bidest zugegeben, um etwaige Verunreinigung der Reagenzien auszuschließen (No Template Control, NTC). Die PCR bzw. Gradienten-PCR erfolgte auf dem Mastercycler Gradient (Eppendorf, Hamburg, Deutschland). Für Programme siehe Anhang S.105.

2.4.3 Gelelektrophorese

Die Gelelektrophorese dient dem Auftrennen von Teilchen entsprechend ihrer elektrischen Ladung und Größe. In der vorliegenden Arbeit wurde sie genutzt, um die Spezifität der PCR-Produkte und die optimalen Annealing-Temperaturen im Rahmen der Bruchpunktsequenzierung zu ermitteln, den Spezifitätsnachweis bei den neu etablierten qPCR-Assays zu erbringen, sowie Amplifikate bei der PCR des *SRY*-Gens nachzuweisen. Hierfür fand ein 2% Agarosegel Verwendung, hergestellt aus 2g NEEO Agarose (Carl Roth, Karlsruhe, Deutschland) auf 100ml TBE Puffer (90 mM Tris, 90mM Borsäure, 2,5 mM EDTA). Vom PCR-Produkt wurden 5 μ l vor dem Auftragen aufs Gel mit 1,5 μ l Probenpuffer aus 80 μ l Loading-Puffer III (30% Glycerol in Wasser, 0,25% Bromophenolblau, 0,25% Xylencyanol) [91] und 20 μ l Midori green direct (Nippon Genetics Europe, Düren, Deutschland) gemischt. Als Referenz und zur Abschätzung der Bandengröße wurde eine 100 bp DNA-Leiter (New England BioLabs, Massachusetts, USA) ebenfalls mit dem Probenpuffer gemischt und auf das Gel aufgetragen. Die Auftrennung der Banden erfolgte in einer horizontalen Elektrophoresekammer (Owl B1A EasyCast, ThermoFisher Scientific, Massachusetts, USA) für 40 bis 70 Minuten bei einer Spannung von 9 bis 11 V/cm. Durch Midori-Green angefärbte DNA konnte in der Blaulicht-Kammer (Fusion SL, Vilber Lourmat, Eberhardzell, Deutschland) dargestellt und fotografiert werden, wobei die Bedienung über die Software Fusion Capt Advance (Vilber Lourmat, Eberhardzell, Deutschland) geschah.

2.4.4 DNA-Sequenzierung

Die spezifische DNA-Sequenzierung der Bruchpunkte erfolgte mit der Didesoxymethode nach Sanger, welche auf dem Kettenabbruch mit Fluoreszenzfarbstoff markierten 2'3'Didesoxyribonukleosid-Triphosphaten (ddNTP) basiert.[92]

Vor der eigentlichen Sequenzierungsreaktion wurden 5 μ l PCR-Produkte mit 0,5 μ l Exonuclease I (ThermoFisher Scientific, Massachusetts, USA) und 1 μ l Shrimp Alkalische

Phosphatase (rSAP) (ThermoFisher Scientific, Massachusetts, USA) einem Exo-SAP-Verdau unterzogen. Dies diente der Aufreinigung der Proben, da die Exonuclease I einzelsträngige DNA (verbliebene Primer) abbaut, während die thermolabile rekombinante rSAP verbliebene dNTPs dephosphoryliert. Die Ansätze wurden zur optimalen Enzymaktivität 15 min bei 37°C im Thermocycler (Nexus X2e, Eppendorf, Hamburg, Deutschland) inkubiert und anschließend für 15 min bei 85°C inaktiviert (siehe Anhang, S.105).

Für die Sequenzierungsreaktion wurden 6,5 µl ExoSAP-Verdau, 1 µl Sequenzierungsprimer (20pmol/µl), 1 µl BigDye v3.1 (ThermoFisher Scientific, Massachusetts, USA), 3,5 µl fünffach konzentrierter Big Dye Puffer (ThermoFisher Scientific, Massachusetts, USA) und 8 µl *Aqua Bide* verwendet. Das BigDye enthält neben dNTPs in geringerer Anzahl auch die für die Kettenabbrüche notwendigen ddNTPs, die mit verschiedenen Fluoreszenzfarbstoffen markiert sind. Die Cycle-Sequencing-Reaktion erfolgte auf dem Thermocycler (Nexus Cycler, Eppendorf, Hamburg, Deutschland) und dem Programm HotDye (siehe Anhang, S.105). Im Anschluss an die Sequenzierungs-PCR mussten die Proben über die NucleoSEQ-Säulen (Macherey-Nagel, Düren, Deutschland) gemäß Protokoll nochmals aufgereinigt und so kleine Moleküle, speziell überschüssige, gefärbte Didesoxynucleotide, entfernt werden.

Im Anschluss wurden die aufgereinigten Produkte mit 20 µl Hi-Di Formamid (ThermoFisher Scientific, Massachusetts, USA) versetzt, hitzedenaturiert und die DNA-Fragmente mittels Kapillarelektrophorese im ABI 3130-Genetic-Analyzer (Applied Biosystems, ThermoFisher Scientific, Massachusetts, USA) aufgetrennt. Zur Auswertung der Elektropherogramme konnten die Programme Sequence-Analysis 6 (Applied Biosystems, ThermoFisher Scientific, Massachusetts, USA), Chromas (Technelysium, Brisbane, Australien) und SeqMan (DNASTar, Wisconsin, USA) genutzt werden.

Die durch Sangersequenzierung ermittelten Bruchpunktsequenzen wurden mit Hilfe des Programms SeqMan (DNASTar, Wisconsin, USA) mit den zuvor aus WGS-Daten abgeleiteten Bruchpunktsequenzen abgeglichen.

2.5 Transkriptom-Sequenzierung

2.5.1 Prinzip

Die RNA-Sequenzierung wurde aus zehn 10 µm Schnitten von ‚Formalin-fixierten Paraffin-eingebettetem‘ (FFPE) Gonadengewebe der Patient:innen 1 und 2 durchgeführt. Die Auftragssequenzierung (Novogene, Hong Kong, China) erfolgte auf der NovaSeq Plattform (Illumina, Kalifornien, USA), während die Erstellung der Sequenzierungsbibliothek mit dem TruSeq stranded total RNA Kit (Illumina, Kalifornien, USA) geschah. Wie bei dem WGS handelt es sich um eine Sequenzierung mit 150 bp Read-Paaren. Insofern soll an dieser Stelle auf die Schritte 1 bis 4 der NGS-Methode nicht nochmals eingegangen werden (siehe S.17ff.), sondern nur auf den davon abweichenden und in Lübeck durchgeführten Teil der Datenauswertung. Die Auswertung der Transkriptomsequenzierung mit dem unten beschriebenen t-SNE und der Gene Set Variation (GSV)-Analyse konnten dank einer Kooperation mit der Gruppe für Systembiologie der Universität zu Lübeck (Prof. Hauke Busch) und einer Beteiligung an dem Lübeck OMICS-Rechencluster durchgeführt werden.

Die Sequenzierung generierte 98 Millionen Reads bei der Probe von P1 und 63 Millionen Reads bei P2. Die Ausrichtung der Reads an die Ensembl cDNA sequence GRCh38 erfolgte mit kallisto [93]. Zur Analyse wurden die Reads der Transkripte den Genen zugeordnet, denen sie entstammen [94]. Damit ein Gen Eingang in die weitere Analyse finden konnte, musste es bei beiden Patientinnen mit über 1 Transkript pro Millionen Transkripte (TPM) vorliegen. Die Daten wurden verglichen mit Transkriptomdaten der GTEx Datenbank [95] und ArrayExpress [96].

2.5.2 t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-Distributed Stochastic Neighbor Embedding ist ein Algorithmus für maschinelles Lernen, der zur Visualisierung hochdimensionaler Daten z.B. in einem zweidimensionalen Streudiagramm dient. Hierbei werden für die Datenpunkte in dem hochdimensionalen Raum anhand ihrer gegenseitigen Entfernungen zuerst Ähnlichkeitswerte berechnet. Diese Ähnlichkeitswerte stellen die Wahrscheinlichkeit dar, dass ein Punkt (x_a) im Algorithmus einen anderen Punkt (x_b) als Nachbarn zugeordnet wird, wenn die Auswahl abhängig von der Wahrscheinlichkeitsdichte einer Gaußkurve mit dem Zentrum am x_a geschieht. Nahegelegene und ähnliche Punkte haben somit einen hohen Wahrscheinlichkeitswert, während weit entfernte einen zu

vernachlässigenden, kleinen Wahrscheinlichkeitswert aufweisen. Diese paarweisen Wahrscheinlichkeiten bilden eine Matrix. In einem nächsten Schritt werden die Punkte zufällig in einem zweidimensionalen Raum verteilt und abermals Ähnlichkeitswerte ausgerechnet. Im Gegensatz zum klassischen SNE [97] wird bei der t-SNE in diesem Schritt statt einer Gaußverteilung eine Studentische t-Verteilung mit schweren Rändern verwendet [98]. Der Algorithmus versucht dann die Abweichung zwischen den Ähnlichkeitswerten des hochdimensionalen und des zweidimensionalen Raums auf ein Minimum zu reduzieren; berechnet mit der Mullback-Leibler-Divergenz. Diese Umwandlung wurde mit insgesamt 641 Transkriptomdaten, inklusive P1 und P2, anhand von 5000 zufällig ausgewählten Genen mit TPM>1 durchgeführt.

2.5.3 Gene Set Variation Analysis (GSVA)

Die GSV-Analyse [99] ermöglicht es, in der Auswertung statt einzelnen Genen ganze Gengruppen und damit Signalwege zu betrachten. Ein einzelnes Gen kann in einer Probe beispielsweise durch Umwelteinflüsse momentanen Schwankungen unterliegen. Für eine robustere Bewertung der Zellaktivität und einen besseren Vergleich mit anderen Proben ist das Betrachten mehrere Gene eines Signalweges vorteilhaft. Die eingesetzten Daten umfassen eine Matrix der Genexpressionen in Form der zugeordneten RNA-Seq-Readzahl ausgewählter Gene einer Probe, sowie eine Liste der Gengruppen (Gene Sets). Als Gengruppen wurde die ‚hallmark gene set collection‘ [100] genutzt. Dabei sind methodisch die Genexpressionen jeder Probe der Größe nach zu sortieren und den Genen entsprechende Rangwerte zuzuteilen, die symmetrisch um Null verteilt sind. Den Rang zu nutzen statt der Zahl der sequenzierten Reads, bringt bei den weiteren Rechnungen die Proben auf eine vergleichbarere Größe. So findet ein Ausgleich zwischen den Datensätzen mit methodisch bedingten Schwankungen der Readzahl aus teils verschiedenen Experimenten statt. Hierbei verleiht das Normalisieren um Null besonders hoch und niedrig exprimierten Genen mehr Gewicht. Um die Aktivität eines Signalwegs zu beurteilen, wird die Rangliste in zwei Gruppen unterteilt (Teil der Gengruppe oder nicht) und mit dem Kolmogorov-Smirnov-Test überprüft, ob diese gleichmäßig verteilt sind. Daraus leitet sich dann die relative Anreicherung (Enrichment Score) einer Gengruppe gegenüber der Expression der Gene außerhalb der Gengruppe ab. Die Berechnung erfolgte für 26 Signalwege/Gengruppen. Die Expression der Patient:innen 1 und 2 wurde mit 6 zufällig ausgewählten ovariellen und testikulären Transkriptomdaten verglichen sowie dem Mittelwert der Transkriptionsdaten der beiden jeweiligen Gewebe der GTEx Datenbank.

2.6 qPCR-Screening auf Xp21.2 Kopiezahlvariationen

Die quantitative Echtzeit-PCR (qPCR) ermöglicht es, gezielt die Kopiezahl bestimmter kleiner DNA-Bereiche zu ermitteln. Im Gegensatz zu einer herkömmlichen PCR wird, während die PCR-Zyklen durchlaufen werden, die zunehmende Amplifikation durch ein steigendes Fluoreszenzsignal erfasst. Hierfür stehen verschiedene chemische Methoden zur Verfügung, unter anderem DNA-Farbstoffe wie SYBR Green oder Förster-Resonanzenergietransfer (FRET)-Sonden. Die Anzahl der nötigen Zyklen, bis die Fluoreszenz das Niveau des Hintergrundsignals übersteigt, ist abhängig von der initial eingesetzten DNA-Menge und wird mit dem C_q-Wert (cycle quantification value) quantifiziert. Je mehr DNA von einem Bereich vorliegt, desto früher wird dieser Schwellenwert überschritten und desto niedriger ist der C_q-Wert. Referenzgene dienen dazu, eventuelle Abweichungen der eingesetzten DNA-Menge auszugleichen. Damit entscheidet in der Auswertung nicht die absolut eingesetzte DNA-Menge, sondern die relative Kopiezahl im Vergleich zum Referenzgen.

2.6.1 Konzipierung der qPCR-Assays

Als Vorlage zur Generierung der Primer sind die NCBI-Gene-Sequenzen (<https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve>) (zuletzt abgerufen 11.05.2021) der Gene *IL1RAPL1*, *MAGEB1*, *NROB1*, *CXorf21*, *GK* und *TAB3* als GenBank-Dateien gespeichert und verwendet worden. Um mit den Assays zukünftig sowohl ein SYBR-Green-System, als auch die FRET-Sonden der ‚Roche Universal ProbeLibrary‘ (Hoffmann-La Roche, Basel, Schweiz) nutzen zu können, wurde zum Design das ‚Universal ProbeLibrary Assay Design Center‘ (Hoffmann-La Roche, Basel, Schweiz) (https://lifescience.roche.com/en_de/brands/universal-probe-library.html#assay-design-center) (zuletzt abgerufen 11.05.2021) genutzt. Die Verwendung dieses Design-Centers sorgte auch dafür, dass alle Primerpaare im Entwicklungsprozess ähnliche Annealing-Temperaturen und Amplikongrößen haben. Alle Primer wurden in den Exonen der sechs obenstehenden Gene berechnet, um Repeat-Elemente in nichtkodierenden Bereichen zu umgehen. Im Anschluss erfolgte die nochmalige Prüfung der Primersequenzen mit einer BLAT-Suche auf ihre Spezifität. Um häufige SNVs in dem Bindungsbereich der Primer und der FRET-Sonden auszuschließen, kam SNPCheck3 (<https://genetools.org/SNPCheck/snp-check.htm>) (zuletzt abgerufen 11.05.2021) zum Einsatz.

Als Referenz können bei dieser Methode die meisten autosomal vererbten Gene mit essenzieller Funktion, die keine Assoziation mit dem Phänotyp aufweisen, verwendet werden.[101] Für diese Arbeit konnte auf bereits etablierte und in der Datenbank RT-PrimerDB [102] zur Verfügung gestellte SYBR-Assays in *ZNF80* (ID 1021) und *GPR15* (ID 1022) zurückgegriffen werden. Als Bezugsquelle für die Primer fungierte die metabion international AG (Planegg, Deutschland).

2.6.2 Etablierung der qPCR-Assays

Für die qPCRs wurden pro Ansatz 10 µl zweifach verdünnter Takyon No Rox SYBR MasterMix dTTP Blue (Eurogentec, Seraing, Belgien) und 1 µl je Primer (2 µM) und 8 µl DNA (bzw. *Aqua bidest* für die Negativkontrollen) hinzugegeben für ein Gesamtvolumen von 20 µl. Die Reaktion lief auf einem LightCycler 96 (Hoffmann-La Roche, Basel, Schweiz) mit Hilfe des im Anhang auf S.105 beschriebenen Programms ab.

Um die Qualität der Assays am Xp21.2-Locus und der Referenzgene zu überprüfen, mussten anhand von Verdünnungsreihen Standardkurven im Bereich der später untersuchten DNA-Menge erstellt und aus diesen die Amplifikations-Effizienz berechnet werden. Die Verdünnungsreihen bestanden aus vier Punkten, die jeweils in Triplikaten gemessen wurden. Die maximale DNA-Menge in den Ansätzen der Standardkurve betrug 20 ng, die in vier Schritten auf 0,3125 ng zu reduzieren war.

Die lineare Standardkurve ergibt sich dann aus dem C_q-Wert gegen den Logarithmus der eingesetzten DNA-Konzentration. Die Steigung der Geraden durch die Datenpunkte entspricht hierbei der Effizienz. In jedem PCR-Zyklus sollte sich die Zielsequenz verdoppeln und insofern im Idealfall eine Effizienz von 2 vorliegen.

Damit also ein Assay für das Screening in die engere Wahl kam, musste die Effizienz $2,00 \pm 0,10$ entsprechen. Die Effizienz wurde darüber hinaus bei den späteren Screening-Läufen zu einer genaueren Berechnung der Ergebnisse miteinbezogen. Eine Aufstellung der ausgewählten Assays und jeweiligen Effizienz ist ebenfalls der Tabelle auf S.104 im Anhang zu entnehmen. Zum Nachweis der Spezifität der Assays bei Verwendung mit SYBR-Green waren

Schmelzkurvenanalysen durchzuführen und das Vorliegen eines einzelnen Reaktionsproduktes durch Gelelektrophorese zu überprüfen.

2.6.3 Validierung der qPCR-Assays

Assays für die Untersuchungen konnten mittels Proben der Patient:in 1 und 2 validiert werden, bei denen bereits bekannte Duplikationen am Xp21.2-Locus vorlagen. Diese Duplikationen umspannten teilweise die von den Assays abgedeckten Bereiche, teilweise nicht. Die Ergebnisse der qPCR waren dabei ausnahmslos im Einklang mit der erwarteten Kopiezahl an der entsprechenden Assay-Position. So zeigten beispielsweise *CXorf21* und *GK* jeweils eine doppelt so hohe Kopiezahl wie *NROB1*. Als Referenz wurde zusätzlich eine männliche und eine weibliche Kontrollprobe mituntersucht, die durchgehend ein einfaches beziehungsweise doppeltes Kopiezahlniveau zeigten entsprechend ihres einfach oder doppelt vorliegenden X-Chromosoms.

2.6.4 Auswertung

Das qPCR-Screening wurde in zwei Stufen durchgeführt und jede Patient:innenprobe initial mit mindestens vier Assays untersucht. *NROB1* als dosisabhängiges Kandidatengen für 46,XY-GD, *CXorf21* um Kopiezahlvariationen stromaufwärts von *NROB1* erfassen zu können und *ZNF80* und *GPR15* als Referenzgene. Es sind immer zwei Referenzgene gemessen worden, falls bei einer der Patient:innen zufällig eine Kopiezahlvariation in einem der beiden Referenzgene vorliegen sollte. Bei auffälligen Ergebnissen wurden die entsprechenden Patient:innenproben nochmals mit diesen Assays, sowie den weiteren in der Umgebung gelegenen Assays untersucht, um das Vorliegen zu bestätigen und eine erste Orientierung bezüglich der Größe des CNV zu erhalten.

2.6.5 Bestimmung der DNA-Konzentration

Um die für die qPCR nötigen Verdünnungen herzustellen, musste die DNA-Konzentration der isolierten Patient:innen-DNA ermittelt werden. Dies geschah photometrisch mit dem Bio-Photometer Plus (Eppendorf, Hamburg, Deutschland) oder mit dem Qubit Fluorometer (ThermoFisher Scientific, Massachusetts, USA) und Qubit dsDNA BR Assay Kit (ThermoFisher Scientific, Massachusetts, USA) gemäß den Herstellerangaben. Letztere Methode hat den Vorteil, dass diese selektiv für doppelsträngige DNA ist und freie Nukleotide, RNA sowie andere

chemische Kontaminationen das Ergebnis nicht verfälschen. Qubit-Messungen sind auch zur Vorbereitung der Proben für das WGS genutzt worden.

2.6.6 Aktenrecherche

Die Sektion für Pädiatrische Endokrinologie und Diabetologie an der Universität zu Lübeck, in der die vorliegende Arbeit angefertigt wurde, blickt auf 25 Jahre Erfahrung in der Forschung und Versorgung im Bereich der Varianten der Geschlechtsentwicklung zurück. Bei der Behandlung von Patient:innen und/oder der Einsendung von DNA-Proben, wird standardmäßig erfragt, ob die Überlassung des Materials zu wissenschaftlichen Zwecken erlaubt ist. Sofern die Patient:innen und/oder Erziehungsberechtigten einer wissenschaftlichen Nutzung zustimmen, werden Patient:innen-Akten zur Forschung archiviert und Patient:innen-DNA pseudonymisiert aufbewahrt. Die Archivierung und wissenschaftliche Nutzung der Patient:innen-Akten und -DNA erfolgt mit Zustimmung der Ethikkommissionen der Universität zu Lübeck und der Christian-Albrechts-Universität zu Kiel (siehe Anhang S. 92)

Zum Zeitpunkt der Nachforschung umfasste der Patient:innenstamm über 2 400 Fälle. Bei allen wurde geprüft, ob diese möglicherweise eine bis dato unerkannte Xp21.2 Kopiezahlvariation als Ursache ihrer 46,XY-DSD haben könnten. Einschlusskriterien für die Auswahl zum anschließenden qPCR-Screenings waren ein 46,XY-Chromosomensatz und weiblicher Phänotyp mit bisher ungeklärter molekulargenetischer Ursache. Auch wenn eine Hypospadie und/oder Mikropenis ein Zeichen einer partiellen Gonadendysgenese sein kann [103], wurden aufgrund der hohen Zahl an Hypospadiepatienten diese vorerst nicht in das Screening mit einbezogen. Auch Patient:innen mit einer unauffälligen MLPA im Bereich von *NROB1* und *CXorf21* waren nicht eingeschlossen. Patient:innen ohne bekannten Karyotyp in den Akten wurden dann ins Screening aufgenommen, wenn diese die Diagnosen Swyer-Syndrom; 17 β -Hydroxysteroid-Dehydrogenase-Mangel ohne *17 β HSD3*-Mutation oder komplette Androgenresistenz ohne nachweisbare Mutation im *AR*-Gen gestellt bekommen hatten.

2.7 Konformationserfassung von Chromosomen durch Hi-C

In den letzten Jahren sind verschiedene Methoden zur Analyse der räumlichen Organisation des im Zellkern befindlichen Chromatins aufgekommen. Die Hi-C-Methode [104] verwendet dabei Hochdurchsatzsequenzierung, um Interaktionspunkte weit entfernter Chromatinbereiche des gesamten Genoms zu erfassen. Vorgängermethoden wie die ‚Chromosome Conformation Capture (3C)‘-Methode [105] misst im Gegensatz dazu lediglich die Interaktionsfrequenz zwischen zwei bestimmten Bereichen.

Hi-C erfolgt standardmäßig in vier Schritten. Zuerst werden die zum entsprechenden Zeitpunkt in dem Gewebe physiologischer Weise vorliegenden Quervernetzungen des Chromatins mit Formaldehyd fixiert, bevor anschließend ein Restriktionsverdau durchgeführt wird, der das Chromatin fragmentiert. Als Drittes werden die Proben verdünnt und dann mit Ligasen versetzt. Durch die Verdünnung werden Ligationen der DNA-Enden innerhalb der über Formaldehyd fixierten Moleküle gegenüber intermolekularen Ligationen bevorzugt. In der Folge entstehen chimäre Sequenzen, die dadurch gekennzeichnet sind, dass sie eventuell auf dem Chromosom weit entfernt liegen, jedoch durch die physiologischen Chromatinfaltung räumlich nah beieinander gelegene Bereiche darstellen. Viertens werden durch die Hochdurchsatzsequenzierung die chimären Sequenzen erfasst und daraus bioinformatisch die Kontaktpunkte berechnet.

Auf dem Campus der Universität zu Lübeck war zum Zeitpunkt der Doktorarbeit die Methode nicht verfügbar und konnte in Ermangelung entsprechender Geräte und Gewebeproben nicht etabliert werden. Daher wurde auf öffentlich verfügbare Hi-C Daten des ‚Penn State 3D Genome Browser‘ zurückgegriffen.[106]

3. Ergebnisse

3.1 Whole Genome Sequencing (WGS)

3.1.1 Patient:in 1

Am Probenmaterial der P1 wurde ein WGS durchgeführt, um die zuvor anhand von Multiplex ligation-dependent probe amplification (MLPA) und Chromosomal microarray (CMA) entdeckte Duplikation zu bestätigen, die genauen Bruchpunkte zu ermitteln und damit zu überprüfen, ob diese tatsächlich das *NROB1* Gen ausspart. Die Analyse der WGS-Daten mit dem Programm Control-FREEC [84] ermittelte daraus 398 Kopiezahlvariationen (CNVs) wovon 10 auf dem X-Chromosom lagen, keine jedoch am erwarteten Xp21.2-Locus. Mittels DELLY [85] konnten 4 761 strukturelle Variationen (SVs) identifiziert werden, wovon 134 auf dem X-Chromosom verortet waren und eine davon auf Xp21.2. Hierbei handelte es sich um eine als Duplikation eingeordnete SV im Bereich ChrX: 30,561,644-30,859,144. Dies entsprach in etwa dem CMA-Ergebnis einer Duplikation im Bereich von ChrX: 30,580,693-30,857,187.

Die Darstellung der Sequenzdaten zwischen den beiden von DELLY berechneten Bruchpunkten im Integrative Genomics Viewer (IGV) [86] zeigte eine deutlich erhöhte Sequenzierungstiefe (Abb I, Anhang S.92). Jeder Base in diesem Bereich waren ungefähr doppelt so viele Sequenzierungsreads zugeordnet wie in den angrenzenden Bereichen des X-Chromosoms. Genau an der Chromosomenposition ChrX:30,561,644 ließ sich ein abrupter Abfall der Sequenzierungstiefe sowie ein relevanter Anteil an chimären Reads feststellen. Bei allen chimären Reads wurde ein Teil der Sequenz dem Bereich ab ChrX:30,561,644 in telomerischer Richtung zugeordnet und der andere Teil an einem ca. 297 200 Basen entfernten Chromosomenbereich verortet. Auch die entgegengesetzten Partner-Reads der bei ChrX:30,561,644 angezeigten chimären Reads waren nicht in der Nähe lokalisiert. Die Partner-Reads wurden im Bereich ab ChrX:30,859,140 in zentromerischer Richtung kartiert. An dieser Basenposition bestand ebenfalls ein Abfall der Sequenzierungstiefe, der somit den gegenüberliegenden Bruchpunkt der Duplikation darstellt. Dieser Bruchpunkt weicht leicht von dem berechneten Ergebnis der DELLY-Analyse (ChrX:30,859,144) ab. Beide Beobachtungen, die der erhöhten Sequenzierungstiefe sowie die an den Abbruchpunkten der Sequenzierungstiefe befindlichen chimären Reads, bestätigen das Vorliegen einer Duplikation in diesem Bereich.

Um andere bekannte molekulargenetische Ursachen als Auslöser der 46,XY-GD in P1 auszuschließen, wurden SNVs und InDels mit Populationshäufigkeit < 1% auf Mutationen in DSD-Kandidatengenen überprüft. Dabei wurden zwei heterozygote missense SNVs mit mutmaßlich schädlicher Auswirkung auf die Proteinstruktur der mit 46,XY-DSD assoziierten Gene von *Steroidogenic Acute Regulatory Protein (STAR)* und *Cytochrome P450 Oxidoreductase (POR)* entdeckt (Tab. 1). Heterozygote Mutationen allein reichen in diesen Genen jedoch vermutlich nicht aus, um den in P1 beschriebenen Phänotyp zu verursachen. Weitere Kopiezahlvariationen in dosissensitiven DSD-Genen und deren 2 Mb-Umgebung konnten nicht identifiziert werden.

Tab. 1. Heterozygote missense SNVs von Patient:in 1 in bekannten DSD-Genen

Gene	dbSNP ID	ExAC All [52]	ExAC European Non-Finish [52]	SIFT [81]	Polyphen2 [82]	MutationTaster [83]
STAR	rs34908868	0.00167	0.00290	D (0.009)	D (1.0)	D (0.996)
POR	rs145782750	0.00159	0.00157	T (0.904)	B (0.156)	D (1.0)

Allelhäufigkeit in einem Gesamtkollektiv (ExAC All) und einem Subkollektiv (ExAC European Non-Finish). Der Zugriff auf die ExAC-Datenbank erfolgte im März 2020. Berechnungen zur Auswirkung auf die Proteinstruktur erfolgten durch SIFT, Polyphen und MutationTaster. D = Schädlich („Deleterious“), T = Verträglich („Tolerated“); B = Benigne („Benign“)

Die Bestätigung der Duplikation 234kb stromaufwärts eines bekannten dosisabhängigen Gens der Geschlechtsentwicklung, *NROB1*, sowie der Ausschluss anderer bekannter Ursachen legt die Pathogenität dieser neu entdeckten Mutation nahe. Eine molekulargenetische Analyse der Mutter von Patient:in 1 zeigte, dass diese ebenfalls Trägerin der 234kb Duplikation ist.

3.1.2 WGS-Analyse weiterer Patient:innen

Um der Frage nachzugehen, ob ähnliche Duplikationen bei anderen Patient:innen mit 46,XY-GD und ungeklärter Ätiologie vorliegen, wurden die WGS-Daten von drei weiteren Patient:innen mit Gonadendysgenese analysiert, woraus die in dieser Arbeit als P2 bezeichnete Patient:in 2 hervorging. Die beiden übrigen Patient:innen wurden aufgrund der Untersuchungsergebnisse nicht in diese Studie eingeschlossen.

Denn: Die Auswertung mit Control-FREEC berechnete 494 (8 auf X-Chromosom), 401 (6) und 497 (2) Kopiezahlvariationen für die drei Patient:innen. Keine davon befand sich am Xp21.2-Locus (ChrX: 29,300,000-31,500,000 [107]). DELLY bestimmte 3 707 (113), 3 633 (105) und 3 677 (124) SVs. Bei den Fällen mit 3 633 (105) und 3 677 (124) DELLY-Ergebnissen fanden sich hierbei Auffälligkeiten im Xp21.2 Bereich. Im Fall mit 3 633 (105) erkannte DELLY eine Duplikation, die sich im IGV jedoch nicht bestätigen ließ. Bei der Patient:in mit 3 677 (124) Treffern (fortan Patient:in 2 bzw. P2 genannt) fanden sich in dem entsprechenden Bereich sechs getrennte SVs, die sich aufteilten auf zwei Inversionen, zwei Deletionen und zwei Duplikationen. Die von DELLY bestimmten Bruchpunkte wurden ebenfalls im IGV visualisiert. Es ließen sich Anomalien in den Bereichen der Bruchpunkte der beiden als Inversionen und einer der beiden als Deletionen eingestuft SVs identifizieren. Inversionen können hingegen im IGV nicht direkt erkannt werden. Stattdessen konnten durch die ungewöhnlich hohe Sequenzierungstiefe und die hervorgehobenen chimären Reads zwischen den von DELLY ausgegebenen Koordinaten zwei Duplikationen und zwei Deletionen detektiert werden (für Positionen siehe Abb. II, Anhang S. 93). Eine der Duplikationen entstammt dabei einem Bereich telomerisch von *NROB1*. Die anderen drei CNVs betreffen Sequenzen, die zentromerisch von *NROB1* gelegen sind. *NROB1* selbst ist jedoch nicht beteiligt.

Bei den Bruchpunkten der anderen von DELLY identifizierten Deletion und den beiden Duplikationen handelte es sich allerdings um Fehleinstufungen des Programmes. Denn alle drei liegen nah beieinander in einem Bereich zwischen *GK* und *TAB3*, einem Bereich, der hauptsächlich aus Simple repeats besteht (ChrX:30,807,585-30,814,379). Im IGV werden die Reads in diesem Bereich mit ungewöhnlich hohen und unterschiedlichen Abweichungen der Read-sequenzen von der Referenzsequenz angezeigt. Daher handelt es sich höchstwahrscheinlich um Sequenzierungsfehler und Fehlkartierung der Reads, die durch repetitive Motive verursacht worden sind. Diese Sequenzierungsfehler und Fehlkartierungen werden

dann vom Programm fälschlich als SV eingestuft. Die von DELLY fälschlich erkannte Duplikation bei der anderen Patient:in aus der WGS-Daten-Screeninggruppe befand sich ebenfalls in diesem Bereich.

Die Auswertung der Exom-Daten der als P2 erfassten Patient:in lieferte eine heterozygote missense Mutation im *Zinc Finger Protein, Multitype 2 (ZFPM2)*, auch bekannt als *Friend of Gata Family Member 2 (FOG2)*. Eine heterozygote Mutationen in diesem Gen konnte bereits mit 46,XY-GD assoziiert werden.[108] Dabei handelt es sich jedoch um eine andere Mutation als die in dieser Arbeit bei P2 vorliegende Mutation. Ferner ist die Einschätzung der Auswirkung auf die Proteinfunktion mit SIFT, PolyPhen und MutationTaster uneinheitlich. Auch scheint die Allelfrequenz mit über 0.4% in der europäischen Bevölkerung als Auslöser für eine so seltenen auftretende Erkrankung sehr hoch. Deswegen muss davon ausgegangen werden, dass es sich hierbei nicht um die eigentliche Ursache der Gonadendysgenese in P2 handelt.

Tab. 2. Heterozygote missense SNV von Patient:in 2 in bekanntem DSD-Gen *ZFPM2*.

Gene	dbSNP ID	ExAC All [52]	ExAC Euro-pean Non-Fin-ish [52]	SIFT [81]	Polyphen2 [82]	MutationTaster [83]
ZFPM2	rs202217256	0.00262	0.00454	T (0.422)	P (0.914)	D (1.0)

T= Verträglich („Tolerated“), P = Möglicherweise Schädlich („Possibly Deleterious“), D = Schädlich („Deleterious“). Der Zugriff auf die ExAC-Datenbank erfolgte im März 2020.

3.2 Bestätigungen der Kopiezahlvariationen

Die WGS-Daten von P2 zeigten Hinweise auf das Vorliegen von Kopiezahlvariationen in der Umgebung von *NROB1*. Doch die bioinformatischen Methoden und Auswertung im IGV lieferten keine übereinstimmenden Ergebnisse zu deren Charakter. Deshalb sollten die Kopiezahlvariationen mit einer unabhängigen Methode bestätigt werden. Hierfür wurde eine aCGH-Untersuchung durchgeführt, wobei sich zwei Duplikationen und eine Deletion nachweisen ließen. Die beiden Duplikationen waren stromauf- und stromabwärts von *NROB1* lokalisiert. Die Duplikation stromabwärts von *NROB1* wurde durch die verstärkte Hybridisierung an 17 Sonden nachgewiesen (Abb. 7) und so auf eine minimale Größe von 389kb (ChrX:29,924,420-30,313,761) bestimmt und umfasste damit die Gene *MAGEB1-4*, sowie einen Teil von *IL1RAPL1*. Die Duplikation stromaufwärts von *NROB1* umfasste 23 aCGH-Sonden und wurde auf eine Mindestgröße von 447kb (ChrX: 30,401,819-30,848,988) bestimmt. Diese Duplikation umfasste die Gene *CXorf21*, *GK* und einen Teil von *TAB3*. Bis auf kleine Abweichungen in der Duplikationsausdehnung, bedingt durch die geringere Auflösung des Arrays, deckten sich die aCGH-Ergebnisse mit den WGS-Ergebnissen.

In dem Bereich zwischen den beiden Duplikationen befanden sich im aCGH vier Sonden ohne erhöhtes Hybridisierungsniveau. Davon zeigte eine Sonde (Abb. 7A, Sonde 1) keine Hybridisierung. Die Chromosomenposition dieser Sonde stimmte mit einer der im IGV erkennbaren Deletionen überein und bestätigte diese somit. Im Bereich der anderen im IGV erkennbaren Deletion befand sich keine aCGH-Sonde, so dass diese nicht einwandfrei gesichert werden konnte. Aufgrund der hohen Übereinstimmung der aCGH-Ergebnisse mit den anderen im IGV erkennbaren Kopiezahlveränderungen bei P1 und P2 ist aber davon auszugehen, dass diese Deletion ebenfalls vorliegt. Ferner ist das abrupte Fehlen von WGS-Reads (Abb. 7B) nur schwer mit anderen Phänomenen zu erklären.

Somit lässt sich zusammenfassen, dass bei P2 zwei Duplikationen und zwei Deletionen am Xp21.2 Locus vorliegen, die jedoch *NROB1* aussparen. Auffällig ist, dass die eine Duplikation eine große Überlappung (287kb) mit der ebenfalls stromaufwärts von *NROB1* gelegenen Duplikation von P1 aufweist. (Abb. II, Anhang S. 93)

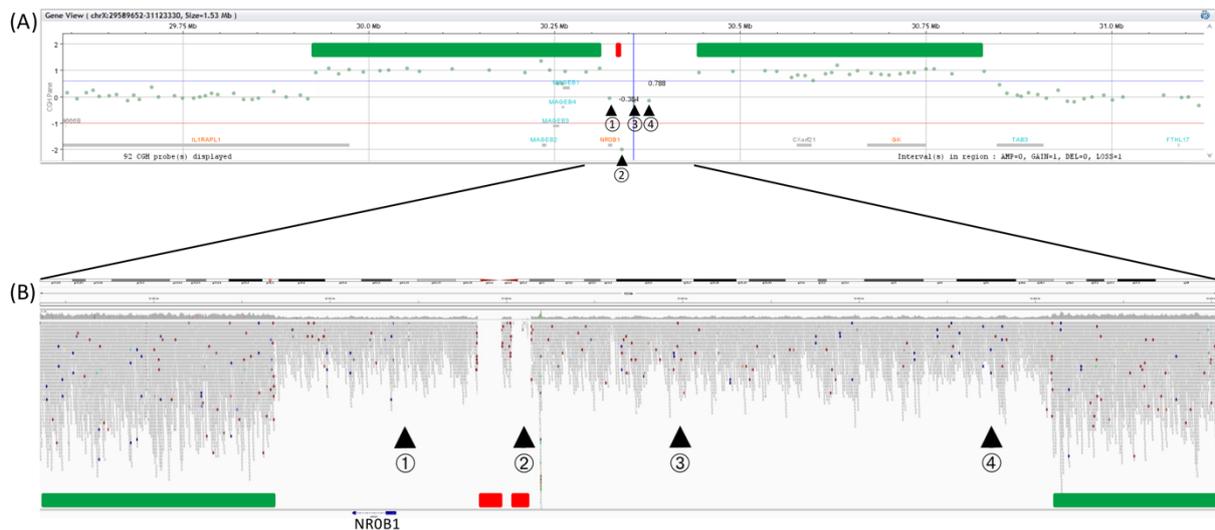


Abb. 7: Kopiezahlanalyse von P2 mittels aCGH und WGS

(A) aCGH-Ergebnis bestätigt das Vorliegen von zwei großen Duplikationen und einer Deletion. Die Duplikationen sind gekennzeichnet durch das höhere Hybridisierungsverhältnis der Patient:innen-DNA an den Sonden (grüne Punkte) und durch grüne Balken hervorgehoben. Die vier nummerierten Sonden zwischen den Duplikationen verdeutlichen, dass es sich um zwei getrennte Duplikationen handelt. Die Sonden 1, 3 und 4 sind im nicht-duplizierten Zwischenbereich, in dem auch *NR0B1* liegt. Sonde 2 ist nicht hybridisiert und bestätigt eine Deletion an diesem Ort (roter Balken). (B) Im IGV dargestellte WGS-Daten im Chromosomenabschnitt ChrX:30,292,562-30,423,695. Der dargestellte Ausschnitt wird jeweils von den duplizierten Segmenten flankiert (grüne Balken), welcher erkennbar ist durch die erhöhte Anzahl an Sequenzierungs-Reads (graue Streifen). Die zwei Deletionen (rote Balken) sind erkennbar durch das Fehlen von Sequenzierungs-Reads. Zum Vergleich sind die Lokalisationen der vier zwischen den Duplikationen liegenden aCGH-Sonden dargestellt.

3.3 Bruchpunktbestimmung

Die Untersuchung der Kopiezahlvariationen mittels CMA & MLPA in P1 und aCGH in P2 lassen klar auf das Vorliegen von Duplikationen am Ort Xp21.2 schießen. Diese Methoden liefern jedoch keine Informationen zur genauen Position, an der die duplizierten Segmente eingebaut wurden. Auch die bioinformatische Auswertung mittels DELLY und Control-FREEC liefert diesbezüglich keine Informationen. Das Gleiche gilt für die Darstellung der WGS-Daten im IGV, welche sich streng an der Referenzsequenz orientiert.

Es stellte sich demnach die Frage, an welchen Orten im Genom die Duplikationen genau eingebaut wurden und ob beispielsweise die von DELLY prognostizierten Inversionen bei P2 tatsächlich vorlagen. Dafür wurden die einzelnen WGS-Reads und ihren Partner-Reads an den bereits zuvor bestimmten zwei bzw. acht Bruchpunkten von P1 und P2 extrahiert und durch Neuordnung der Reads der Aufbau der strukturellen Variationen abgeleitet. Die so

ermittelten Sequenzübergänge an den Bruchpunkten wurden danach mittels Sanger-Sequenzierung bestätigt.

Bei P1 konnte dabei eine Tandemduplikation ermittelt werden (Abb.8). Der duplizierte Bereich (chrX:30,561,644-30,859,140) und das Duplikat befinden sich in der DNA-Sequenz der Patient:in 1 unmittelbar hintereinander gereiht. Getrennt werden Original und Duplikat lediglich durch eine 27bp große Insertion.

Bei P2 ist die Anordnung durch das Vorliegen von vier verschiedenen Kopiezahlvariationen komplizierter. Beide duplizierten Segmente sind stromaufwärts von *NROB1* inseriert (Abb. 9). Entgegen der Prognose von DELLY ist jedoch nur eines der Segmente invertiert, wie sich den neu ausgerichteten WGS-Reads und der Sanger-Sequenzierung entnehmen lässt.

Die Duplikation (ChrX:30,401,819-30,848,988) von *CXorf21*, *GK* und von einem Teil von *TAB3* ist 9,2kb stromaufwärts von *NROB1* invertiert eingesetzt (Duplikation I). Der Punkt, an dem die Referenzsequenz in die Duplikation I übergeht (ChrX: chrX:30,336,745), stellt auch den Beginn der Deletion I dar.

Duplikation I und II werden durch einen 1,2kb invertierten Sequenzbereich getrennt, der der Referenzsequenz zwischen den beiden Deletionen (ChrX:30,339,452-30,340,605) entspricht. Die Duplikation II (ChrX:29,924,420-30,313,761) von *MAGEB1-4* und von einem Teil von *IL1RAPL1* entspricht in der Ausrichtung hingegen der jener Gene der Referenzsequenz. Das zentromerische Ende dieses duplizierten Segmentes geht wieder in die Referenzsequenz über und stellt das Ende der zweiten Deletion dar (ChrX:30,342,785). Durch WGS der Mutter-DNA von P2 zeigte sich, dass sie die gleiche SV trägt. Es handelt sich bei P2 demnach nicht um eine de novo Mutation.

Das Aufdecken der Sequenzanordnung in P1 und P2 zeigt, dass die beiden Patient:innen nicht nur eine große Überlappung eines duplizierten Bereichs mit den Genen *CXorf21* und *GK* aufweisen, sondern auch bei beiden Patient:innen die Duplikationen in unmittelbarer Nähe und stromaufwärts von *NROB1* inseriert sind. Bei P2 zeigt sich dies jedoch in invertierter Richtung. Es ließ sich außerdem demonstrieren, dass sich mit unserem Vorgehen aus WGS-Rohdaten erfolgreich und zuverlässig die Anordnung von strukturellen Variationen ableiten lässt.

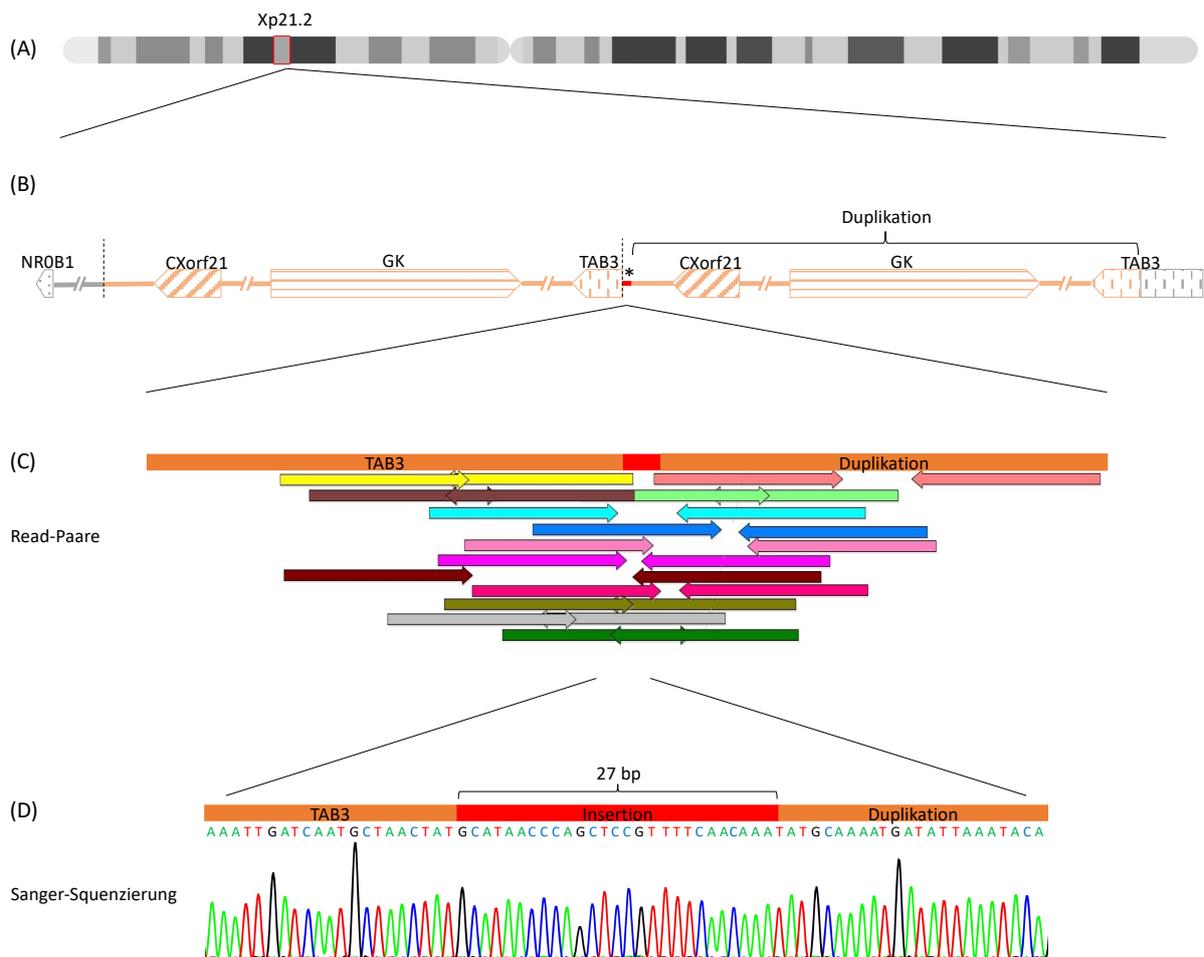


Abb. 8: Ermittlung der Strukturellen Variation in P1

(A) Orientierende Abbildung des X-Chromosoms. Die rote Markierung entspricht dem Xp21.2-Lokus. (B) Schematische Darstellung der Sequenzanordnung in P1 stromaufwärts von *NROB1*. Rechtsseitig befindet sich das Zentromer, linksseitig das Telomer. Die Gene sind durch gerichtete Pfeile symbolisiert, welche der Leserichtung der Gene entsprechen. Die Abstände zwischen den Genen sind nicht maßstabsgetreu. Original und Duplikat sind in Tandem arrangiert und orange dargestellt; nicht beteiligte Sequenzbereiche sind grau. Mit Sternchen * gekennzeichnet ist die 27bp Insertion (rot), welche Original und Duplikat trennt. (B) Relative Position extrahierter und neu ausgerichteter Partner-Reads, die den Bruchpunkt ChrX:30,859,140 überspannen. Zusammengehörige Partner-Reads wurden mit derselben Farbe gekennzeichnet. (C) Bestätigung der aus WGS hergeleiteten Bruchpunktsequenz durch Sanger-Sequenzierung. Das Elektropherogramm zeigt den Sequenzübergang von *TAB3* zu der 27bp Insertion und einem stromabwärts von *CXorf21* gelegenen Bereich.

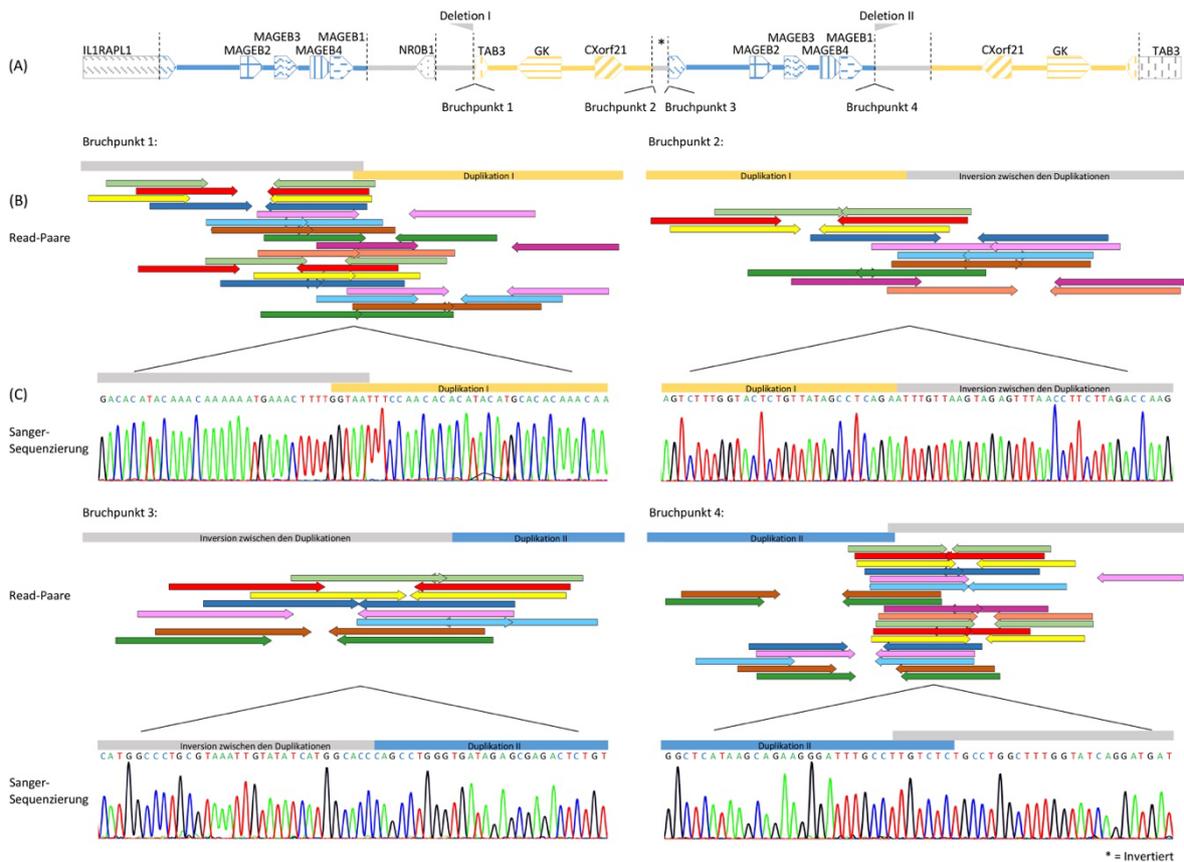


Abb. 9: Ermittlung der strukturellen Variation in P2.

(A) Schematische Darstellung der Sequenzanordnung in der Umgebung von *NROB1* in P2; Darstellung entspricht im Grundprinzip Abb. 8. Der gelbe Bereich stellt die Duplikation I dar, die invertiert zwischen Bruchpunkt 1 (Brp1; ChrX:30,336,745) und Bruchpunkt 2 (Brp2; ChrX: 30,340,605) eingesetzt ist. Das Sternchen * kennzeichnet einen 1.2kb großen und ebenfalls invertierten Bereich der Referenzsequenz zwischen Brp2 und Bruchpunkt 3 (Brp3; ChrX: 30,339,452), der die beiden Duplikationen trennt. Das graue Fähnchen mit der Beschriftung Deletion I markiert das Fehlen des Sequenzbereichs zwischen Brp 1 und Brp 3. Die Duplikation II ist blau dargestellt. Originalsequenz und Duplikat entsprechen sich in ihrer Ausrichtung. Die zwischen Brp 2 und Brp 4 (Brp4; ChrX:30,342,785) liegende Referenzsequenz ist ebenfalls deletiert (Deletion II). (B) Die Sequenzen der an den Bruchpunkten 1 und 4 beteiligten DNA-Abschnitte zeigten Homologien, wodurch eine auf die Base genaue Definition des Bruchpunktes nicht möglich war. Dies ist durch die Überlappung der gelben und grauen bzw. blauen und grauen Blanken dargestellt. (C) Bestätigung der Bruchpunktsequenzen erfolgte durch Sanger-Sequenzierung.

3.4 Transkriptom-Sequenzierung

Bei den strukturellen Variationen bei P1 und P2 handelt es sich um einen bisher unbekanntes Mutationstyp bei Patient:innen mit 46,XY-GD. Beiden Patient:innen wurden im Verlauf ihrer Behandlung die Gonaden entfernt und diese im Rahmen einer pathologischen Aufarbeitung Formalin-fixiert in Paraffin-eingebettet (FFPE) und im Anschluss verwahrt. Um die Wirkung dieser andersgearteten Mutation auf die Expression und die in den Gonaden aktiven Signalwege besser zu verstehen, wurde eine RNA-Sequenzierung aus dem FFPE-Material durchgeführt.

Beim Vergleich der Transkriptome von P1 und P2 mit Transkriptionsdaten von Ovarien (n=133) und Hoden (n=156) aus der GETx Datenbank sowie ArrayExpress-Transkriptionsdaten aus sieben verschiedenen Organen (Großhirn (n=66); Kleinhirn (n=59); Herz (n=46); Niere (n=37); Leber (n=50); Ovar (n=18); Hoden Säugling (n=29); Hoden Erwachsene (n=10)) zeigte sich im t-SNE eine starke gewebeabhängige Gruppierung der Transkriptome. Diese Gruppierung war unabhängig von der Quelle, der die Daten entstammten. Die Daten von P1 und P2 sind ausgesprochen nah beieinander gruppiert, was für einen gemeinsamen Pathomechanismus mit entsprechender ähnlicher Auswirkung auf die Transkription in den Gonaden spricht. P1 und P2 liegen dabei zwischen den Transkriptomen aus Ovarien und Hoden der GETx Datenbank. Sie liegen jedoch näher an den Datenpunkten der Hodenproben, speziell solchen, bei denen eine verringerte Spermatogenese angegeben wird (Abb. 10).

Eine Gegenüberstellung der Expression ausgewählter Gene der Geschlechtsentwicklung im Hoden und Ovarien zeigt, wie zu erwarten, eine deutlich differentielle Expression zwischen den Geweben. P1 und P2 zeigen dabei zumeist ein intermediäres Expressionsverhalten. Zwei charakteristische, gegensinnige Gene für die weibliche und männliche Geschlechtsentwicklung und langfristige Aufrechterhaltung des Ovar- bzw. Hodengewebes sind *forkhead box L2 (FOXL2)* und *doublesex and mab-3 related transcription factor 1 (DMRT1)*. *FOXL2*, welches auch nach der embryonalen Geschlechtsentwicklung weiter die Hodenentwicklung unterdrückt [109, 110], wird selektiv im Ovar und nicht im Hoden exprimiert. Umgekehrt ist *DMRT1* für den Erhalt von Sertoli-Zellen entscheidend [111, 112] und selektiv im Hoden exprimiert. Für beide Gene zeigen die Patient:innen im Vergleich zu den GTEX-Transkriptomdaten des Ovars und Hodens der anderen untersuchten Teenager ein intermediäres Expressions-

verhalten (Abb. III, Anhang S. 94). Wobei P2 sowohl bei *FOXL2* als auch *DMRT1* eine höhere Expression als P1 aufweist.

Zur übersichtlichen Darstellung und zum Verständnis der in P1 und P2 aktiven Signalwege wurde die Gene Set Variation Analyse anhand des *Hallmark gene sets* der MSigDB [100] durchgeführt. Als Vergleich wurden sechs zufällige ovarielle und testikuläre Proben der GTEx und ArrayExpress-Datenbanken (E-MTAB-6814) [96] sowie die gemittelten ovariellen und testikulären Expressionsdaten der GTEx Datenbank verwendet. Bei den Patient:innen zeigt sich dabei Aktivität sowohl von charakteristischerweise eher ovariellen als auch von testikulären Signalwegen. Deutlichstes Beispiel ist hier die Spermatogenese, welche im Hoden, nicht jedoch im Ovar stattfindet (Abb. IV, Anhang S. 95). In den beiden Patient:innen zeigt sich zwar eine niedrige, im Vergleich zum Ovar aber erhöhte Aktivität dieses Signalweges. In umgekehrte Richtung sind der NOTCH- und TGF- β -Signalweg zu nennen, die im Ovar stärker hervortreten (Abb. IV, Anhang S. 95) als im Hoden und ebenfalls bei beiden Patient:innen intermediär aktiv sind. Bemerkenswert ist, dass – obwohl der MYC-Signalweg physiologischerweise in beiden Geweben exprimiert wird – dieser in P1 und P2 fast fehlt.

In der Transkriptomanalyse aus FFPE-Gonadengewebe zeigt sich, dass die Patient:innen postnatal ein ähnliches intermediäres Expressionsverhalten haben und keine klare Zuordnung des dysgenetischen Gonadengewebes in Richtung Hoden oder Ovargewebe möglich ist. Das auffällige Fehlen des MYC-Signalweges, welcher vielfältige zelluläre Prozesse wie Proliferation, Differenzierung und Apoptose steuert [113], könnte ein Hinweis auf eine Wirkung der Mutation auf das Zellwachstum sein.

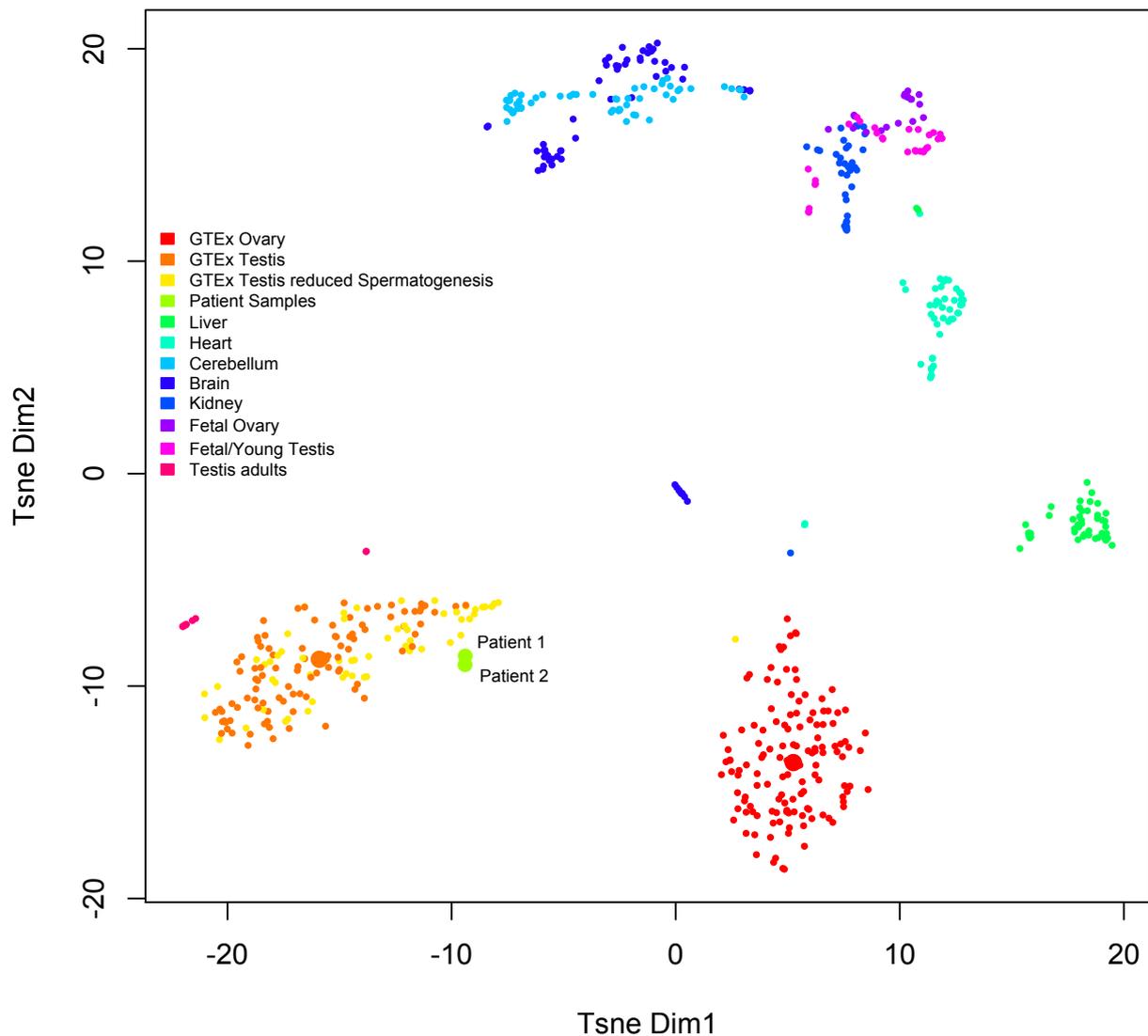


Abb. 10: t-SNE zum Vergleich der Expressionsdaten von P1 und P2 mit gesunden Kontrollgeweben

In der zweidimensionalen t-SNE Darstellung sind die Patient:innenproben (große grüne Punkte) im Vergleich zu GETx-Daten aus Ovarien (rot) und Hoden (orange) bzw. Hoden mit geringer Spermatogenese (gelb) dargestellt. Die großen roten und orangen Punkte sind der Mittelwert der entsprechenden Transkriptomdaten eines Gewebes. Zusätzlich wurden ArrayExpress-Daten (E-MTAB-6814) für die Organe Leber, Herz, Großhirn, Kleinhirn und Niere miteinbezogen. Die Berechnung erfolgte mit 608 Gewebeproben anhand von 5.000 verschiedenen Genen. Diese wurden zufällig aus 13.005 Genen ausgewählt, die in P1 und P2 eine TPM>1 aufwiesen.

3.5 Screening weiterer Patient:innen auf Xp21.2 Kopiezahlvariationen

Die in P1 und P2 aufgedeckten strukturellen Variationen stellen die ersten beschriebenen mit 46,XY-GD assoziierten Xp21.2-Duplikationen dar, die nicht *NROB1* beinhalten. Lediglich eine weitere Deletion stromaufwärts von *NROB1* wurde bereits bei einer Patient:in mit 46,XY-GD beschrieben.[114] Ferner ist auch die Gesamtzahl der in den letzten 20 Jahren auf englisch publizierten *NROB1* beinhaltenden, nicht-syndromalen Duplikationen mit einer Anzahl von sieben recht gering. Daher stellte sich die Frage, ob Kopiezahlvariationen an diesem Genort als Ursache von 46,XY-GD unterdiagnostiziert sind aufgrund einer zu seltenen Testung auf Xp21.2 CNVs, oder im Falle von Duplikationen stromaufwärts von *NROB1* diese durch fehlendes Augenmerk bei der Auswertung nicht auffallen. Eventuell neu ermittelte Fälle könnten zu einem besseren Krankheitsverständnis beitragen.

Um dies zu überprüfen, waren in einem ersten Schritt 2 417 Patient:innenakten nach den in der Methodik beschriebenen Einschlusskriterien auszuwerten. Hiervon erfüllten 151 Patient:innen die Einschlusskriterien für das zu diesem Zwecke etablierte Screening mittels qPCR. Hinzu kamen zum Screening drei Patient:innenproben der Christian-Albrechts-Universität (CAU) / Kiel sowie elf Patient:innenproben des All India Institute Of Medical Sciences (AIIMS), New Delhi / Indien, welche die Einschlusskriterien erfüllten.

Die qPCR zeigte bei acht Patient:innen auffällige Ergebnisse. Alle Auffälligkeiten deuteten auf Kopiezahlvermehrungen ohne Deletionen hin. Bei einer Patient:in ließ sich die Duplikation in Folgeversuchen nicht bestätigen. Zwei Fälle stellten sich nach Rücksprache mit den behandelnden Ärzten und Ärztinnen als Patient:innen mit 46,XX-Chromosomensatz heraus. Zwei weitere Patient:innen sind bereits bei Ledig et al. publizierte Fälle von *NROB1* einschließenden Duplikationen, wovon eine syndromal ist.[67] Bei zwei der drei verbleibenden Fälle ließ sich mittels *SRY*-PCR ein 46,XY-Chromosomensatz nicht bestätigen (Abb. V, Anhang S. 96).

Abbildung 11 zeigt eine übersichtliche Darstellung dieses Vorgehens.

Somit blieb ein Fall (fortan P3 genannt) einer durch das Screening neu entdeckten Xp21.2 Kopiezahlvermehrung. Die qPCR Ergebnisse des initialen Screenings und allen weiteren Bestätigungsläufen sind am ehesten mit einem dreifachen Vorliegen von *NROB1* und dessen

3.6. Bestätigung und Charakterisierung der Triplikation von P3

Die qPCR, als einfache Screeningmethode, lieferte Hinweise auf das Vorliegen einer Kopiezahlvermehrung von P3, ließ aber einige Fragen unbeantwortet. Zum einen, ob es sich tatsächlich um eine Triplikation oder doch eine Duplikation oder gar einen Fehler der neu etablierten Methode handelt. Außerdem erlaubt eine qPCR nur eine sehr grobe Schätzung der Größe von CNVs und kann in diesem Fall nicht vorhersagen, ob der Bereich *MAGEB1* bis *TAB3* durchgehend betroffen ist oder es Unterbrechungen gibt. Auch bleibt dabei unklar, wo die zusätzliche(n) Kopie(n) inseriert sind. Um diesen Fragen zu klären, musste zunächst ein aCGH und danach ein WGS durchgeführt werden.

Im aCGH ließ sich eine Triplikation bestätigen, denn im Vergleich zu normalen 46,XY-Kontrollproben zeigte sich ein Hybridisierungsverhältnis von 1,5 (eine Kopie =0; zwei Kopien =1). Bis zu den Enden der Triplikation wiesen 67 Sonden ein einheitliches erhöhtes Hybridisierungsniveau auf, was gegen eine Unterbrechung spricht. Die Triplikation umspannt damit 1,2Mb inklusive der Gene *MAGEB1-4*, *NROB1*, *CXorf21*, *GK* und *TAB3* sowie einem Teil von *IL1RAPL1* (Abb. 13).

Die bioinformatische Analyse der WGS-Daten mittels Control-FREEC identifizierte 436 CNVs, davon 13 auf dem X-Chromosom und vier auf Xp21.2. Alle vier wurden als Kopiezahlvermehrung (gain) beschrieben. DELLY erkannte 8 261 SVs, mit 205 auf dem X-Chromosom und zehn auf Xp21.2. Dabei handelte es sich um fünf Translokationen, drei Duplikationen und zwei Deletionen. Es gab keine basengenaue Übereinstimmung der Ergebnisse der zwei Programme. Die Visualisierung aller vermeintlichen Bruchpunkte im IGV bestätigte das Vorliegen einer Kopiezahlvariation des Bereichs chrX:29,849,782-31,088,713 und schloss die anderen Bruchpunkte als Programmfehler aus. Der Bereich entsprach bis auf eine Base genau einer durch DELLY ermittelten Duplikation. Alle vier von Control-FREEC eingestuftten ‚gains‘ liegen im Triplikationsbereich, das Programm hat diese aber nicht klar als ein durchgehendes Geschehen mit korrekten Bruchpunkten erkannt.

Viele der weiteren von DELLY und Control-FREEC identifizierten Bruchpunkte und davon abgeleiteten Kopiezahlvariationen befanden sich in der bereits bei P2 diskutierten simple-repeat-reichen Region ChrX:30,807,585-30,814,379. Die fünf von DELLY ermittelten

Translokationen wurden alle an demselben Ort auf Chromosom 2 verortet, zusammen mit vielen Partner-Reads von anderen Chromosomen. Das deutet darauf hin, dass es sich hierbei um eine Region hoher Fehlkartierung von Reads handelt, die zu falsch positiven Ergebnissen des SV-Erkennungsalgorithmus führen.

Control-FREEC beschreibt die CNVs im Triplikationsbereich lediglich als *gain*, wohingegen DELLY diese als Duplikation beschreibt. Letzteres liegt daran, dass das Programm nicht auf die Einstufung von Triplikationen ausgelegt ist. Beim Vergleich der Sequenzierungstiefe des betroffenen Bereichs chrX:29,849,782-31,088,713 mit dessen Umgebung besteht ungefähr ein Verhältnis von 3:1 (~ 60:20). Gemeinsam mit dem aCGH bestätigt dies den initialen Verdacht durch das qPCR-Screening, dass bei der Patient:in eine Triplikation vorliegt.

Wie bereits bei P1 und P2, wurde auch hier durch das Extrahieren der einzelnen Partner-Reads an den Bruchpunkten und deren Neuordnung die tatsächliche Anordnung der strukturellen Variation in der Patient:in abgeleitet und im Anschluss mittels PCR und Sanger-Sequenzierung bestätigt. Das Original und die replizierten Segmente sind in Tandem arrangiert und lediglich am Übergang durch eine 64bp lange Zwischensequenz getrennt (Abb. 14B). Die Zwischensequenz besteht aus 49bp, die aus dem *DMD*-Gen stammen, sowie 15bp mit unklarem Ursprung.

Auch bei P3 wurden die WGS-Daten weiterführend zum Ausschluss anderer bekannten DSD-Ursachen ausgewertet. Dabei fiel eine heterozygote missense SNV im *estrogen receptor 2 (ESR2)* auf (Tabelle 3). Homo- und heterozygote Mutationen in *ESR2* wurden zuletzt mit partieller und kompletter 46,XY-GD assoziiert.[115] Die in P3 gefundene Mutation ist jedoch noch nicht beschrieben und liegt nicht in unmittelbarer Nähe einer bisher charakterisierten Mutation mit vermuteter pathologischer Auswirkung. Vor dem Hintergrund bereits vielfach beschriebener *NROB1*-Duplikationen und deren anerkannter Ätiologie bei 46,XY-GD bleibt die *NROB1*-Triplikation der wahrscheinlichste Auslöser des Phänotyps von P3. Die SNV in *ESR2* dürfte maximal eine additive Wirkung haben.

Tab. 3: Heterozygote missense SNV von Patient:in 3 in bekanntem DSD-Gen *ESR2*.

Gene	dbSNP ID	ExAC All [52]	ExAC European Non-Finish [52]	SIFT [81]	Polyphen2 [82]	MutationTaster [83]
ESR2	rs367855747	0.0002	0.0004	D (0.021)	D (0.975)	D (0.999)

D = Schädlich („Deleterious“). Der Zugriff auf die ExAC-Datenbank erfolgte im März 2020.

Um die Entstehung der Triplikation besser zu verstehen, wurde die Mutter von P3 ebenfalls mittels qPCR und aCGH untersucht (Abb. VI, Anhang S. 97). Hierbei zeigte sich ebenfalls ein dreifaches Vorliegen von *NROB1* und der benachbarten Gene. Aufgrund der physiologisch bei 46,XX vorliegenden doppelten Kopiezahl ist demnach davon auszugehen, dass sie lediglich eine Duplikation trägt und keine Triplikation. Die 64bp lange Zwischensequenz am Übergang der beiden beteiligten Segmente konnte bei der Mutter ebenfalls mittels Sanger-Sequenzierung nachgewiesen werden.

In P3 liegt demnach eine Triplikation als wahrscheinlichste Ursache der 46,XY-GD vor, die neben *NROB1* auch die Bereiche der Duplikationen von P1 und P2 umfasst. Ebenfalls in Analogie zu den anderen beiden Patient:innen wurden die replizierten Bereiche in unmittelbarer Nähe zur Originalsequenz eingebaut und haben einen maternalen Ursprung. Denn die Triplikation ist zwar de novo entstanden, doch auf dem Boden einer bereits existierenden Duplikation des entsprechenden Bereichs.

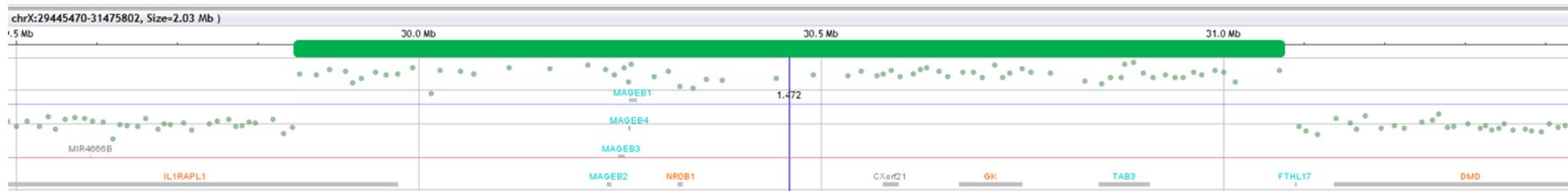


Abb. 13: ArrayCGH Ergebnis von P3 in der Region von *NROB1*

Die Sonden zeigen einen Mindestbereich der Triplikation von ChrX:29,851,537-31,069,736 ohne erkennbare Unterbrechungen. Der Triplikationsbereich endet telomerisch in *IL1RAPL1* und zentromerisch kurz vor *FTHL17*.

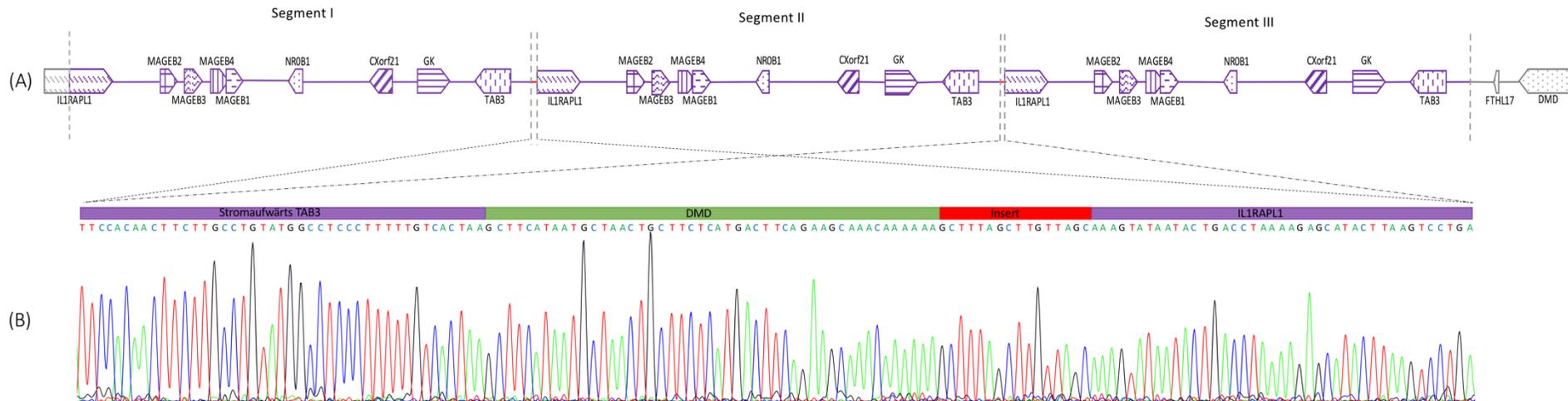


Abb. 14: Ermittlung der strukturellen Variation in P3

(A) Schematische Darstellung der Sequenzanordnung in der Umgebung von *NROB1* in P3. Die Darstellung entspricht im Grundprinzip den Abb. 8 & 9. Die triplizierten Segmente sind in Tandem arrangiert und lila dargestellt. Nicht beteiligte Bereiche sind grau. Die aus den chirmären WGS-Read-Paaren rekonstruierte Bruchpunktsequenz zeigt, dass zwischen den von der Replikation betroffenen Bereichen, zwei identische 64bp lange Zwischensequenzen liegen. Da die Übergänge von Segment I zu Segment II und von Segment II zu Segment III deckungsgleich sind, konnten die extrahierten chimären Read-Paare nicht genau einem der beiden Übergänge zugeordnet werden. Auf ihre Darstellung wird hier folglich verzichtet. (B) Bestätigung der Bruchpunktsequenz durch Sanger-Sequenzierung.

3.7. Identifikation von Isolator-Bindungsstellen

Neben den Duplikationen von P1 und P2 ist nur eine weitere Kopiezahlvariation stromaufwärts von *NROB1* bei einem Fall mit 46,XY-GD beschrieben. Hierbei handelt es sich jedoch um eine Deletion.[114] Smyk et al. gehen beim Pathomechanismus der Deletion von der Entfernung eines Repressor-Elements von *NROB1* aus.[114] Hierfür wurden damals vier konservierte regulatorische Regionen im Bereich der Deletion als mögliche Repressoren postuliert. Alle vier Regionen sind auch von der Duplikation in P2 eingeschlossen [114] und müssten somit eine Verstärkung der Repression hervorrufen, die nicht im Einklang mit dem bei beiden Patient:innen beschriebenen Phänotyp steht. Denn der Phänotyp ist auch bei den Duplikationen eher mit einer verstärkten *NROB1*-Expression in der Embryonalphase zu erklären. Aufgrund der großen Überlappung der Deletion von Smyk et al. [114] und der Duplikation von P1 stellt sich die Frage, wie eine Deletion und Duplikation eine analoge pathologische Wirkung haben können?

Vor diesem Hintergrund entstand die Vermutung, es könnte sich hier um die Alteration eines Isolators handeln. Ein solcher Isolator müsste dann sowohl durch Deletionen als auch entsprechende Duplikationsanordnungen umgangen werden. Um die Frage zu klären, ob es einen entsprechenden Isolator in dieser Region gibt, wurden Chip-Seq-Daten zur Bindung von Transkriptionsfaktoren der ENCODE-Datenbank [116, 117] im Überlappungsbereich der stromaufwärts von *NROB1* gelegene Duplikation von P1, P2 und Deletion von Smyk et al. untersucht. Dieser Überlappungsbereich von 35kb (ChrX:30,561,644-30,596,386) ist auch in der Triplikation von P3 vorhanden.

Die Chip-Seq-Daten zeigen in diesem Bereich Bindungsstellen der Proteine CTCF, SMC3 und RAD21 (Abb. 15). Der CCCTC binding factor (CTCF) ist ein Zinkfinger-Protein und der einzige in Wirbeltieren bekannte Isolator.[118] Ein Isolator besitzt nach herkömmlicher Definition einerseits die Fähigkeit, die Inaktivierung eines Gens durch die Ausbreitung von Heterochromatin zu verhindern und andererseits Enhancer-Promotor-Interaktionen zu unterbinden.[119, 120] Wenn ein CTCF zwischen Enhancer und Promotor bindet, ließ sich sowohl in Plasmiden als auch im nativen Chromatin nachweisen, dass die Wirkung des Enhancers auf den entsprechenden Promotor aufgehoben ist.[118, 121-125] Der dem Phänomen zugrunde liegende Mechanismus wird mit dem Ausbilden von Chromatinschleifen erklärt, an deren Basis sich auf beiden Seiten des schleifenbildenden Chromatinstrangs ein CTCF befindet.[120] DNA-

Bereiche innerhalb der Schleife bilden eine Domäne und können miteinander in Berührung kommen. Regulatorische Elemente, die sich mit einem Gen in einer gemeinsamen Chromatinschleife befinden, können demnach potentiell auf dieses Gen wirken. DNA-Bereiche innerhalb einer Schleife haben jedoch kaum Kontaktpunkte mit Bereichen außerhalb der Schleife und dort gelegenen Enhancern. Diese assoziierenden, in Schleifen geformten DNA-Bereiche wurden nach ihrem experimentellen Nachweis als ‚Topological Domains‘ [126] oder ‚Topologically Associating Domains‘ (TADs) [127] benannt. TADs reichen in ihrer Größe bis in den Megabasenbereich und Gene innerhalb einer TADs zeigen ähnliche Expressionsmuster.[126-128] Die Störung einer TAD durch Veränderung der TAD-Grenze kann zu ektopten DNA-Kontakten und veränderter Genexpression führen.[127, 129, 130] Neben CTCF sind auch die Proteine RAD21 und Structural Maintenance of Chromosomes 3 (SMC3) charakteristisch für solche TAD-Grenzen.[131, 132]

Die im Intron 2 von *CXorf21* identifizierten Bindungsstellen von CTCF, RAD21 und SMC3 stellen somit eine mögliche Grenze dar, die zwei regulativ verschiedene DNA-Bereiche voneinander trennt.

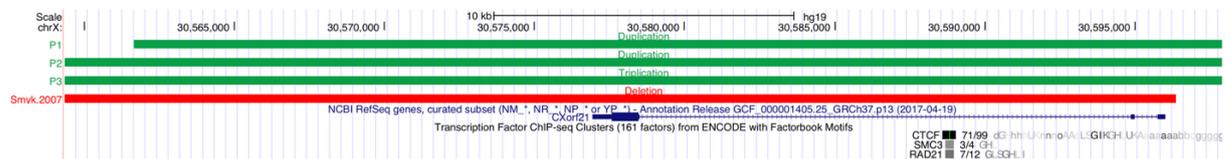


Abb. 15: Isolator-Bindungsstellen in minimaler Überlappungsregion

Daten der ENCODE-Datenbank [116, 117] sind im UCSC-Browser [107] für den Ausschnitt der 35kb Überlappungsregion zwischen P1 und der Deletion von Smyk et al. dargestellt. Im Überlappungsbereich befindlichen Intron 2 von *CXorf21* finden sich konservierte Bindungsstellen der Proteine CTCF, SMC3 und RAD21.

3.8. Hi-C Analyse

Ein Großteil der gebunden CTCFs sind nicht an der Ausbildung von Schleifen/TADs beteiligt.[133] Um zu untersuchen, ob die in Intron 2 von *CXorf21* identifizierten CTCF-, SMC3- und RAD21-Bindungsstelle mit einer TAD-Grenze zusammenfallen, wurden im Penn State 3D Genome Browser [106] gepoolte, öffentlich zugängliche Hi-C-Daten ausgewertet. Zum Zeitpunkt der Arbeit gab es keine verfügbaren Hi-C-Daten aus Gonaden- oder Hodengewebe. Entsprechend wurden zur Orientierung die Hi-C-Daten aller verfügbaren Zelllinien genutzt.

Ein Großteil der 36 betrachteten Zelllinien zeigt eine TAD-Grenze zwischen *NROB1* und *GK*, die in die bereits beschriebene 35kb Überlappungsregion fällt (Abb 16). Lediglich in den Zelllinien IMR90 (fötale Lungenfibroblasten), Aorta, Spleen, SJCRH30 (Rhabdomyosarkom) und K562 (Leukozyten) fehlte eine Trennung der Domänen in dem Bereich. Im Thymus liegt *NROB1* zwischen den berechneten TADs. Dennoch ist die TAD-Grenze insgesamt zwischen den Zelllinien ungewöhnlich konserviert. Das verbreitete Vorliegen einer TAD-Grenze in dem Bereich des zuvor identifizierten CTCFs legt insofern eine isolatorische Wirkung durch Schleifenbildung an dieser Stelle nahe.



Abb. 16: TAD-Konformation in verschiedenen menschlichen Zelllinien am *NROB1*-Locus

Hi-C-Daten des Penn State 3D Genome Browsers [106] wurden im UCSC-Browser [107] visualisiert. Grüne Balken stellen die Bereiche der Kopiezahlvariationen von P1, P2, P3 dar. Der rote Balken ist der deletierte Bereich bei Smyk et al. [114]. In blau sind jeweils die *NROB1* beinhaltenden TADs dargestellt, sowie ihre zentromerische und telomerische Nachbar-TADs für 36 unterschiedliche menschliche Zelllinien.

3.9 Identifikation putativer Enhancer-Elemente

Die in fast allen Zelllinien in einem gemeinsamen TAD liegenden Gene *NROB1* und *MAGEB1-4* (Abb. 16) zeigen ein ähnliches, hodendominiertes Transkriptionsverhalten (GETx Datenbank; Zugriff März 2020).[134] Um besser zu verstehen, wovon der in Intron 2 von *CXorf21* identifizierte Isolator die Gene des *NROB1*-TAD eventuell abschirmt, wurden Enhancer-Elemente und deren putativen Zielgene der GeneHancer-Datenbank [135] in der Region stromaufwärts von *NROB1* ausgewertet.

In der Region 500kb zentromerisch der *NROB1*-TAD Grenze finden sich zwei Enhancer-Cluster. Eins in der Region um *CXorf21* und *GK* (Abb. 17) das andere um *TAB3*. Das Enhancer-Cluster um *TAB3* befindet sich in vielen Zelllinien jedoch in einem eigenständigen TAD mit *TAB3*, *FTHL17* und *DMD* (Abb. 16). So sind diese Enhancer ihrerseits durch einen eigenen Isolator von einer Kommunikation mit *NROB1* abgeschirmt. *CXorf21* und *GK* liegen hingegen in dem Zwischenbereich zweier TADs und sind demnach von *NROB1* nur durch einen einzigen Isolator bzw. die TAD-Grenze getrennt. Gene in einem TAD-Grenzbereich sind häufig durch ubiquitäre und kontinuierliche Expression im gesamten Entwicklungsverlauf gekennzeichnet.[128] Diese Beobachtung deckt sich mit GTEx-Expressionsdaten [134], da *CXorf21* und speziell *GK* in deutlich mehr Zelllinien exprimiert werden als *NROB1* oder *MAGEB1-4*. Enhancer, die auf die Gene in der Zwischenregion wirken, führen also zu einer deutlich stärkeren und weniger selektiven Expression als dies für *NROB1* physiologisch ist, denn *NROB1* ist ein Entwicklungsgen, das normalerweise zeitlich und geweblich differenziert exprimiert wird. Ein Kontakt mit diesen Enhancern könnte demnach zu einer übermäßigen Expression von *NROB1* in der entscheidenden Phase der embryonalen *SOX9*-Aktivierung führen. Die dadurch entstehende, mangelnde *SOX9*-Expression steht dann einer erfolgreichen Sertolizell- und männlichen Geschlechtsentwicklung im Wege.

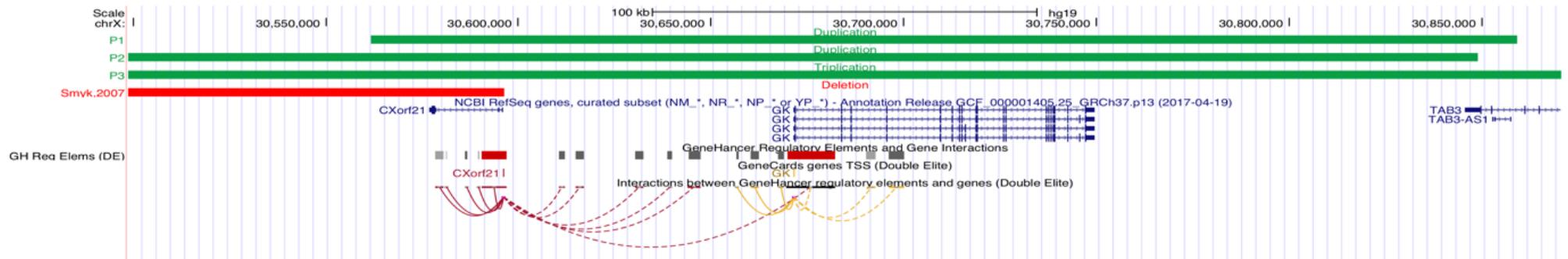


Abb. 17: GeneHancer Promotoren und Enhancer stromaufwärts von *NROB1*

Daten der GeneHancer-Datenbank [135] wurden im UCSC-Browser [107] visualisiert. Rote Balken in der GH Reg Elems (DE) -Spalte stellen Promotoren dar. Graue Balken sind putative Enhancer-Elemente. Die Intensität des Grautons spiegelt hierbei das Konfidenzmaß des prognostizierten Enhancers wider.

3.10 Potentieller Pathomechanismus der Kopiezahlvariationen am Xp21.2 Locus durch Disruption Topologisch Assoziierender Domänen (TADs)

Die Herleitung der genauen Anordnung der strukturellen Variationen und die Lokalisierung eines Isolators mit vermutlich TAD-Grenzen definierender Wirkung sowie die Identifikation putativer Enhancer-Elemente ermöglichte eine Erklärung des potentiellen Pathomechanismus der in dieser Arbeit beschriebenen Kopiezahlvariationen am Xp21.2-Locus. Dabei wurde das Loop-Extrusion-Model angewendet.[131, 136]

In der Vergangenheit wurde gezeigt, dass SVs, die TAD-Grenzen entfernen oder verschieben, Einfluss auf langstreckige, regulative Interaktionen haben können, mit der Folge von pathologischen Entwicklungsstörungen [123, 137, 138] und Onkogenese [139-141]. Der in *CXorf21* befindliche CTCF, dessen Bindungsmotiv Teil der in P1, P2, P3 beschriebenen Kopiezahlvariationen ist, ist als Isolator wahrscheinlich an der Ausbildung einer solchen TAD-Grenze beteiligt. Das CTCF Protein bindet mit einer direktionalen Ausrichtung an eine entsprechend orientierte Bindungssequenz. Dabei bildet sich in aller Regel eine TAD zwischen zwei entgegengesetzten CTCF-Motiven.[132, 142, 143] Daher können sowohl das Entfernen als auch das Invertieren des CTCFs zu einem Aufheben oder Verschieben der TAD-Grenzen führen [131, 144]. Das in *CXorf21* identifizierte CTCF-Motiv ist in Richtung von *NROB1* ausgerichtet, was dazu passt, dass es sich hierbei um die zentromere Begrenzung der *NROB1*-TAD handelt (Abb. VII, Anhang S.98). Dieser CTCF trennt *NROB1* von den jenseits der TAD-Grenze befindlichen Enhancern der Gene *CXorf21* und *GK*. Durch die SVs kann es jetzt zu einem Umgehen der isolierenden Wirkung des CTCF kommen. Die Enhancer gelangen mit *NROB1* in eine gemeinsame Domäne und können damit zu unphysiologischen Interaktionen und zu ektopter Expression führen. Für dieses Konzept wurde in der Vergangenheit der Name ‚enhancer adoption‘ geprägt.[138]

Die in P1 vorliegende Tandemduplikation führt zur Insertion eines zusätzlichen CTCF-Motivs (Abb. VII, Anhang S. 98). Das Loop-Extrusion-Model besagt, dass jedoch nur eins der zwei CTCFs mit dem entgegengesetzten CTCF am anderen TAD-Ende interagieren und die TAD-Grenze ausbilden kann.[131, 136] Man nimmt an, dass es sich beim Loop-Extrusion um einen dynamischen Prozess handelt [145], bei dem der Structural Maintenance of Chromosomes (SMC) -Komplex (Cohesin) ATP-abhängig in beide Richtungen am Chromosom entlang

wandert und dabei Chromatinschleifen ausbildet. Dadurch werden zunehmend weiter entfernte DNA-Abschnitte in räumliche Nähe gebracht – ein entscheidender Vorgang, damit weit entfernte regulative Elemente zu den zugehörigen Promotoren finden können. Dieser Prozess findet so lange statt, bis der Komplex dissoziiert und den DNA-Strang verlässt oder auf ein entsprechend ausgerichtetes CTCF stößt (Abb. 18 I). Es wird davon ausgegangen, dass die Chromatinwanderung in die entgegengesetzte Richtung fortgesetzt wird, wenn das Cohesin in der einen Richtung eine Grenze (CTCF) erreicht hat.[136, 146] Durch das regelmäßige Beladen und Dissoziieren von DNA ist dies ein sehr dynamischer Prozess und die TADs haben keine durchgehend feste Struktur. Vielmehr handelt es sich bei den TAD-Grenzen um eine Position, an denen das Cohesin eine erhöhte Verweildauer hat und somit die daraus resultierende Konformation als Zustand der Chromatinfaltung favorisiert ist.[147]

In P1 ergeben sich daher zwei mögliche TAD-Konfigurationen mit erhöhter Verweildauer des Cohesins. Welche sich ausbildet, hängt davon ab, an welcher Position das Cohesin auf den Chromatinstrang beladen wird. Wenn sich die TAD mit dem proximal zu *NROB1* gelegenen CTCF ausbildet, entspräche dies im Grunde der Konfiguration im Gesunden mit aktiver Isolation zwischen *NROB1* und den Enhancern (Abb. 18 II P1b). In Situationen, in denen sich die TAD-Grenze am distalen CTCF ausbildet, befänden sich die Enhancer von *CXorf21* und *GK* mit *NROB1* in einer gemeinsamen Domäne und es könnte so zur Enhancer adoption kommen (Abb. 18 II P1a).

Die Inversion von P2 hebt die isolierende Wirkung des CTCFs im duplizierten Bereich auf. Im Gegensatz dazu behalten Enhancer innerhalb der Duplikation ihre Wirkung unabhängig von ihrer Ausrichtung. Im Gegensatz zu P1 leitet sich also nur eine TAD-Konformation ab. In dieser sind die Enhancer und *NROB1* in einer gemeinsamen Domäne (Abb. 18 III P2).

Von der Triplikation bei P3 ist nicht nur der CTCF und die Enhancer zwischen *CXorf21* und *GK* betroffen, sondern auch *NROB1* selbst. Von den mithin drei vorliegenden CTCFs in den verschiedenen *CXorf21*-Kopien kann immer nur einer die TAD-Grenze ausbilden (Abb. 18 IV P3a-c). Bei den anderen zwei Bereichen kommt es demnach zu einer fehlenden Isolation zwischen *NROB1* und den hinter den CTCF liegenden Enhancern (Abb. VII, Anhang S. 98).

Bei der von Smyk et al. [114] beschriebenen Deletion kommt es zur Entfernung des Isolators, die putativen Enhancer sind jedoch nicht von der Deletion betroffen. Somit hebt die Deletion

die Abschirmung vor diesen Enhancern auf (Abb. VII, Anhang S. 98). Vielfach wurde beschrieben, dass die Deletionen einer TAD-Grenze zu vermehrten Kontakten außerhalb der vorher bestehenden TAD-Grenze führen [127, 129-131].

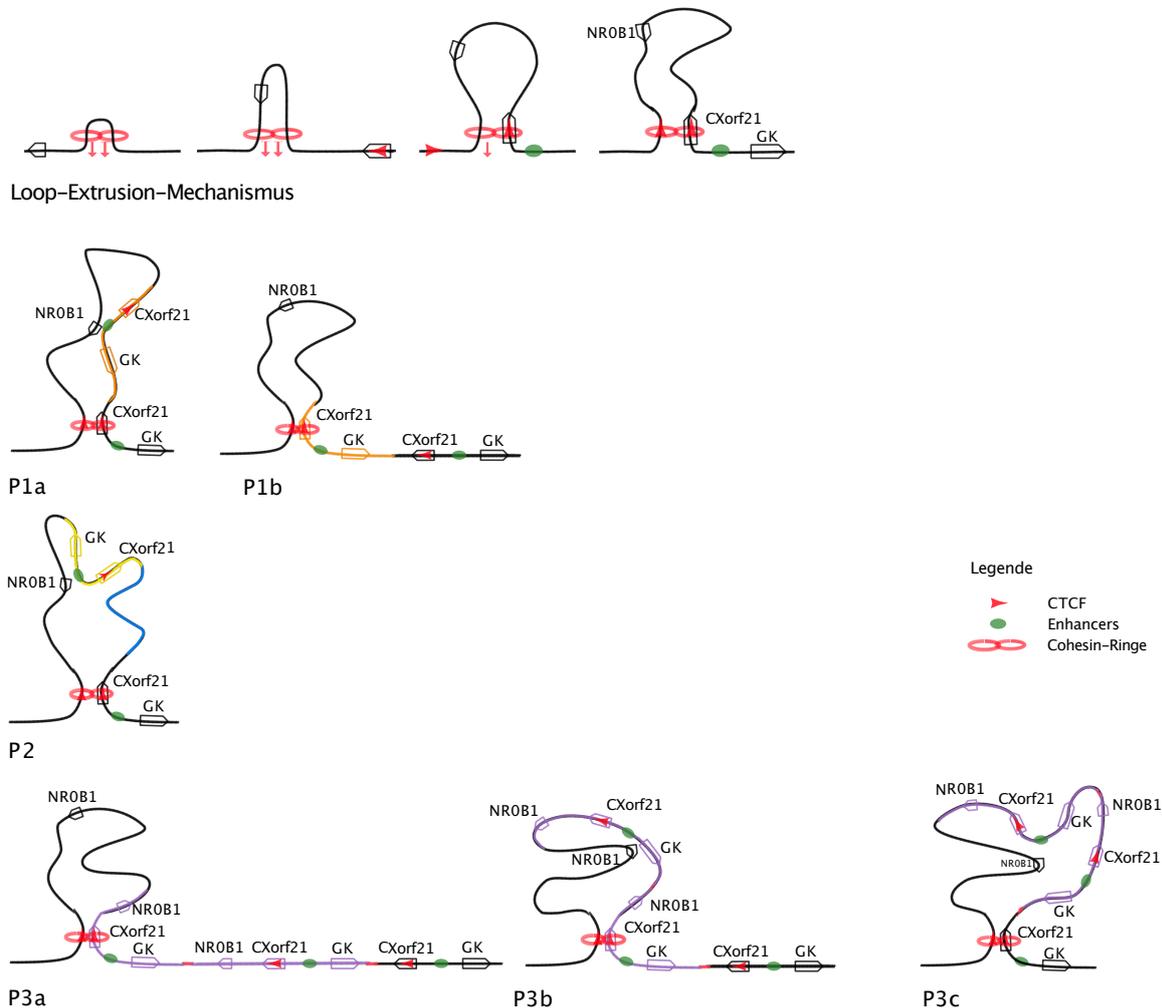


Abb. 18: Auswirkung struktureller Variationen auf mögliche DNA-Faltung in P1 bis P3 im Vergleich zur regulären Anordnung

Topologisch assoziierende Domänen (TAD) sind vereinfacht als Schleifen des Chromatinstrangs dargestellt. Gene sind durch Pfeile symbolisiert, deren Ausrichtung der Leserichtung des Gens entspricht. Im Sinne der Übersichtlichkeit sind jeweils ausschließlich die Gene *NROB1*, *CXorf21* und *GK* dargestellt. I) Ausbildung der regulären *NROB1*-TAD im Sinne des Loop-Extrusion-Modells.[131, 136] Die TAD wird durch den Cohesin-Komplex ausgebildet, dessen zwei Cohesin-Ringe (rote Kreise) sich in entgegengesetzter Richtung entlang des DNA-Strangs bewegen. Erreicht einer der Cohesin-Ringe einen entsprechend ausgerichteten CTCF (roter Pfeil), kommt der Prozess auf einer Seite zum Stehen. Der andere Cohesin-Ring bewegt sich hingegen weiter entlang der DNA bis dieser ebenfalls einen korrekt ausgerichteten CTCF erreicht und damit die Grenzen der TAD, bis zum Dissoziieren des Cohesin-Komplexes, fixiert sind. Regulär sind die zwischen *CXorf21* und *GK* gelegenen Enhancer (grüner Punkt) nicht mit *NROB1* in einem gemeinsamen TAD und dadurch voneinander abgeschirmt. II-IV) Duplizierte bzw. triplizierte DNA-Bereiche der P1, P2 und P3 sind als farbige Segmente gekennzeichnet und enthalten jeweils auch das vervielfältigte CTCF-Motiv in *CXorf21*. Durch das doppelte oder dreifache Vorliegen des CTCFs ergeben sich zum Teil mehrere mögliche TAD-Konfigurationen, abhängig davon an welcher Stelle der Cohesin-Komplex initial an die DNA andockt. Nur wenn bei einer TAD-Konfiguration die zwischen *CXorf21* und *GK* gelegenen Enhancer mit *NROB1* und dessen Promotor in eine gemeinsame TAD gelangen, können diese miteinander in Kontakt treten.

4. Diskussion

Von Duplikationen am Xp21.2-Locus, die *NROB1* einschließen, ist bekannt, dass sie eine mögliche Ursache für 46,XY-Gonadendysgenesien darstellen. Die neu aufgedeckten Fälle von Duplikationen stromaufwärts von *NROB1* bei Patient:innen mit 46,XY-GD stehen nicht im Einklang mit dem bisherigen Verständnis des Pathomechanismus, der auf einer doppelten *NROB1*-Gendosis beruht. WGS im Rahmen dieser Arbeit erlaubte eine genaue Charakterisierung des Aufbaus der in den Patient:innen vorliegenden SVs und den Ausschluss anderer bekannter Ursachen für ihre Varianten der Geschlechtsentwicklung. Umgekehrt legte auch die Transkriptomsequenzierung nahe, dass bei P1 und P2 eine gemeinsame Ursache für deren Varianten der Geschlechtsentwicklung vorliegt. Schließlich führten qPCR-Untersuchungen von 151 ähnlichen Patient:innen zu der erstmaligen Identifizierung einer Xp21.2-Triplikation und deckte auf, dass es sich bei Kopiezahlvariationen an diesem Genort um eine seltene Ursache von 46,XY-DSD handelt.

In Zusammenschau der Kopiezahlvariationen der Patient:innen P1 bis 3 mit einer zuvor publizierten Deletion [114] in dem Chromosomenbereich zeigte sich eine gemeinsame Überlappungsregion rund um *CXorf21*. Die Auswertung von Chip-Seq [116], GeneHancer [135] und Hi-C Daten [106] in dem entsprechenden Bereich ermöglichte es, einen neuen Mechanismus zu postulieren, wie die genannten Mutationen unabhängig der *NROB1*-Gendosis bzw. Kopiezahl wirken.

4.1. Whole Genome Sequencing

4.1.1 Möglichkeiten und Grenzen bei der Erkennung von strukturellen Variationen

Durch die gezielte Auswertung eines vorher bestimmten Chromosomenbereichs war es möglich, aus den WGS-Daten der Patient:innen 1 bis 3 den Nachweis von Kopiezahlveränderungen zu erbringen. Eine weniger vordefinierte Auswertung nur anhand der von den Programmen DELLY [85] und Control-FREEC [84] identifizierten SVs wäre hingegen deutlich schwieriger gewesen. Control-FREEC wies eine ausgesprochen geringe Sensitivität auf und erkannte die Kopiezahlveränderungen der Patient:innen gar nicht oder nicht korrekt. Auch DELLY offenbarte bei P2 nicht alle Bruchpunkte und stufte ein Segment

fälschlich als Inversion ein. Zusätzlich zeigten beide Programme eine hohe Zahl falsch positiver Ergebnisse.

In den vergangenen Jahren sind viele Auswertungsprogramme entwickelt worden, die teils unterschiedliche Ansätze zur SV-Erkennung nutzen.[148-153] Bei zukünftigen Analysen wäre daher die Erstellung oder Nutzung eines bereits konzipierten ‚Ensemble-Algorithmus‘ [154, 155] möglich. Bei einem Ensemble-Algorithmus werden die Ergebnisse verschiedener Programme miteinander abgeglichen und übereinstimmende Ereignisse ausgegeben. Genregionen, die in verschiedenen Programmen als auffällig markiert werden, könnten dann in der Auswertung priorisiert werden.

Bei der Auswertung fiel auf, dass die einzelnen Programme bei verschiedenen Patient:innen immer wieder an ähnlichen Stellen falsch positive Ergebnisse lieferten. Somit ist zukünftig auch eine Subtraktionsmethode denkbar, bei der die SV-Ergebnisse einer Patient:in mit denen einer oder mehrerer Kontrollproben abgeglichen werden und nur die Ergebnisse hervorgehoben werden, die exklusiv bei der Patient:in vorkommen. Dies entfernt neben häufigen falsch positiven Bereichen auch in der Bevölkerung verbreitet vorkommende nicht-pathogene SVs.

Grundsätzlich ist das für diese Arbeit verwendete Short-read-WGS anderen Methoden zur Kopiezahlidentifizierung, wie beispielweise Microarrays, überlegen, da es eine höhere Auflösung hat und auch balancierte SVs erkannt werden können, sowie deren genaue Anordnung. Da die Erkennung aber zu großen Teilen auf den durch SVs ausgelöste Abweichungen zwischen den Reads und der Referenzgenomsequenz beim Mapping basiert, versprechen NGS-Methoden des Long-read-sequencing (mit Readsequenzlängen im Kilobasenbereich), langfristig zuverlässiger zu sein. Als Beispiel sind hier die Bestrebungen der Firmen PacBio (Kalifornien, USA) [156] oder Oxford Nanopore Technologies (Oxford, UK) [157] zu nennen.

4.1.2 Anwendung zum Ausschluss anderer Ursachen der 46,XY GD

Die Forschungen für die vorliegende Arbeit konzentrierten sich neben Kopiezahlvariationen auch auf SNVs und InDels in kodierenden Bereichen, um bekannte Ursachen einer 46,XY-GD auszuschließen. Entscheidende Parameter zur Bewertung der Mutationen war unter anderem die Allelhäufigkeit in der Gesamtbevölkerung. Für exonische Mutationen liegen dafür mittlerweile umfangreiche Datensätze vor. Es ist zu bedenken, dass diese Datenbanken bewusst oder unbewusst auch Daten von kranken Patient:innen mit ähnlichem Phänotyp beeinhaltend können.[53] Dies gilt insbesondere, da in der Vergangenheit DSD-Patient:innen häufig undiagnostiziert geblieben sind. Insofern können diese Patient:innen mit krankheitsauslösenden Mutationen in die Datenbanken eingegangen sein und diese verzerren. Aufgrund der Seltenheit des Erscheinungsbildes DSD ist dieser Effekt aber vermutlich zu vernachlässigen.

Der zweite Parameter zur Einschätzung der SNVs und InDels waren Prognosen zur Auswirkung auf die Proteinstruktur.[81-83] Die Ergebnisse der verschiedenen Programme waren häufig nicht einheitlich. Denn die Berechnungen der Programme bauen vielfach nicht auf einem präzisen Verständnis der molekularen Veränderungen auf, sondern auf dem Vermögen in einem Referenzdatensatz richtige Zuordnungen herzustellen.[158] Dies führt eher zu falsch positiven Ergebnissen [159], also einer allzu schwerwiegenden Einschätzung der SNV-Auswirkung. Da diese Programme jedoch zum Ausschluss einer Krankheitsrelevanz von SNVs genutzt wurden, hat dieses Aggravieren demnach weniger Auswirkungen. Das bedeutet aber auch, dass zum aktuellen Zeitpunkt weiterhin nur funktionelle Untersuchungen im Labor eine definitive Aussage zur tatsächlichen Auswirkung der Mutationen erlauben. Darüber hinaus fehlt es an Informationen und am Verständnis, um Mutationen in nicht-kodierenden Bereichen wirklich angemessen bewerten zu können. Für exonische Bereiche sind dank der inzwischen weit verbreiteten Anwendung des WES umfangreiche Datenbanken zu Allelhäufigkeiten von Mutationen aufgebaut worden. Da die Verbreitung von WGS dem WES noch hinterherhinkt, sind die Angaben zur Häufigkeit der SNVs und InDels in nicht-exonischen Bereiche hingegen noch weniger repräsentativ. Ferner ist auch die Berechnung der funktionellen Auswirkungen von nicht-proteinkodierenden Mutationen noch schwieriger als bei exonischen. Zwar gibt es erste Ansätze, welche die funktionellen Annotation auch in diesen Bereichen ermöglichen [160, 161], doch bedarf es für eine verlässliche Bewertung eines deutlich besseren Verständnisses der regulativen Elemente und Mechanismen im menschlichen Genom.

Eine vielversprechende Stärke des WGS ist die Möglichkeit einer fortlaufenden Auswertung bei seltenen Erkrankungen, um dem stetig wachsenden Krankheitsverständnis Rechnung tragen zu können. In der Vergangenheit musste, wenn beispielsweise ein neues krankheitsrelevantes Gen entdeckt wurde, dieses mühsam bei potentiell betroffenen Patient:innen erst sequenziert werden. Dank eines WGS liegen die meisten genomischen Informationen bereits vor und es bedarf keiner weiteren Probenentnahmen und Versuche sondern lediglich eine Anpassung der Auswertung.

4.1.3 Anwendung zur Diagnosestellung bei DSD-Patient:innen

Als im Sommer 2018 die Forschungsanalysen für die vorliegende Arbeit begannen, handelte es sich um eines der ersten Projekte, das Whole Genome Sequencing (WGS) bei DSD-Patient:innen durchführte. Neben der Sequenzierung von P1 wurden auch Sequenzierungen von Patient:innen durchgeführt, bei denen es im Vorfeld keine Hinweise auf die auslösende Mutation gab. Aus diesen Untersuchungen ging Patient:in 2 hervor. Aktuell deutlich verbreiteter ist noch die Anwendung des Whole Exome Sequencing (WES) im Rahmen der molekulargenetischen Abklärung von DSD und anderen seltenen Erkrankungen.

Beim Whole **Exome** Sequencing werden jedoch ausschließlich die exonischen und teilweise die untranslatierten Bereiche (UTRs) eines Gens angereichert und sequenziert. Dafür benötigt das WES im Gegensatz zum Whole **Genome** Sequencing nur ungefähr 2% der Sequenzierungskapazität und ermöglicht die Detektion von krankheitsrelevanten Mutationen, die direkt Einfluss auf die Proteinkodierung oder Splice-sites nehmen.[53] Darauf entfallen schätzungsweise 85% der krankheitsauslösenden Mutationen.[162] 15% der krankheitsauslösenden Mutationen bleiben vermutlich unerkannt. Strukturelle Variationen können dabei, soweit die Bruchpunkte nicht in zwei Exonen liegen, gar nicht aufgedeckt werden. WES ist somit mithin unaufwändiger und schneller, aber eben auch eingeschränkt bei offenen Fragestellungen.

WGS, also Whole **Genome** Sequencing, eröffnet entscheidend größere Möglichkeiten auch für die korrekte Diagnosestellung im Zusammenhang mit DSD.[163-165] Die noch umfangreichere Datenmenge stellt Diagnostiker jedoch auch vor neue Herausforderungen, denen sich auch die vorliegende Arbeit gegenüber sah. Bei fast 100 GB generierten Daten pro

Patient:in ergaben sich dabei neben neuen Anforderungen an die digitale Infrastruktur auch höhere Anforderungen an die Auswerter:innen, um in der Masse der Daten eine krankheitsbezogene, zielgenaue Bewertung zu ermöglichen.

Im Vergleich zur klassischen Sanger-Sequenzierung sind beim WES und WGS (allgemein beim NGS) die einzelnen Sequenzlängen kürzer, aber auch die Fehlerraten etwas höher. Die Fehlerrate der Illumina HiSeq Plattform wird mit 0,1% angegeben [166], welche bioinformatisch noch auf 0,06% reduziert werden kann [167]. Eine besondere Schwierigkeit stellen hier AT- und GC-reiche Bereiche dar.[168] Bei einer Genomlänge von über 3 Milliarden Basen [169] ergeben sich hieraus potentiell noch immer über eine Million falsche Ergebnisse. Dies beeinträchtigt weniger die Sensitivität, vielmehr provoziert es eher eine erhöhte Rate an falsch positiven SNVs.[170] Dementsprechend behält die Sangersequenzierung aktuell ihre Relevanz, um mittels WGS aufgedeckte SNVs zu bestätigen [53]. WGS stellt somit noch keine Alleinlösung dar.

4.2 Beschränkungen der Erkennungsmöglichkeiten des qPCR-Screenings

Das Screening auf Kopiezahlvariationen am Xp21.2-Lokus über qPCR nutzte ein SYBR Green-System. Bei der Methode werden unselektiv alle doppelsträngigen Reaktionsprodukte angefärbt, was zu Verzerrungen der Ergebnisse führt, sollten die Primer unspezifische Amplifikate bilden. Um dem entgegenzuwirken, wurde bereits bei der Auswahl der Primer besonderes Augenmerk auf die Spezifität ihrer Bindungsstellen gelegt. Darüber hinaus wurden Nebenprodukte der Amplifikation sowie Primerdimere mittels Durchführung von Schmelzkurven-Analysen, Gelelektrophoresen und entsprechenden Effizienzwerten ausgeschlossen.

Die neu etablierte qPCR-Untersuchung demonstrierte, dass sie Kopiezahlunterschiede von *NROB1* und dessen Umgebung aufdecken kann, indem sie erstens eine neuen Triplikation bei P3 erkannte, zweitens zwei Patient:innen mit 46,XX-Chromosomensatz herausfand sowie drittens zwei bereits zuvor publizierte Fälle von *NROB1*-Duplikationen [67] blind bestätigte. Gleichzeitig gab es bei insgesamt 165 untersuchten Proben nur ein falsch positives Ergebnis. Zum Vorliegen von falsch negativen Ergebnissen kann keine Aussage getroffen werden. qPCRs

zeichnen sich jedoch durch eine hohe Sensitivität aus und können schon kleine Unterschiede in der Kopiezahl nachweisen. Hohe Sensitivität ist speziell bei der Reversen-Transkriptase-qPCR (RT-qPCR) zum Nachweis von Gentranskripten nützlich. Hier ist es prinzipiell möglich, bei entsprechender Anzahl an Replikaten, Kopiezahlunterschiede vom Faktor x1.1 aufzudecken.[171]

Eine Schwäche der qPCR-Untersuchungen bleibt jedoch die Limitation, dass bei einem Großteil der Patient:innen-Proben nur zwei Genorte am Xp21.2-Locus untersucht wurden: für *NROB1* und *CXorf21*. Daher war die Auflösung dieses Screenings bei Kopiezahlveränderungen stromaufwärts von *NROB1* gering, denn es konnten nur solche erkannt werden, die *CXorf21* einschlossen. Aus Ressourcengründen wurde in dieser Arbeit auf eine flächendeckende Anwendung aller Assays an dem Locus im Screening verzichtet. Die in der angrenzenden Umgebung von *NROB1* und *CXorf21* etablierten Assays wurden erst zur weiteren Überprüfung primär auffälliger Ergebnisse genutzt. Prinzipiell ist aber vorstellbar, das Screening nochmals in einer höheren Auflösung, d.h. mit den weiteren Assays zu wiederholen. So könnten auch Kopiezahlvariationen in der Umgebung ausgeschlossen werden. Ferner wäre auch eine Erweiterung der Einschlusskriterien zu erwägen. In dieser Arbeit wurden bewusst sehr enge Einschlusskriterien gewählt und auf die Überprüfung von Patient:innen mit einer Hypospadie oder einem Mikropenis verzichtet, um ein Kollektiv mit einem möglichst hohen Anteil an Patient:innen mit kompletten Gonadendysgenesien auszulesen. Eventuell könnten solche Fehlbildungen gleichfalls Ausdruck einer partiellen Gonadendysgenese in Folge einer Alteration von *NROB1* oder dessen Umgebung sein. Das Auslassen dieser Patient:innengruppe könnte somit die Mutationen seltener erscheinen lassen, als sie es in der Kohorte der Patient:innen mit GD tatsächlich sind.

4.3 RNA-Sequenzierung

4.3.1 Einfluss der Lagerungsdauer auf die FFPE-RNA-Sequenzierung

Die zur RNA-Sequenzierung genutzten FFPE Gonadenproben waren mehrere Jahre alt. Dies birgt sowohl die Gefahr der RNA-Degradierung, als auch der Kontamination durch Bakterien.[172] Und tatsächlich zeigten sich in den Sequenzierungen beider Proben bakterielle Transkripte. Von den 102 Millionen Reads der Patient:in 1 und 66 Millionen Reads der Patient:in 2 konnten 2,1 bzw. 1,9 Millionen dem humanen Transkriptom zugeordnet werden. Die restlichen Reads entstammten entweder bakterieller Verunreinigung oder waren

humane Transkripte, die durch die Fixierung so weit verändert waren, dass ein Mapping nicht mehr möglich war.

Das Risiko einer Kontamination oder RNA-Degradierung ist bei RNA-Sequenzierungen aus frisch gefrorenen Gewebeproben deutlich geringer, welches somit die bevorzugte Konservierung von Gewebe für diese Untersuchung darstellt. In Vergleichsuntersuchungen anhand verschiedener Tumorgewebe konnte jedoch gezeigt werden, dass die RNA-Sequenzierungen aus FFPE-Gewebe und frisch gefrorenem Gewebe vergleichbare Ergebnisse hervorbringt. Diese wurden vor der Sequenzierung zum Teil ebenfalls viele Jahre gelagert.[173-176] Es ist demnach davon auszugehen, dass die für diese Arbeit durchgeführten RNA-Sequenzierungen trotz der langen Lagerdauer belastbare Informationen lieferten und einen Vergleich mit den Ergebnissen anderer Transkriptom-Sequenzierungen zulassen.

Dennoch unterstreichen die Schwierigkeiten bei der RNA-Sequenzierung den Stellenwert von frisch gefrorenen Gewebeproben und des Biobankings bei operativen Eingriffen. Entsprechend gewonnenes Material öffnet nicht nur den Weg für zukünftige Transkriptomanalysen, sondern auch für viele andere molekulare Analysen, die sich im Verlauf der Patient:innenbetreuung als wertvolle Informationen herausstellen können.

4.3.2 Expression gonadenspezifischer Gene

In Folge der Duplikationen zeigte sich in den Transkriptomen der Gonaden der Patient:innen P1 und P2 ein intermediäres Expressionsprofil zwischen ovariellm und testikulärem Gewebe. Dies lässt sich im t-SNE (Abb. 10) nachvollziehen, bei dem ovarielle und testikuläre Proben klar voneinander getrennte Gruppen bilden, zwischen denen die Datenpunkte von P1 und P2 liegen. Auch beim Vergleich der Expression einzelner Gene (Abb. III, Anhang S. 94) lässt sich dies bestätigen. Die jeweils für Hoden bzw. Ovar charakteristischen Gene *DMRT1* und *FOXL2* werden bei beiden Patient:innen intermediär exprimiert. Auch die Expression von *NROB1* liegt nicht im Interquartilabstand des einen oder anderen Gewebes und lässt sich nicht einem von beiden klar zuordnen.

Der in dieser Arbeit postulierte Mechanismus einer unphysiologisch erhöhten *NROB1* Expression ist aus diesen Daten nicht abzulesen, weil *NROB1* ein für das erwachsene Hodengewebe

eher erniedrigtes Expressionsniveau aufweist (Abb. III, Anhang S. 94). Bei der Bewertung dieser Daten ist es aber wichtig zu bedenken, dass es sich um Analysen aus postnatalen Gonaden handelt. Bei Menschen mit einem 46,XY-Karyotyp ist in der frühen Entwicklung *NROB1* normalerweise niedrig exprimiert [29], erst später im Zuge der Spermatogenese wird es wichtig und muss hoch exprimiert sein [177]. Das lässt sich auch an den GTEx-Vergleichsdaten von erwachsenen Hoden und Ovarien erkennen (Abb. III, Anhang S. 94). Die Expression von *NROB1* liegt physiologisch im erwachsenen Hoden über der vom Ovar – in der Embryonalphase wäre dies jedoch genau umgekehrt.

Folglich braucht es mindestens zwei verschiedene Regulationsmechanismen, je nach Zeitpunkt der Entwicklung. Bei den Patient:innen 1 und 2 könnte der Kontakt mit den unphysiologischen *GK*- und *CXorf21*-Enhancern die Expression von *NROB1* während der Phase der embryonalen Gonadenentwicklung steigern, und zwar stark genug, um eine erfolgreiche Hodenentwicklung zu unterdrücken. Zu einem späteren Zeitpunkt nach der Geburt steht *NROB1* beim Mann dann unter dem Einfluss eines regulären Enhancers. Durch die strukturellen Variationen bei P1 und P2 ist aber auch das normale Regulationsumfeld gestört, wodurch die Expression von *NROB1* dann wiederum im Vergleich eher zu niedrig sein dürfte, weil die physiologischen *NROB1*-Enhancer schlechter wirken können. Die erkennbar etwas niedrigere Expression des für die Spermatogenese wichtigen *NROB1* könnte auch eine Erklärung sein, wieso die beiden Patient:innen im t-SNE eher in die Nähe der Kontrollproben mit verringerter Spermatogenese gruppiert wurden. Ferner ist zu bedenken, dass durch die frühe embryonale Fehlregulation von *NROB1* und folgende Fehlentwicklung der Gonaden sich kein funktionsfähiges testikuläres Gewebe ausbilden konnte, welches zur Spermatogenese fähig wäre.

4.3.3 Expression des MYC Signalweges

Auffällig in beiden Patient:innen (P1 und P2) ist die niedrige Expression des MYC-Signalweges. Eine Varianzanalyse der Reactome Gengruppen [178] ‚Cell Cycle, Cell Cycle Mitotic‘ und ‚Cell Cycle Checkpoints‘ errechnete zwischen den Patient:innen und ovariellen und testikulären Kontrollproben signifikante Unterschiede ($p=0,005$; $p=0,010$; $p=0,027$). Die Varianzanalyse ergab zunächst noch keine Auskunft darüber, welche Probengruppe die deutlich abweichenden Werte produzierte. Doch die mittlere Differenz zwischen den Probengruppen zeigte, dass diese auf die Expressionsdaten der Patient:innen zurückzuführen war. Somit liegt in P1 und

P2 eine charakteristisch niedrigere Expression dieser drei für den Zellzyklus wichtigen Gengruppen vor.

Somit wäre zu überlegen, ob eine solche Xp21.2-Duplikation zu einem Stopp der Entwicklung aller oder eines Teils der gonadalen Zelllinien führt. *NROB1*-knockdown-Versuche mit Small-hairpin-RNA in Gebärmutterkrebszellen zeigten, dass dies zu einer Verminderung des Zellwachstums führte. Einen Hinweis auf einen möglichen Mechanismus lieferte die begleitende Herunterregulierung von *Cyclin D1* und *c-myc*. [179] Ähnliches ließ sich bei einem *NROB1*-knockdown in Ewing-Sarkom-Zellen [180] und bei einem *NROB1*-Knockout in embryonalen Stammzellen der Maus [181] zeigen, wo diese Versuche ebenfalls zu einer Proliferationsabnahme führten. Im Gegensatz dazu wurde in Zellen des hepatozellulären Karzinoms gezeigt, dass eine verminderte *NROB1*-Expression zu einer vermehrten Proliferation führt. [182] Es scheint dennoch gut möglich, dass die geringe Ausprägung des MYC-Signalweges und der Zellproliferation auf eine veränderte *NROB1*-Wirkung zurückzuführen ist. Die Transkriptomdaten aus FFPE-Material können hier aber nur erste Hinweise liefern. Dieses Phänomen bedarf einer umfangreicheren In-vitro-Untersuchung.

4.3.4 Bedarf weiterer RNA-Sequenzierungs-Vergleichsdaten

Um die Theorie einer durch ektope *NROB1*-Expression ausgelösten 46,XY-Gonadendysgenese bei stromaufwärts von *NROB1* gelegenen Kopiezahlvariationen auf transkriptioneller Ebene einwandfrei zu überprüfen, bräuchte es Gonadenproben mit entsprechender Mutation aus der Embryonalphase. Die Entnahme ist jedoch ethisch und praktisch am Menschen nicht durchführbar. Dennoch könnte zukünftig schon ein Vergleich der Transkriptome von P1 und P2 mit anderen Patient:innen in dieser Hinsicht aufschlussreich sein. So zum Beispiel mit Gonadengewebe von Patient:innen mit Duplikationen die *NROB1* einschließen oder Patient:innen mit 46,XY-GD ohne Kopiezahlveränderungen am Xp21.2-Locus.

4.4 ChIP-Seq

4.4.1 Bewertung als mögliche TAD-Grenze des identifizierten Isolators

In ChIP-Seq (chromatin immunoprecipitation coupled with deep sequencing)-Daten [116, 117] zeigt sich eine zwischen vielen menschlichen Zelllinien konservierte CTCF, SMC3 und RAD21 Bindungsstelle im Intron 2 von *CXorf21*. Es ist die einzige dieser Bindungsstellen im

Überlappungsbereich der Duplikationen der Patient:innen 1 bis 3 und der Deletion, die von Smyk et al. [114] beschrieben wurde. Zwar bedeutet der bloße Nachweis einer CTCF-Bindung an einen DNA-Abschnitt nicht, dass dieser auch an der Ausbildung einer TAD-Grenze beteiligt ist [133], doch gibt es bestimmte Charakteristiken, die solche grenzausbildenden CTCFs häufig auszeichnen. Dazu gehört die Empfindlichkeit der Region gegenüber DNase I und eine hohe evolutionäre Konservierung der CTCF-Bindungsstelle.[183] In dem entsprechenden Bereich der CTCF-Bindungsstelle in *CXorf21* zeigt sich in ENCODE [116] bei 41 aus 125 Ziellinien eine DNase I-Hypersensitivität (Abb. VIII, Anhang S. 99).

In der Maus gibt es neben dem *NROB1* ortholog (*NrOb1*) auch ein *CXorf21* ortholog mit der Bezeichnung *5430427019Rik*. Dieses Gen hat im Gegensatz zum Menschen nur ein Intron. Während in intronischen Sequenzen zwischen Mensch und Maus gewöhnlich nur Sequenzähnlichkeiten von ungefähr 35% vorliegen [184], zeigt sich im Bereich der 25bp langen menschlichen CTCF-Bindungssequenz eine Homologie von 92%. Umgekehrt zeigt auch das CTCF Protein im Bereich seiner DNA-Bindungsdomäne eine hohe Konservierung zwischen Maus und Mensch.[185] Bei Berücksichtigung der weiteren Umgebung der intronischen Bindungssequenz ergibt sich eine Homologie von 77% zu dem entsprechenden Bereich im Mausgenom (ChrX:83,124,728 – 83,124,782; mm9) (Abb.19). Dies ist bemerkenswert für einen eigentlich nicht-kodierenden Bereich zweier Spezies, die vor 80 Millionen Jahren den letzten Vorfahren geteilt haben [186] und entspricht eher den Homologiewerten von protein-kodierenden Bereichen [187]. Dies legt eine physiologisch wichtige Funktion dieser Sequenz nahe, was es durchaus plausibel erscheinen lässt, dass diese Bindungsstelle für das Ausbilden einer TAD-Grenze und die Expressionsregulation entscheidend ist.

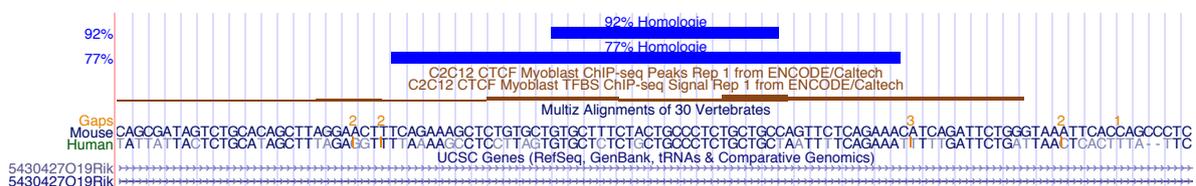


Abb. 19: Homologievergleich zwischen Maus und Menschen im Intron von *5430427019Rik* beziehungsweise *CXorf21*

Sowohl in der Maus als auch im Menschen zeigt sich in diesem Bereich eine CTCF-Bindungsstelle. Für die Maus ist dies repräsentativ durch die braun hervorgehobenen ChIP-seq-Daten aus Myoblasten dargestellt. Der Bereich mit 92% Homologie umspannt 25 Basen mit 23 übereinstimmenden Basen. Eine 77% Homologie ist durch das Übereinstimmen von 42 Basen in einem Bereich von 55 Basen gegeben.

4.4.2 Übertragbarkeit auf gonadales Gewebe

Für eine genauere Bewertung bräuchte es aber vor allem entsprechende ChIP-Seq Daten aus Hodengewebe bzw. noch besser Gonadengewebe, um zu untersuchen, ob dieses Sequenzmotiv auch in diesen Geweben durch ein CTCF gebunden ist. Die Übertragbarkeit von CTCF-ChIP-Seq-Daten anderer Gewebe auf Hodengewebe wird vor allem durch das CTCF paralog BORIS erschwert. BORIS ist das einzige um die CTCF-Bindungsstellen konkurrierende Protein und wird ausschließlich im Hodengewebe exprimiert und kann potentiell bei der TAD-Ausbildung interferieren.[188] Inwieweit dieses für die Gametogenese wichtige Protein jedoch überhaupt in den Patient:innen exprimiert wird und tatsächlich um die Bindungsstellen konkurriert bleibt unklar. Die Daten lassen auch keine Rückschlüsse zu, mit welchem entgegengesetzten CTCF die Bindungsstelle in *CXorf21* auf der anderen TAD-Seite eventuell interagiert.

4.5 Hi-C Analyse

4.5.1 Einteilung in konstitutive und fakultative TADs

Mit den im PennState 3D Genome Browser [106] zur Verfügung gestellten Hi-C-Ergebnissen ließ sich die räumliche Organisation und Interaktionen am Xp21.2-Locus in verschiedenen humanen Zelllinien darstellen. In einem Großteil der humanen Zelllinien konnte eine TAD-Grenze im Bereich des zuvor ermittelten Isolators in *CXorf21* aufgedeckt werden. Vielfach ist beschrieben, dass TADs und TAD-Grenzen zwischen Zelllinien und sogar zwischen verschiedenen Säugetieren konserviert sein können.[126, 189-191] Daraus ergibt sich eine mögliche Einteilung in konstitutive und fakultative TADs [192], welche entweder konstante oder über verschiedene Zelllinien und die zeitliche Entwicklung hinweg flexible Organisationsstrukturen darstellen.[193] Im Falle einer konstitutiven TAD-Grenze am Xp21.2-Locus im *CXorf21*-Gen wäre anzunehmen, dass diese auch im Hoden- bzw. Gonadengewebe vorliegt. Diese Einteilung als konstitutive TAD-Grenze lässt sich anhand der Hi-C-Daten aus den verschiedenen Zelllinien aber nicht zweifelsfrei vornehmen. Denn es zeigt sich nicht in allen Zelllinien eine TAD-Grenze in diesem Bereich. Der Umstand, dass fünf der 36 Zelllinien (IMR90/Lungenfibroblasten, Spleen/Milz, SJCRH30/Rhabdomyosarkom, K562/Leukozyten, Aorta) keine Grenze aufweisen, kann aber auch auf methodische Ungenauigkeiten zurückzuführen sein.

4.5.2 Schwankungsanfälligkeit von Hi-C-Daten und TAD-Bereichen

Eine der fünf Zelllinien, in denen zwischen *NROB1* und *GK* im 3D Genome Browser keine TAD Grenze vermerkt ist, ist beispielsweise IMR90 (Abb. 16). In dieser Zelllinie wird die TAD-Grenze jenseits von *GK* verortet. Wenn man hingegen Hi-C-Daten zur Zelllinie IMR90 im ‚Chorogenome Navigator‘ [128] darstellt, zeigt sich sehr wohl eine TAD-Grenze in der Nähe von *CXorf21*. Und das, obwohl beide Plattformen auf den gleichen Ursprungsdatensatz von Rao et al. [191] zurückgreifen. Diese Diskrepanz kann unter anderem auf unterschiedliche Algorithmen und deren Unschärfen zurückzuführen sein, die zur Identifikation der TAD-Grenzen verwendet wurden. Für die Berechnung von TADs stehen verschiedene Programme zur Verfügung, die jedoch bei gleichen zugrundeliegenden Daten zum Teil sehr unterschiedliche Vorhersagen zur Anzahl und Größe der Domänen machen.[194] Um dieses Problem der divergierenden Prognosen zu lösen, braucht es in Zukunft vor allem ein besseres biologisches und mathematisches Verständnis, was ein TAD ist, um die Programme weiter verbessern zu können und einen größeren Konsens in den Ergebnissen zu generieren. Bei den ebenfalls im Chorogenome Navigator abrufbaren Zelllinien GM12878 und K562 zeigte sich, wie im 3D Genome Browser, eine TAD-Trennung im Bereich zwischen *NROB1* und *CXorf21*. Die Positionen der Grenzen stimmen aber ebenfalls nicht genau überein.

Beim Vergleich von Hi-C-Daten ist auch zu bedenken, dass es aktuell noch keine einheitlichen Standards zur Qualitätskontrolle gibt.[195] Darüber hinaus lässt sich beobachten, dass es methodisch bedingte Schwankungen der Ergebnisse gibt, die teilweise zu einer geringen Reproduzierbarkeit führen. [196] Es ist auch zu bedenken, dass es sich bei der Ausbildung der TADs um einen sehr dynamischen Prozess handelt und das Hi-C somit eine Momentaufnahme der Chromatinkonformation darstellt [197]. Die Konformation einer einzelnen Zelle kann beispielsweise durch Prozesse wie die Mitose stark verzerrt werden.[198] Da außer beim Single-cell-Verfahren beim Hi-C viele Zellen eines Gewebes auf einmal untersucht werden, ist das Ergebnis der Analyse vermutlich weniger Ausdruck einer stabilen Ausprägung in allen Zellen als vielmehr eine Art Durchschnittswert der gepoolten Zellgruppe.

All diese Faktoren, die zu einer Schwankung beitragen können, machen es dementsprechend schwierig, anhand isolierter Hi-C-Daten eindeutige Schlüsse zu ziehen. Umso wichtiger ist die Zusammenschau mit ChIP-Seq-, Genom- oder Transkriptomdaten sowie der Abgleich mit etwaigen Vergleichs-Hi-C-Daten, etwa aus dem 3D Genome Browser.

4.5.3 Übertragbarkeit auf Patient:innenproben

Für einen definitiven Nachweis der alterierten TAD-Struktur im Gonadengewebe durch die Xp21.2-Kopiezahlvariationen bräuchte es eine Analyse der Chromatinkonformation aus entsprechendem Patient:innengewebe, das die fraglichen strukturellen Variationen aufweist. Bei den Patient:innen müssten, gemäß Theorie, dann mehr Interaktionen über die ehemalige TAD-Grenze hinaus vorliegen. Es entsteht somit ein größeres TAD. Neben der Größe wären aber speziell auch die genauen Kontakte innerhalb dieses TADs von großem Interesse. TADs können nochmals in Sub-Strukturen unterteilt werden, sogenannte Sub-TADs.[199] Diese stellen die eigentlichen regulativen Kontakte innerhalb einer Domäne her, wie Promoter-Enhancer-Interaktionen.[200] Diese sind weniger konserviert und zeigen größere zellabhängige Schwankungen [199, 201]. Aufgrund der geringen Auflösung können diese Kontakte in aller Regel jedoch nicht vollständig durch Hi-C erfasst werden und bedürfen zu ihrer Detektion anderer Methoden wie dem konventionellen Chromosome conformation capture (3C) [105].[195] Eine hochauflösende Untersuchung der DNA-Faltung in nur einem ausgewählten Bereich würde es erlauben, tatsächliche Interaktionen zwischen den *CXorf21*- und *GK*-Enhancern sowie dem *NROB1*-Promotor offenzulegen, bzw. das Interaktionsprofil der Patient:innen an dem Genort mit der normalen Regulationsarchitektur in Kontrollproben zu vergleichen.

4.6 GeneHancer

4.6.1 Verlässlichkeit putativer Enhancer-Elemente

Die postulierten Enhancer im Bereich der Gene *CXorf21* und *GK* beruhen auf Ergebnissen der GeneHancer-Datenbank. Die genaue Identifikation von Enhancern stellt eine Herausforderung dar und basiert häufig auf der Identifikation von Gruppen konservierter Sequenzen [202], Histonmodifikationen und DNase-Hypersensitivität [203] oder Transkripten von Enhancer-RNA (eRNA) [204].[135] Eine große Stärke von GeneHancer ist, dass dabei Information aus vier verschiedenen Quellen [117, 203, 205, 206] gebündelt werden, was die Verlässlichkeit der Vorhersagen erhöht.[135] Dadurch sind in der Datenbank insgesamt ca. 285 000 Enhancer-Kandidaten gesammelt, die in Ränge unterteilt werden, abhängig davon, in wie vielen Quellen sie aufgeführt sind. Enhancer, die in mindestens zwei Datensätzen vertreten sind, werden als ‚Elite‘-Enhancer bezeichnet und stellen mit höherer Wahrscheinlichkeit einen tatsächlichen Enhancer dar. Wenn diese Enhancer zusätzlich mit verschiedenen Methoden übereinstimmende Zuordnungen aufweisen, auf welches Gen sie wirken, dann werden sie als ‚Double-

elite-Enhancers' bezeichnet. Von diesen Double-elite-Enhancers gibt es insgesamt ca. 30.000 und eine Fokussierung auf genau diese Enhancer entfernt einen erheblichen Anteil an Rauschen/Ungenauigkeit. Bei allen 14 in Abb. 17 dargestellten Enhancern handelt es sich um Double-elite-Enhancer.

Bisher gibt es nur wenige experimentell validierte Enhancer [207] und die Daten bleiben putativ. So bestehen zwei Ebenen der Unsicherheit: Zum einen, ob es sich bei den dargestellten Bereichen tatsächlich um Enhancer handelt und zum anderen, ob die vorgenommene Zuordnung von Enhancer und Zielgen korrekt ist. Die exakte Zuordnung kann sehr schwer sein, weil Enhancer zum Teil Kilobasen voneinander entfernt liegen können und durch entsprechende DNA-Faltung auch über dazwischenliegende Gene hinweg wirken können.[208, 209] Da es in dieser Region eine beträchtliche Anzahl an vorhergesagten Enhancern gibt, die außerdem eine verhältnismäßig hohe Verlässlichkeit in ihrer Vorhersage aufweisen, kann man davon ausgehen, dass tatsächlich jenseits der TAD-Grenze in *CXorf21* mehrere Enhancer lokalisiert sind. Die genaue Zuordnung ist zwar nicht sicher, aber von nachrangiger Bedeutung, denn Enhancer können in aller Regel nicht über TAD-Grenzen hinwegwirken [119, 210]. Und da der Bereich von *CXorf21* und *GK* mit den Enhancern in beide Richtungen von TAD-Grenzen (Abb. 16) flankiert ist, scheint die Zuordnung von Enhancer und Genen sehr plausibel; nicht zuletzt auch in Ermanglung an erkennbaren alternativen Zuordnungsmöglichkeiten.

4.6.2 Enhancerwirkungen sind abhängig von vorliegenden Transkriptionsfaktoren

Was aus den Daten nicht hervorgeht, sind die an die Enhancer bindenden Transkriptionsfaktoren (TF), welche für die Rekrutierung der RNA-Polymerase II an dem Promotor erforderlich sind.[211] Ein Schlüsselement der Regulation ist dabei, dass die Enhancer nicht durchgehend von TFs besetzt sind und im zeitlichen Verlauf von wechselnden TFs gebunden werden können.[212, 213] Für die Rekrutierung der RNA-Polymerase II müssen überdies mehrere an Enhancer gebundene Transkriptionsfaktoren gemeinsam auf einen Promoter wirken.[211]

Es ist somit wichtig zu bedenken, dass das bloße Vorliegen eines Enhancers und eines Gens in einem TAD und ggf. auch deren räumliche Nähe zu einem sub-TAD dementsprechend nicht gleichbedeutend damit sind, dass es zwangsläufig zu einer transkriptionellen Aktivierung kommt. Daher wäre denkbar, dass die *GK*- und *CXorf21*-Enhancer nur zu einem bestimmten

Zeitpunkt auf *NROB1* wirken können, wenn eine entsprechende begünstigende Konstellation an Transkriptionsfaktoren exprimiert wird.

4.7 Schlussfolgerung

Die Ergebnisse der vorliegenden Arbeit verschieben das Augenmerk bei der Bewertung von Xp21.2-Kopiezahlvariationen weg von *NROB1* selbst, hin zu einem 35kb-Bereich stromaufwärts von *NROB1* und einem sich daraus ergebenden gemeinsamen Modell der 46,XY-GD Pathogenese (Abb. 18). Bemerkenswerterweise ist dieser Bereich nicht nur in den drei Kopiezahlvermehrungen der in dieser Arbeit vorgestellten Patient:innen 1-3 und der Deletion von Smyk et al. [114] zu finden, sondern ist zusätzlich auch als einziger Bereich in allen publizierten, mit 46,XY-GD assoziierten, Xp21.2-Duplikationen vorhanden (Abb. 20).[64-67, 214-216] Innerhalb des 35kb Fragmentes liegt wiederum *Cxorf21*, wodurch es prinzipiell zu einem Kandidaten-Gen als Auslöser der GD avanciert. Im Gegensatz zu *NROB1* gibt es aber keine Erkenntnisse, die einen aktiven Einfluss von *CXorf21* auf die Geschlechtsentwicklung nahelegen. Stattdessen fand sich in dem Intron von *CXorf21* ein Isolator, dessen abschirmende Wirkung von Enhancern durch strukturelle Variationen umgangen werden kann (Abb. 18). Die mögliche Folge dürfte eine veränderte *NROB1* Expression in der Embryonalphase sein mit der Konsequenz einer verminderten SF1-vermittelten *SOX9*-Expression und einer daraufhin beeinträchtigten Hodenentwicklung. Damit im Einklang steht, dass bei allen drei Patient:innen, sowie der von Smyk et al. und weiteren Autoren publizierten SVs, diese einen maternalen Ursprung hatten und sich bei den XX-Träger:innen kein offensichtlicher Phänotyp präsentierte [65, 114, 214].

Weil der Überlappungsbereich in allen bekannten Fällen vorliegt, ist der postulierte Mechanismus auch auf alle anderen Fälle zu übertragen. Die bisher vorherrschende Erklärung des Pathomechanismus sollte mithin überdacht werden. Denn in der Vergangenheit wurde der Pathomechanismus schlicht mit dem doppelten Vorliegen des *NROB1*-Gens erklärt. Diese Erklärung würde damit aber im Widerspruch zu experimentellen Ergebnissen aus der Maus stehen.[217] In transgenen Mäusen konnte gezeigt werden, dass eine bloße doppelt erhöhte Expression von *NROB1* nicht ausreicht, um einen weiblichen Phänotyp bei 46,XY-Mäusen hervorzurufen. Stattdessen kommt es erst bei einer vielfach erhöhten Expression von *NROB1* zu einer Geschlechtsumkehr der Mäuse.[217] Dies spricht dafür, dass eine zusätzliche Kopie

allein nicht ausreicht, sondern diese Kopie zusätzlich einer deutlichen Fehlregulation unterliegen muss, um die Dysgenese auszulösen – wie dies beispielsweise durch die Disruption topologisch assoziierender Domänen der Fall sein kann. Eine Übertragbarkeit von Ergebnissen aus der Maus auf den Menschen ist im Bereich der Geschlechtsentwicklung freilich nur bedingt gegeben. Ein Beispiel ist das *SF1*-Gen, dessen heterozygote Mutationen im Menschen zu einer 46,XY-GD führen [218-220], in der Maus hingegen zu gesunden und fertilen 46,XY-Nachkommen [221]. Ein pathologischer Phänotyp erscheint erst in Mäusen mit homozygoten *SF1*-Mutationen.[221] Auch bei Mutationen von *NROB1* zeigen sich zwischen Maus und Mensch zum Teil deutliche Differenzen im Phänotyp, abhängig vom genetischen Hintergrund der Mäuse. So können im Mausstamm C57BL/6J Teildeletionen von *NrOb1* bei 46,XY-Chromosomensatz einen weiblichen Phänotyp verursachen.[222, 223] 46,XY-*NrOb1*-Knockouts in DBA/2J- und B6xD2-Stämmen entwickeln sich hingegen meist männlich, sind aber infertil.[177, 224]. Letztere beiden Stämme ähneln damit dem beim Menschen beschriebenen Erscheinungsbild. Patient:innen mit *NROB1*-Mutationen bei 46,XY-Karyotyp zeigen nach der Geburt in der Regel keine Geschlechtsauffälligkeiten, leiden später aber an Infertilität.[225, 226]

Durch die Identifikation anderer Fälle sowie die weitere Untersuchung der 35kb großen Überlappungsregion kann die hier postulierte neue Erklärungsweise überprüft und ggf. ausgebaut werden. Häufig stehen jedoch primär SNVs und InDels im Vordergrund molekulargenetischer Untersuchungen bei isolierten genetischen Erkrankungen. Diese momentane Fokussierung dürfte daran liegen, dass sich WES in der Diagnostik zunehmender Beliebtheit erfreut, was zu einer Konzentration speziell auf Protein-Kodierung führt. Und selbst bei der Auswertung von WGS-Daten erhalten SNVs häufig eine größere Aufmerksamkeit als SVs, weil diese direkt aus der Sequenz ablesbar sind und keine weitere Daten-Prozessierung erfordern.[57] So ist die Suche nach SVs bei syndromalen Krankheitsbildern deutlich verbreiteter. Dadurch werden auch Fälle von großen, syndromalen Duplikationen entdeckt, die *NROB1* einschließen.[67, 227-230] Bleibt nur zu hoffen, dass mit zunehmender Erfahrung im Umgang mit NGS-Verfahren auch bald strukturelle Variationen eine größere Aufmerksamkeit erlangen.

Mit einer flächendeckenden Anwendung von WGS und dem Aufbau größerer Datenbanken wird es interessant sein zu sehen, ob in Zukunft Xp21.2-Duplikationen – mit und ohne *NROB1*-Beteiligung – in 46,XY-Patient:innen ohne Gonadendysgenese entdeckt werden können. Entsprechend der hier vorgestellten Wirkungsweise ist jedoch nicht die Duplikation allein

entscheidend, sondern die Frage, wo genau ein dupliziertes Segment eingebaut wurde. Beispielsweise könnte eine Duplikation ohne *NROB1*, die jedoch den Isolatorbereich in *CXorf21* beinhaltet, weit entfernt vom ursprünglichen Xp21.2-Locus inseriert sein und dadurch keine Veränderung der *NROB1*-TAD hervorrufen. Ähnlich könnte eine Duplikation inklusive *NROB1* in eine regulative Wüste eingebaut werden, in der diese zusätzliche Kopie ohne Funktion und Möglichkeit einer ektopten Expression bleibt. Beide SVs könnten in konventionellen Verfahren wie dem aCGH oder MLPA analog als eine pathogene Mutation erscheinen, ohne aber den Phänotyp überhaupt zu beeinflussen.

Die Ergebnisse dieser Arbeit zeigen, wie wichtig es ist, in der molekulargenetischen Diagnostik den Blick nicht nur auf die gängigen Kandidatengene zu richten, sondern auch den nicht-kodierenden Bereich genau zu untersuchen. Dies ermöglicht, neue Ursachen genetischer Erkrankungen zu entdecken und – wie im Fall von DSD – darüber hinaus ein besseres Verständnis grundlegender Entwicklungsprozesse zu erlangen, wie etwa bei der menschlichen Geschlechtsentwicklung.

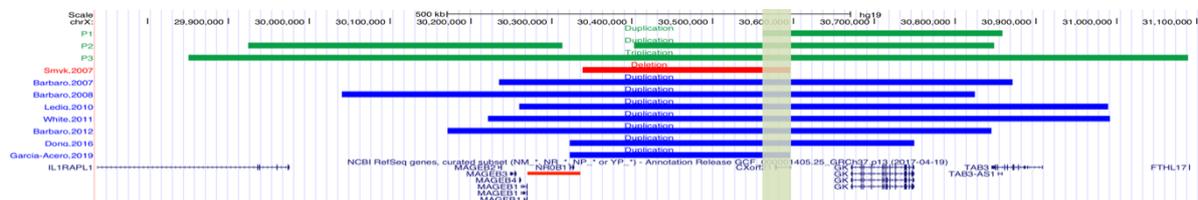


Abb. 20: Gegenüberstellung bisher bekannter Xp21.2-Kopiezahlvariationen

Die grünen Balken stellen die Duplikationen und Triplikation der Patient:innen 1 bis 3 aus dieser Arbeit dar. Die von Smyk et al. [114] publizierte Deletion ist mit einem roten Balken dargestellt, wohingegen mit 46,XY-GD assoziierte Duplikationen [64-67, 214-216] durch blauen Balken wiedergegeben werden. Der grün schattierte Bereich hebt den Überlappungsbereich aller bekannten Xp21.2-Kopiezahlveraiationen hervor, in dem *CXorf21* liegt. *NROB1* ist rot unterstrichen hervorgehoben.

4.8 Ausblick

Zur weiteren Überprüfung des in dieser Arbeit vorgestellten Krankheitsmechanismus ergeben sich künftig verschiedene Wege. Erstens kann über die Identifikation eines homologen Sequenzbereichs zwischen Maus und Mensch im *CXorf21*-Intron die abschirmende Wirkung des dort gelegenen CTCFs in der Maus mit dem CRISPR-Cas9-System [231] überprüft werden. Zu diesem Zweck entstand eine Kooperation mit dem MRC Harwell Institute, Harwell /UK, um eine Mauslinie mit deletiertem Isolatorbereich zu generieren. Sollten sich diese Mäuse trotz 46,XY-Karyotyp mit weiblichem Phänotyp präsentieren, wäre dies ein gewichtiger Hinweis zur Stützung dieser Theorie. Hierdurch wären weitere Versuche zur Ermittlung der Chromatin-Konformation, der Bindung von Transkriptionsfaktoren und von Promotor-Enhancer-Interaktionen möglich. Die Schwierigkeiten, Ergebnisse aus der Maus auf den Menschen zu übertragen, wurden allerdings oben bereits diskutiert.

Zweitens: Von den Patient:innen 1 und 2 liegen FFPE-Proben der postnatalen Gonaden vor, die in dieser Arbeit zur Transkriptomanalyse verwendet wurden. Eine kürzlich neu vorgestellte Methode, genannt Fix-C, verspricht Hi-C-Untersuchungen aus FFPE-Material möglich zu machen.[232] Nach Auskunft der Firma Dovetail Genomics (Kalifornien, USA), die die Methode vertreibt, ist die Qualität der Technik zum aktuellen Zeitpunkt aber noch nicht für das zuverlässige Ermitteln von TADs geeignet. Langfristig wird jedoch erwartet, dass es zuverlässig möglich sein wird, TADs zu erkennen (Korrespondenz April 2019 und März 2020).

Im Rahmen einer Gonadektomie von Patient:in 3 am UKSH konnten gonadale Gewebeproben gewonnen und im Interdisziplinären Centrum für Biobanking Lübeck eingefroren werden. Aus diesem Material wären sowohl Transkriptom, konventionelle Hi-C, als auch CHIP-Seq-Untersuchungen möglich, die im Rahmen weiterer Forschungsprojekte und Kooperationen durchgeführt werden könnten.

Bei den Gonadengewebeproben, ob FFPE oder gefroren, stellt sich jedoch das bereits diskutierte Problem, dass diese Proben lange nach den kritischen Phasen der embryonalen Geschlechtsentwicklung entnommen wurden. Folglich können sie kein genaues Abbild eventuell pränataler Fehlregulationen geben. Zu überlegen wären daher Experimente mit induzierten pluripotenten Stammzellen aus humanen Fibroblasten. Diese Fibroblasten

könnten entweder von Patient:innen stammen, die die entsprechenden SVs tragen oder von gesunden Probanden, deren Zellen dann mittels CRISPR-Cas9-Mutationen am Xp21.2-Locus eingebracht werden. In den letzten Jahren konnten mehrere Protokolle für die Generierung von Gonaden und Hodenzellen aus IPS ermittelt werden. [233-236]

In einem ersten Schritt konnten von P3 und ihrer Mutter Fibroblasten sowie Lymphozyten gewonnen und in Kultur genommen werden. Von P1 konnten ebenfalls Lymphozyten gewonnen werden. Im Rahmen einer Kooperation mit dem Institut für Humangenetik des UKSH, Lübeck und dem Institut für Medizinische Genetik an der Charité, Berlin sollen aus diesen Zellkulturen erste orientierende Hi-C-Daten erhoben werden. Prinzipiell könnte einer Hi-C-Analyse der mütterlichen Zellen die Inaktivierung des X-Chromosoms und die damit verbundene dreidimensionale Umstrukturierung im Wege stehen. Von den *NROB1*-Nachbargenen *IL1RAPL1* [237] und *CXorf21* [238] ist jedoch bekannt, dass sie der X-Inaktivierung entgehen können. Dementsprechend ist es vorstellbar, dass selbiges auch für *NROB1* gilt und die Analyse davon unbeeinflusst bleibt.

Für die klinische Praxis wird es wichtig sein, weitere Fälle von Xp21.2-Kopiezahlvariationen zu identifizieren, um den Patient:innen die Ungewissheit einer unklaren Diagnose nehmen zu können. Und durch steigende Fallzahlen ließe sich das Verständnis dieses Erkrankungsbildes und eventueller Komorbiditäten weiter ausbauen und damit hoffentlich auch die Therapiemöglichkeiten verbessern.

5. Zusammenfassung

Duplikationen am Xp21.2-Lokus sind eine bekannte Ursache von 46,XY-Gonadendysgenesien. Die pathologische Wirkung dieser Duplikationen wurde bisher durch das doppelte Vorliegen des Gens *NROB1* erklärt. Die vorliegende Arbeit hat nun untersucht, ob auch Xp21.2-Duplikationen, die *NROB1* aussparen, zu einer Gonadendysgenese (GD) führen können. Durch die genaue genetische Charakterisierung einer initial entdeckten Patient:in mit Duplikation stromaufwärts von *NROB1*, sowie der Identifikation weiterer Patient:innen, sollten Hinweise gefunden werden, ob der Pathomechanismus einer Xp21.2-Duplikationen wirklich zwingend an die Duplikation von *NROB1* gekoppelt ist. Die Suche nach weiteren Patient:innen erfolgte sowohl über Whole Genome Sequencing (WGS), als auch einen eigens dafür etablierten qPCR-Screening und wurde über ArrayCGH bestätigt. Durch das WGS konnten andere bekannte Ursachen der GD ausgeschlossen sowie die exakte Größe und Anordnung der Kopiezahlvariationen ermittelt werden. Die Anordnungen ließen sich durch Sanger-Sequenzierungen verifizieren.

Insgesamt wurden drei nicht-verwandte Patient:innen identifiziert, bei denen Xp21.2-Kopiezahlvariationen die wahrscheinlichste Ursache ihrer 46,XY-Gonadendysgenese darstellen. Zwei wiesen überlappende Duplikationen stromaufwärts von *NROB1* auf, während sich bei einer dritten Patient:in eine große *NROB1* umspannende Triplikation präsentierte. WGS und Sanger-Sequenzierungen zeigten dabei jeweils eine Insertion der Kopiezahlfragmente in der Nähe von *NROB1*. Um die Wirkung des bisher unbekanntes Mutationstyps stromaufwärts von *NROB1* weiter zu untersuchen, wurde das Transkriptom der Gonaden zweier betroffenen Patient:innen sequenziert. Es zeigte sich ein intermediäres Expressionsverhalten mit ovariellen und testikulären Eigenschaften sowie u.a. stark reduzierter Expression von Genen der Spermatogenese. Der Vergleich der drei Patient:innen mit bereits publizierten Fällen von 46,XY-Gonadendysgenese und Kopiezahlveränderungen am Xp21.2-Lokus ergab einen 35kb großen, allen Proben gemeinsamen Bereich stromaufwärts von *NROB1*.

Bei In-silico-Analysen fand sich in dem 35kb-Bereich ein Isolator, der *NROB1* regulativ abschirmt, aber durch strukturelle Variationen ggf. umgangen wird. So könnten eine anormale *NROB1*-Expression mit dadurch reduzierter *SOX9*-Expression in beeinträchtigter Hodenentwicklung und Gonadendysgenese bei den Patient:innen resultieren, selbst wenn keine *NROB1*-Duplikation vorliegt. Diese Arbeit verschiebt somit das Augenmerk bei der Bewertung von Xp21.2-Kopiezahlvariationen weg von *NROB1* und hin zu einem stromaufwärts gelegenen Bereich und bietet einen neuen Erklärungsansatz für die Wirkweise der Kopiezahlvariationen.

6. Literaturverzeichnis

1. Eggers, S. und A. Sinclair, *Mammalian sex determination-insights from humans and mice*. Chromosome Res, 2012. **20**(1): p. 215-38.
2. Biason-Lauber, A., *Human Sex Development: from Basic Science to Clinical Practice and Back*. *Pediatr Endocrinol Rev*, 2017. **15**(1): p. 8-20.
3. Piprek, R.P., M. Kloc, und J.Z. Kubiak, *Early Development of the Gonads: Origin and Differentiation of the Somatic Cells of the Genital Ridges*. *Results Probl Cell Differ*, 2016. **58**: p. 1-22.
4. Wilhelm, D., S. Palmer, und P. Koopman, *Sex determination and gonadal development in mammals*. *Physiol Rev*, 2007. **87**(1): p. 1-28.
5. Müller, W. und M. Hassel, *Entwicklungsbiologie und Reproduktionsbiologie des Menschen und bedeutender Modellorganismen*. 6. Auflage ed. 2018, Berlin: Springer Spektrum.
6. Sinclair, A.H., et al., *A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif*. *Nature*, 1990. **346**(6281): p. 240-4.
7. Biason-Lauber, A. und M.C. Chaboissier, *Ovarian development and disease: The known and the unexpected*. *Semin Cell Dev Biol*, 2015. **45**: p. 59-67.
8. Wolpert, L., et al., *Principles of Development: Das Original mit Übersetzungshilfen*. 3. Auflage ed. 2007, Berlin: Springer Spektrum.
9. Hiort, O. *Typical sex development*. 2021 19.01.2021; Verfügbar unter: https://www.uptodate.com/contents/typical-sex-development?source=mostViewed_widget (Zuletzt aufgerufen: 06.08.2021).
10. Hughes, I.A., et al., *Consensus statement on management of intersex disorders*. *Arch Dis Child*, 2006. **91**(7): p. 554-63.
11. *Stellungnahme Intersexualität des Deutschen Ethikrat*. 2012; Verfügbar unter: https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/DER_StnIntersex_Deu_Online.pdf (Zuletzt aufgerufen: 23.07.2021).
12. *Gesetz zur Änderung der in das Geburtenregister einzutragenden Angaben* Bundesgesetzblatt Jahrgang 2018 Teil I 2018; Verfügbar unter: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&start=//*/%5B@attr_id=%27bgbl118s2635.pdf%27%5D/#_bgbl_%2F%2F*%5B%40attr_id%3D%27bgbl118s2635.pdf%27%5D_1589112173716 (Zuletzt aufgerufen: 23.07.2021).
13. Nordenvall, A.S., et al., *Population based nationwide study of hypospadias in Sweden, 1973 to 2009: incidence and risk factors*. *J Urol*, 2014. **191**(3): p. 783-9.
14. Lee, P.A., et al., *Global Disorders of Sex Development Update since 2006: Perceptions, Approach and Care*. *Horm Res Paediatr*, 2016. **85**(3): p. 158-80.
15. Sax, L., *How common is intersex? a response to Anne Fausto-Sterling*. *J Sex Res*, 2002. **39**(3): p. 174-8.
16. Thyen, U., et al., *Epidemiology and initial management of ambiguous genitalia at birth in Germany*. *Horm Res*, 2006. **66**(4): p. 195-203.
17. Bashamboo, A. und K. McElreavey, *Mechanism of Sex Determination in Humans: Insights from Disorders of Sex Development*. *Sex Dev*, 2016. **10**(5-6): p. 313-325.
18. Hiort, O., et al., *Management of disorders of sex development*. *Nat Rev Endocrinol*, 2014. **10**(9): p. 520-9.
19. Cools, M., et al., *Caring for individuals with a difference of sex development (DSD): a Consensus Statement*. *Nat Rev Endocrinol*, 2018. **14**(7): p. 415-429.

20. Speiser, P.W., et al., *Congenital Adrenal Hyperplasia Due to Steroid 21-Hydroxylase Deficiency: An Endocrine Society Clinical Practice Guideline*. J Clin Endocrinol Metab, 2018. **103**(11): p. 4043-4088.
21. Eggers, S., et al., *Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort*. Genome Biol, 2016. **17**(1): p. 243.
22. Arboleda, V.A., D.E. Sandberg, und E. Vilain, *DSDs: genetics, underlying pathologies and psychosexual differentiation*. Nat Rev Endocrinol, 2014. **10**(10): p. 603-15.
23. Ito, M., R. Yu, und J.L. Jameson, *DAX-1 inhibits SF-1-mediated transactivation via a carboxy-terminal domain that is deleted in adrenal hypoplasia congenita*. Mol Cell Biol, 1997. **17**(3): p. 1476-83.
24. Lala, D.S., et al., *Activation of the orphan nuclear receptor steroidogenic factor 1 by oxysterols*. Proc Natl Acad Sci U S A, 1997. **94**(10): p. 4895-900.
25. Chaboissier, M.C., et al., *Functional analysis of Sox8 and Sox9 during sex determination in the mouse*. Development, 2004. **131**(9): p. 1891-901.
26. Eggers, S., T. Ohnesorg, und A. Sinclair, *Genetic regulation of mammalian gonad development*. Nat Rev Endocrinol, 2014. **10**(11): p. 673-83.
27. Ludbrook, L.M., et al., *Excess DAX1 leads to XY ovotesticular disorder of sex development (DSD) in mice by inhibiting steroidogenic factor-1 (SF1) activation of the testis enhancer of SRY-box-9 (Sox9)*. Endocrinology, 2012. **153**(4): p. 1948-58.
28. Sekido, R., et al., *SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors*. Dev Biol, 2004. **274**(2): p. 271-9.
29. Hanley, N.A., et al., *SRY, SOX9, and DAX1 expression patterns during human sex determination and gonadal development*. Mech Dev, 2000. **91**(1-2): p. 403-7.
30. Sekido, R. und R. Lovell-Badge, *Sex determination and SRY: down to a wink and a nudge?* Trends Genet, 2009. **25**(1): p. 19-29.
31. Arango, N.A., R. Lovell-Badge, und R.R. Behringer, *Targeted mutagenesis of the endogenous mouse Mis gene promoter: in vivo definition of genetic pathways of vertebrate sexual development*. Cell, 1999. **99**(4): p. 409-19.
32. Wilson, M.J., et al., *The transcription factors steroidogenic factor-1 and SOX9 regulate expression of Vanin-1 during mouse testis development*. J Biol Chem, 2005. **280**(7): p. 5917-23.
33. Bradford, S.T., et al., *The cerebellin 4 precursor gene is a direct target of SRY and SOX9 in mice*. Biol Reprod, 2009. **80**(6): p. 1178-88.
34. Wilhelm, D., et al., *SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development*. J Biol Chem, 2007. **282**(14): p. 10553-60.
35. Kim, Y., et al., *Fgf9 and Wnt4 act as antagonistic signals to regulate mammalian sex determination*. PLoS Biol, 2006. **4**(6): p. e187.
36. Moniot, B., et al., *The PGD2 pathway, independently of FGF9, amplifies SOX9 activity in Sertoli cells during male sexual differentiation*. Development, 2009. **136**(11): p. 1813-21.
37. Gonen, N., et al., *Sex reversal following deletion of a single distal enhancer of Sox9*. Science, 2018. **360**(6396): p. 1469-1473.
38. Nachtigal, M.W., et al., *Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression*. Cell, 1998. **93**(3): p. 445-54.
39. Kidokoro, T., et al., *Influence on spatiotemporal patterns of a male-specific Sox9 activation by ectopic Sry expression during early phases of testis differentiation in mice*. Dev Biol, 2005. **278**(2): p. 511-25.

40. Bashamboo, A., et al., *Anomalies in human sex determination provide unique insights into the complex genetic interactions of early gonad development*. Clin Genet, 2017. **91**(2): p. 143-156.
41. Rahmoun, M., et al., *In mammalian foetal testes, SOX9 regulates expression of its target genes by binding to genomic regions with conserved signatures*. Nucleic Acids Res, 2017. **45**(12): p. 7191-7211.
42. She, Z.Y. und W.X. Yang, *Molecular mechanisms involved in mammalian primary sex determination*. J Mol Endocrinol, 2014. **53**(1): p. R21-37.
43. Sekido, R. und R. Lovell-Badge, *Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer*. Nature, 2008. **453**(7197): p. 930-4.
44. Griffiths, A.J.F., J.H. Miller, und D.T. Suzuki, *An Introduction to Genetic Analysis*. 7th Edition. 2000, New York: W. H. Freeman.
45. Eilbeck, K., et al., *The Sequence Ontology: a tool for the unification of genome annotations*. Genome Biol, 2005. **6**(5): p. R44.
46. Scacheri, C.A. und P.C. Scacheri, *Mutations in the noncoding genome*. Curr Opin Pediatr, 2015. **27**(6): p. 659-64.
47. Lappalainen, T., et al., *Genomic Analysis in the Age of Human Genome Sequencing*. Cell, 2019. **177**(1): p. 70-84.
48. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing*. Nature, 2010. **467**(7319): p. 1061-73.
49. Caspar, S.M., et al., *Clinical sequencing: From raw data to diagnosis with lifetime value*. Clin Genet, 2018. **93**(3): p. 508-519.
50. Sun, H. und G. Yu, *New insights into the pathogenicity of non-synonymous variants through multi-level analysis*. Sci Rep, 2019. **9**(1): p. 1667.
51. Kimchi-Sarfaty, C., et al., *A "silent" polymorphism in the MDR1 gene changes substrate specificity*. Science, 2007. **315**(5811): p. 525-8.
52. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans*. Nature, 2016. **536**(7616): p. 285-91.
53. Petersen, B.S., et al., *Opportunities and challenges of whole-genome and -exome sequencing*. BMC Genet, 2017. **18**(1): p. 14.
54. Audi, L., et al., *GENETICS IN ENDOCRINOLOGY: Approaches to molecular genetic diagnosis in the management of differences/disorders of sex development (DSD): position paper of EU COST Action BM 1303 'DSDnet'*. Eur J Endocrinol, 2018. **179**(4): p. R197-R206.
55. de Calais, F.L., et al., *A study of splicing mutations in disorders of sex development*. Sci Rep, 2017. **7**(1): p. 16202.
56. Hellwinkel, O.J., et al., *A unique exonic splicing mutation in the human androgen receptor gene indicates a physiologic relevance of regular androgen receptor transcript variants*. J Clin Endocrinol Metab, 2001. **86**(6): p. 2569-75.
57. Ho, S.S., A.E. Urban, und R.E. Mills, *Structural variation in the sequencing era*. Nat Rev Genet, 2020. **21**(3): p. 171-189.
58. Collins, R.L., et al., *An open resource of structural variation for medical and population genetics*. bioRxiv, 2019: p. 578674.
59. Croft, B., T. Ohnesorg, und A.H. Sinclair, *The Role of Copy Number Variants in Disorders of Sex Development*. Sex Dev, 2018. **12**(1-3): p. 19-29.
60. Kim, G.J., et al., *Copy number variation of two separate regulatory regions upstream of SOX9 causes isolated 46,XY or 46,XX disorder of sex development*. J Med Genet, 2015. **52**(4): p. 240-7.

61. Shankara Narayana, N., et al., *Painful ovulation in a 46,XX SRY -ve adult male with SOX9 duplication*. *Endocrinol Diabetes Metab Case Rep*, 2017. **2017**.
62. Ohnesorg, T., et al., *A duplication in a patient with 46,XX ovo-testicular disorder of sex development refines the SOX9 testis-specific regulatory region to 24 kb*. *Clin Genet*, 2017. **92**(3): p. 347-349.
63. Dangle, P., et al., *Female-to-male sex reversal associated with unique Xp21.2 deletion disrupting genomic regulatory architecture of the dosage-sensitive sex reversal region*. *J Med Genet*, 2017. **54**(10): p. 705-709.
64. White, S., et al., *Copy number variation in patients with disorders of sex development due to 46,XY gonadal dysgenesis*. *PLoS One*, 2011. **6**(3): p. e17793.
65. Barbaro, M., et al., *Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene*. *J Clin Endocrinol Metab*, 2007. **92**(8): p. 3305-13.
66. Barbaro, M., et al., *Gene dosage imbalances in patients with 46,XY gonadal DSD detected by an in-house-designed synthetic probe set for multiplex ligation-dependent probe amplification analysis*. *Clin Genet*, 2008. **73**(5): p. 453-64.
67. Ledig, S., et al., *Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci*. *Hum Reprod*, 2010. **25**(10): p. 2637-46.
68. Weischenfeldt, J., et al., *Phenotypic impact of genomic structural variation: insights from and for human disease*. *Nat Rev Genet*, 2013. **14**(2): p. 125-38.
69. Spielmann, M., D.G. Lupianez, und S. Mundlos, *Structural variation in the 3D genome*. *Nat Rev Genet*, 2018. **19**(7): p. 453-467.
70. Kazazian, H.H., Jr. und J.V. Moran, *Mobile DNA in Health and Disease*. *N Engl J Med*, 2017. **377**(4): p. 361-370.
71. Batista, R.L., et al., *Mobile DNA in Endocrinology: LINE-1 Retrotransposon Causing Partial Androgen Insensitivity Syndrome*. *J Clin Endocrinol Metab*, 2019. **104**(12): p. 6385-6390.
72. Iafrate, A.J., et al., *Detection of large-scale variation in the human genome*. *Nat Genet*, 2004. **36**(9): p. 949-51.
73. Seo, J.S., et al., *De novo assembly and phasing of a Korean human genome*. *Nature*, 2016. **538**(7624): p. 243-247.
74. Shi, L., et al., *Long-read sequencing and de novo assembly of a Chinese genome*. *Nat Commun*, 2016. **7**: p. 12065.
75. Bardoni, B., et al., *A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal*. *Nat Genet*, 1994. **7**(4): p. 497-501.
76. Illumina. *An introduction to Next-Generation Sequencing Technology*. 2017; Verfügbar unter: https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf (Zuletzt aufgerufen: 23.07.2021).
77. Illumina. *Illumina Sequencing Technology*. Technology Spotlight 2010; Verfügbar unter: https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf (Zuletzt aufgerufen: 23.07.2021).
78. Li, H. und R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. *Bioinformatics*, 2009. **25**(14): p. 1754-60.
79. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data*. *Nat Genet*, 2011. **43**(5): p. 491-8.

80. Wang, K., M. Li, und H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. *Nucleic Acids Res*, 2010. **38**(16): p. e164.
81. Vaser, R., et al., *SIFT missense predictions for genomes*. *Nat Protoc*, 2016. **11**(1): p. 1-9.
82. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. *Nat Methods*, 2010. **7**(4): p. 248-9.
83. Schwarz, J.M., et al., *MutationTaster2: mutation prediction for the deep-sequencing age*. *Nat Methods*, 2014. **11**(4): p. 361-2.
84. Boeva, V., et al., *Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data*. *Bioinformatics*, 2012. **28**(3): p. 423-5.
85. Rausch, T., et al., *DELLY: structural variant discovery by integrated paired-end and split-read analysis*. *Bioinformatics*, 2012. **28**(18): p. i333-i339.
86. Thorvaldsdottir, H., J.T. Robinson, und J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration*. *Brief Bioinform*, 2013. **14**(2): p. 178-92.
87. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. **25**(16): p. 2078-9.
88. Kallioniemi, A., et al., *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors*. *Science*, 1992. **258**(5083): p. 818-21.
89. Kent, W.J., *BLAT--the BLAST-like alignment tool*. *Genome Res*, 2002. **12**(4): p. 656-64.
90. Mullis, K., et al., *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. *Cold Spring Harb Symp Quant Biol*, 1986. **51 Pt 1**: p. 263-73.
91. Sambrook, J., E.F. Fritsch, und T. Maniatis, *Molecular Cloning - A Laboratory Manual*. Vol. I. 1998, Cold Spring Harbour, New York: Cold Spring Harbor Laboratory Press.
92. Sanger, F., S. Nicklen, und A.R. Coulson, *DNA sequencing with chain-terminating inhibitors*. *Proc Natl Acad Sci U S A*, 1977. **74**(12): p. 5463-7.
93. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification*. *Nat Biotechnol*, 2016. **34**(5): p. 525-7.
94. Sonesson, C., M.I. Love, und M.D. Robinson, *Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences*. *F1000Res*, 2015. **4**: p. 1521.
95. Carithers, L.J., et al., *A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project*. *Biopreserv Biobank*, 2015. **13**(5): p. 311-9.
96. Athar, A., et al., *ArrayExpress update - from bulk to single-cell expression data*. *Nucleic Acids Res*, 2019. **47**(D1): p. D711-D715.
97. Hinton, G.E. und S.T. Roweis, *Stochastic Neighbor Embedding*. *Advances in Neural Information Processing Systems*, 2002. **15**: p. 833-840.
98. van der Maaten, L., *Accelerating t-SNE using Tree-Based Algorithms*. *J Mach Learn Res*, 2014(15:3221-3245).
99. Hanzelmann, S., R. Castelo, und J. Guinney, *GSVA: gene set variation analysis for microarray and RNA-seq data*. *BMC Bioinformatics*, 2013. **14**: p. 7.
100. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection*. *Cell Syst*, 2015. **1**(6): p. 417-425.
101. D'Haene, B., J. Vandesompele, und J. Hellemans, *Accurate and objective copy number profiling using real-time quantitative PCR*. *Methods*, 2010. **50**(4): p. 262-70.
102. Pattyn, F., et al., *RTPrimerDB: the real-time PCR primer and probe database*. *Nucleic Acids Res*, 2003. **31**(1): p. 122-3.
103. Wong, Y.S., et al., *Incidence and diagnoses of disorders of sex development in proximal hypospadias*. *J Pediatr Surg*, 2018. **53**(12): p. 2498-2501.

104. van Berkum, N.L., et al., *Hi-C: a method to study the three-dimensional architecture of genomes*. J Vis Exp, 2010(39).
105. Dekker, J., et al., *Capturing chromosome conformation*. Science, 2002. **295**(5558): p. 1306-11.
106. Wang, Y., et al., *The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions*. Genome Biol, 2018. **19**(1): p. 151.
107. Kent, W.J., et al., *The human genome browser at UCSC*. Genome Res, 2002. **12**(6): p. 996-1006.
108. Bashamboo, A., et al., *Mutations in the FOG2/ZFPM2 gene are associated with anomalies of human testis determination*. Hum Mol Genet, 2014. **23**(14): p. 3657-65.
109. Boulanger, L., et al., *FOXL2 is a female sex-determining gene in the goat*. Curr Biol, 2014. **24**(4): p. 404-8.
110. Uhlenhaut, N.H., et al., *Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation*. Cell, 2009. **139**(6): p. 1130-42.
111. Ledig, S., et al., *Partial deletion of DMRT1 causes 46,XY ovotesticular disorder of sexual development*. Eur J Endocrinol, 2012. **167**(1): p. 119-24.
112. Matson, C.K., et al., *DMRT1 prevents female reprogramming in the postnatal mammalian testis*. Nature, 2011. **476**(7358): p. 101-4.
113. Stefan, E. und K. Bister, *MYC and RAF: Key Effectors in Cellular Signaling and Major Drivers in Human Cancer*. Curr Top Microbiol Immunol, 2017. **407**: p. 117-151.
114. Smyk, M., et al., *Male-to-female sex reversal associated with an approximately 250 kb deletion upstream of NROB1 (DAX1)*. Hum Genet, 2007. **122**(1): p. 63-70.
115. Baetens, D., et al., *Biallelic and monoallelic ESR2 variants associated with 46,XY disorders of sex development*. Genet Med, 2018. **20**(7): p. 717-727.
116. Davis, C.A., et al., *The Encyclopedia of DNA elements (ENCODE): data portal update*. Nucleic Acids Res, 2018. **46**(D1): p. D794-D801.
117. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome*. Nature, 2012. **489**(7414): p. 57-74.
118. Bell, A.C., A.G. West, und G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators*. Cell, 1999. **98**(3): p. 387-96.
119. Wallace, J.A. und G. Felsenfeld, *We gather together: insulators and genome organization*. Curr Opin Genet Dev, 2007. **17**(5): p. 400-7.
120. Dorman, E.R., A.M. Bushey, und V.G. Corces, *The role of insulator elements in large-scale chromatin structure in interphase*. Semin Cell Dev Biol, 2007. **18**(5): p. 682-90.
121. Chung, J.H., M. Whiteley, und G. Felsenfeld, *A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila*. Cell, 1993. **74**(3): p. 505-14.
122. Bell, A.C. und G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene*. Nature, 2000. **405**(6785): p. 482-5.
123. Lupianez, D.G., et al., *Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions*. Cell, 2015. **161**(5): p. 1012-1025.
124. Szabo, P., et al., *Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function*. Curr Biol, 2000. **10**(10): p. 607-10.
125. Shrimali, S., et al., *An ectopic CTCF-dependent transcriptional insulator influences the choice of Vbeta gene segments for VDJ recombination at TCRbeta locus*. Nucleic Acids Res, 2012. **40**(16): p. 7753-65.

126. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398): p. 376-80.
127. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre*. Nature, 2012. **485**(7398): p. 381-5.
128. Ramirez, F., et al., *High-resolution TADs reveal DNA sequences underlying genome organization in flies*. Nat Commun, 2018. **9**(1): p. 189.
129. Gomez-Marin, C., et al., *Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders*. Proc Natl Acad Sci U S A, 2015. **112**(24): p. 7542-7.
130. Narendra, V., et al., *CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation*. Science, 2015. **347**(6225): p. 1017-21.
131. Sanborn, A.L., et al., *Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes*. Proc Natl Acad Sci U S A, 2015. **112**(47): p. E6456-65.
132. de Wit, E., et al., *CTCF Binding Polarity Determines Chromatin Looping*. Mol Cell, 2015. **60**(4): p. 676-84.
133. Merckenschlager, M. und E.P. Nora, *CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation*. Annu Rev Genomics Hum Genet, 2016. **17**: p. 17-43.
134. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project*. Nat Genet, 2013. **45**(6): p. 580-5.
135. Fishilevich, S., et al., *GeneHancer: genome-wide integration of enhancers and target genes in GeneCards*. Database (Oxford), 2017. **2017**.
136. Fudenberg, G., et al., *Formation of Chromosomal Domains by Loop Extrusion*. Cell Rep, 2016. **15**(9): p. 2038-49.
137. Ibn-Salem, J., et al., *Deletions of chromosomal regulatory boundaries are associated with congenital disease*. Genome Biol, 2014. **15**(9): p. 423.
138. Lettice, L.A., et al., *Enhancer-adoption as a mechanism of human developmental disease*. Hum Mutat, 2011. **32**(12): p. 1492-9.
139. Hnisz, D., et al., *Activation of proto-oncogenes by disruption of chromosome neighborhoods*. Science, 2016. **351**(6280): p. 1454-1458.
140. Weischenfeldt, J., et al., *Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking*. Nat Genet, 2017. **49**(1): p. 65-74.
141. Viswanathan, S.R., et al., *Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing*. Cell, 2018. **174**(2): p. 433-447 e19.
142. Guo, Y., et al., *CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function*. Cell, 2015. **162**(4): p. 900-10.
143. Vietri Rudan, M., et al., *Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture*. Cell Rep, 2015. **10**(8): p. 1297-309.
144. Lupianez, D.G., M. Spielmann, und S. Mundlos, *Breaking TADs: How Alterations of Chromatin Domains Result in Disease*. Trends Genet, 2016. **32**(4): p. 225-237.
145. Rao, S.S.P., et al., *Cohesin Loss Eliminates All Loop Domains*. Cell, 2017. **171**(2): p. 305-320 e24.
146. Goloborodko, A., J.F. Marko, und L.A. Mirny, *Chromosome Compaction by Active Loop Extrusion*. Biophys J, 2016. **110**(10): p. 2162-8.
147. Szabo, Q., F. Bantignies, und G. Cavalli, *Principles of genome folding into topologically associating domains*. Sci Adv, 2019. **5**(4): p. eaaw1668.
148. Chen, K., et al., *BreakDancer: an algorithm for high-resolution mapping of genomic structural variation*. Nat Methods, 2009. **6**(9): p. 677-81.

149. Abyzov, A., et al., *CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing*. *Genome Res*, 2011. **21**(6): p. 974-84.
150. Korbel, J.O., et al., *PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data*. *Genome Biol*, 2009. **10**(2): p. R23.
151. Handsaker, R.E., et al., *Discovery and genotyping of genome structural polymorphism by sequencing on a population scale*. *Nat Genet*, 2011. **43**(3): p. 269-76.
152. Layer, R.M., et al., *LUMPY: a probabilistic framework for structural variant discovery*. *Genome Biol*, 2014. **15**(6): p. R84.
153. Michaelson, J.J. und J. Sebat, *forestSV: structural variant discovery through statistical learning*. *Nat Methods*, 2012. **9**(8): p. 819-21.
154. Wong, K., et al., *Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly*. *Genome Biol*, 2010. **11**(12): p. R128.
155. Lam, H.Y., et al., *Detecting and annotating genetic variations using the HugeSeq pipeline*. *Nat Biotechnol*, 2012. **30**(3): p. 226-9.
156. Eid, J., et al., *Real-time DNA sequencing from single polymerase molecules*. *Science*, 2009. **323**(5910): p. 133-8.
157. Lu, H., F. Giordano, und Z. Ning, *Oxford Nanopore MinION Sequencing and Genome Assembly*. *Genomics Proteomics Bioinformatics*, 2016. **14**(5): p. 265-279.
158. Ancien, F., et al., *Prediction and interpretation of deleterious coding variants in terms of protein structural stability*. *Sci Rep*, 2018. **8**(1): p. 4480.
159. Miosge, L.A., et al., *Comparison of predicted and actual consequences of missense mutations*. *Proc Natl Acad Sci U S A*, 2015. **112**(37): p. E5189-98.
160. Wells, A., et al., *Ranking of non-coding pathogenic variants and putative essential regions of the human genome*. *Nat Commun*, 2019. **10**(1): p. 5241.
161. Ritchie, G.R., et al., *Functional annotation of noncoding sequence variants*. *Nat Methods*, 2014. **11**(3): p. 294-6.
162. Majewski, J., et al., *What can exome sequencing do for you?* *J Med Genet*, 2011. **48**(9): p. 580-9.
163. Saunders, C.J., et al., *Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units*. *Sci Transl Med*, 2012. **4**(154): p. 154ra135.
164. Farnaes, L., et al., *Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization*. *NPJ Genom Med*, 2018. **3**: p. 10.
165. French, C.E., et al., *Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children*. *Intensive Care Med*, 2019. **45**(5): p. 627-636.
166. Fox, E.J., et al., *Accuracy of Next Generation Sequencing Platforms*. *Next Gener Seq Appl*, 2014. **1**.
167. Pfeiffer, F., et al., *Systematic evaluation of error rates and causes in short samples in next-generation sequencing*. *Sci Rep*, 2018. **8**(1): p. 10950.
168. Goodwin, S., J.D. McPherson, und W.R. McCombie, *Coming of age: ten years of next-generation sequencing technologies*. *Nat Rev Genet*, 2016. **17**(6): p. 333-51.
169. Lander, E.S., et al., *Initial sequencing and analysis of the human genome*. *Nature*, 2001. **409**(6822): p. 860-921.
170. Bentley, D.R., et al., *Accurate whole human genome sequencing using reversible terminator chemistry*. *Nature*, 2008. **456**(7218): p. 53-9.

171. Weaver, S., et al., *Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution*. *Methods*, 2010. **50**(4): p. 271-6.
172. Zhao, Y., et al., *Robustness of RNA sequencing on older formalin-fixed paraffin-embedded tissue from high-grade ovarian serous adenocarcinomas*. *PLoS One*, 2019. **14**(5): p. e0216050.
173. Hedegaard, J., et al., *Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue*. *PLoS One*, 2014. **9**(5): p. e98187.
174. Esteve-Codina, A., et al., *A Comparison of RNA-Seq Results from Paired Formalin-Fixed Paraffin-Embedded and Fresh-Frozen Glioblastoma Tissue Samples*. *PLoS One*, 2017. **12**(1): p. e0170632.
175. Li, P., et al., *Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq*. *BMC Genomics*, 2014. **15**: p. 1087.
176. Graw, S., et al., *Robust gene expression and mutation analyses of RNA-sequencing of formalin-fixed diagnostic tumor samples*. *Sci Rep*, 2015. **5**: p. 12335.
177. Yu, R.N., et al., *Role of Ahch in gonadal development and gametogenesis*. *Nat Genet*, 1998. **20**(4): p. 353-7.
178. Jassal, B., et al., *The reactome pathway knowledgebase*. *Nucleic Acids Res*, 2020. **48**(D1): p. D498-D503.
179. Liu, X.F., et al., *DAX1 promotes cervical cancer cell growth and tumorigenicity through activation of Wnt/beta-catenin pathway via GSK3beta*. *Cell Death Dis*, 2018. **9**(3): p. 339.
180. Garcia-Aragoncillo, E., et al., *DAX1, a direct target of EWS/FLI1 oncoprotein, is a principal regulator of cell-cycle progression in Ewing's tumor cells*. *Oncogene*, 2008. **27**(46): p. 6034-43.
181. Fujii, S., et al., *Nr0b1 is a negative regulator of Zscan4c in mouse embryonic stem cells*. *Sci Rep*, 2015. **5**: p. 9146.
182. Jiang, H.L., et al., *DAX-1 inhibits hepatocellular carcinoma proliferation by inhibiting beta-catenin transcriptional activity*. *Cell Physiol Biochem*, 2014. **34**(3): p. 734-42.
183. Zhang, R., et al., *Predicting CTCF-mediated chromatin loops using CTCF-MP*. *Bioinformatics*, 2018. **34**(13): p. i133-i141.
184. Batzoglou, S., et al., *Human and mouse gene structure: comparative analysis and application to exon prediction*. *Genome Res*, 2000. **10**(7): p. 950-8.
185. Heger, P., et al., *The chromatin insulator CTCF and the emergence of metazoan diversity*. *Proc Natl Acad Sci U S A*, 2012. **109**(43): p. 17507-12.
186. Novacek, M.J., *Mammalian phylogeny: shaking the tree*. *Nature*, 1992. **356**(6365): p. 121-5.
187. Makalowski, W., J. Zhang, und M.S. Boguski, *Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences*. *Genome Res*, 1996. **6**(9): p. 846-57.
188. Lobanenkova, V.V. und G.E. Zentner, *Discovering a binary CTCF code with a little help from BORIS*. *Nucleus*, 2018. **9**(1): p. 33-41.
189. Dixon, J.R., et al., *Chromatin architecture reorganization during stem cell differentiation*. *Nature*, 2015. **518**(7539): p. 331-6.
190. Krefting, J., M.A. Andrade-Navarro, und J. Ibn-Salem, *Evolutionary stability of topologically associating domains is associated with conserved gene regulation*. *BMC Biol*, 2018. **16**(1): p. 87.

191. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping*. Cell, 2014. **159**(7): p. 1665-80.
192. Sexton, T. und G. Cavalli, *The role of chromosome domains in shaping the functional genome*. Cell, 2015. **160**(6): p. 1049-59.
193. Kong, S. und Y. Zhang, *Deciphering Hi-C: from 3D genome to function*. Cell Biol Toxicol, 2019. **35**(1): p. 15-32.
194. Dali, R. und M. Blanchette, *A critical assessment of topologically associating domain prediction tools*. Nucleic Acids Res, 2017. **45**(6): p. 2994-3005.
195. Hansen, P., et al., *Computational Processing and Quality Control of Hi-C, Capture Hi-C and Capture-C Data*. Genes (Basel), 2019. **10**(7).
196. Nagano, T., et al., *Comparison of Hi-C results using in-solution versus in-nucleus ligation*. Genome Biol, 2015. **16**: p. 175.
197. Mota-Gomez, I. und D.G. Lupianez, *A (3D-Nuclear) Space Odyssey: Making Sense of Hi-C Maps*. Genes (Basel), 2019. **10**(6).
198. Yu, M. und B. Ren, *The Three-Dimensional Organization of Mammalian Genomes*. Annu Rev Cell Dev Biol, 2017. **33**: p. 265-289.
199. Phillips-Cremins, J.E., et al., *Architectural protein subclasses shape 3D organization of genomes during lineage commitment*. Cell, 2013. **153**(6): p. 1281-95.
200. Berlivet, S., et al., *Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs*. PLoS Genet, 2013. **9**(12): p. e1004018.
201. Jin, F., et al., *A high-resolution map of the three-dimensional chromatin interactome in human cells*. Nature, 2013. **503**(7475): p. 290-4.
202. Pennacchio, L.A., et al., *In vivo enhancer analysis of human conserved non-coding sequences*. Nature, 2006. **444**(7118): p. 499-502.
203. Zerbino, D.R., et al., *The ensembl regulatory build*. Genome Biol, 2015. **16**: p. 56.
204. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers*. Nature, 2010. **465**(7295): p. 182-7.
205. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues*. Nature, 2014. **507**(7493): p. 455-461.
206. Visel, A., et al., *VISTA Enhancer Browser--a database of tissue-specific human enhancers*. Nucleic Acids Res, 2007. **35**(Database issue): p. D88-92.
207. Zacher, B., et al., *Accurate Promoter and Enhancer Identification in 127 ENCODE and Roadmap Epigenomics Cell Types and Tissues by GenoSTAN*. PLoS One, 2017. **12**(1): p. e0169249.
208. Sanyal, A., et al., *The long-range interaction landscape of gene promoters*. Nature, 2012. **489**(7414): p. 109-13.
209. Rubtsov, M.A., et al., *Chromatin structure can strongly facilitate enhancer action over a distance*. Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17690-5.
210. Phillips, J.E. und V.G. Corces, *CTCF: master weaver of the genome*. Cell, 2009. **137**(7): p. 1194-211.
211. Nikolov, D.B. und S.K. Burley, *RNA polymerase II transcription initiation: a structural view*. Proc Natl Acad Sci U S A, 1997. **94**(1): p. 15-22.
212. Wilczynski, B. und E.E. Furlong, *Dynamic CRM occupancy reflects a temporal map of developmental progression*. Mol Syst Biol, 2010. **6**: p. 383.
213. Jakobsen, J.S., et al., *Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network*. Genes Dev, 2007. **21**(19): p. 2448-60.
214. Barbaro, M., et al., *Multigeneration Inheritance through Fertile XX Carriers of an NROB1 (DAX1) Locus Duplication in a Kindred of Females with Isolated XY Gonadal Dysgenesis*. Int J Endocrinol, 2012. **2012**: p. 504904.

215. Dong, Y., et al., *Targeted next-generation sequencing identification of mutations in patients with disorders of sex development*. BMC Med Genet, 2016. **17**: p. 23.
216. Garcia-Acero, M., et al., *Gene dosage of DAX-1, determining in sexual differentiation: duplication of DAX-1 in two sisters with gonadal dysgenesis*. Mol Biol Rep, 2019.
217. Swain, A., et al., *Dax1 antagonizes Sry action in mammalian sex determination*. Nature, 1998. **391**(6669): p. 761-7.
218. Correa, R.V., et al., *A microdeletion in the ligand binding domain of human steroidogenic factor 1 causes XY sex reversal without adrenal insufficiency*. J Clin Endocrinol Metab, 2004. **89**(4): p. 1767-72.
219. Mallet, D., et al., *Gonadal dysgenesis without adrenal insufficiency in a 46, XY patient heterozygous for the nonsense C16X mutation: a case of SF1 haploinsufficiency*. J Clin Endocrinol Metab, 2004. **89**(10): p. 4829-32.
220. Hasegawa, T., et al., *Testicular dysgenesis without adrenal insufficiency in a 46,XY patient with a heterozygous inactive mutation of steroidogenic factor-1*. J Clin Endocrinol Metab, 2004. **89**(12): p. 5930-5.
221. Miyado, M., et al., *The p.R92W variant of NR5A1/Nr5a1 induces testicular development of 46,XX gonads in humans, but not in mice: phenotypic comparison of human patients and mutation-induced mice*. Biol Sex Differ, 2016. **7**: p. 56.
222. Park, S.Y., et al., *A phenotypic spectrum of sexual development in Dax1 (Nr0b1)-deficient mice: consequence of the C57BL/6J strain on sex determination*. Biol Reprod, 2008. **79**(6): p. 1038-45.
223. Fernandes-Freitas, I., et al., *Live Birth in Sex-Reversed XY Mice Lacking the Nuclear Receptor Dax1*. Sci Rep, 2020. **10**(1): p. 1703.
224. Bouma, G.J., et al., *Gonadal sex reversal in mutant Dax1 XY mice: a failure to upregulate Sox9 in pre-Sertoli cells*. Development, 2005. **132**(13): p. 3045-54.
225. Suntharalingham, J.P., et al., *DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease*. Best Pract Res Clin Endocrinol Metab, 2015. **29**(4): p. 607-19.
226. Ostermann, S., et al., *Importance of genetic diagnosis of DAX-1 deficiency: example from a large, multigenerational family*. Horm Res, 2006. **65**(4): p. 163-8.
227. Sanlaville, D., et al., *Functional disomy of Xp including duplication of DAX1 gene with sex reversal due to t(X;Y)(p21.2;p11.3)*. Am J Med Genet A, 2004. **128A**(3): p. 325-30.
228. Sukumaran, A., et al., *Duplication of dosage sensitive sex reversal area in a 46, XY patient with normal sex determining region of Y causing complete sex reversal*. J Pediatr Endocrinol Metab, 2013. **26**(7-8): p. 775-9.
229. Simioni, M., et al., *A Cytogenomic Approach in a Case of Syndromic XY Gonadal Dysgenesis*. Sex Dev, 2016. **10**(1): p. 23-7.
230. Gimelli, G., et al., *Gene dosage of the spermidine/spermine N(1)-acetyltransferase (SSAT) gene with putrescine accumulation in a patient with a Xp21.1p22.12 duplication and keratosis follicularis spinulosa decalvans (KFSD)*. Hum Genet, 2002. **111**(3): p. 235-41.
231. Doudna, J.A. und E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.
232. Troll, C.J., et al., *Structural Variation Detection by Proximity Ligation from Formalin-Fixed, Paraffin-Embedded Tumor Tissue*. J Mol Diagn, 2019. **21**(3): p. 375-383.
233. Chen, X., et al., *Differentiation of human induced pluripotent stem cells into Leydig-like cells with molecular compounds*. Cell Death Dis, 2019. **10**(3): p. 220.
234. Rodriguez Gutierrez, D., W. Eid, und A. Biason-Lauber, *A Human Gonadal Cell Model From Induced Pluripotent Stem Cells*. Front Genet, 2018. **9**: p. 498.

235. Eozenou, C., K. McElreavey, und A. Bashamboo, *Development of testicular organoids to understand Disorders of Sex Development*. *Horm Res Paediatr*, 2019. **91**((suppl 1)): p. 116.
236. Knarston, I.M., et al., *An In Vitro Differentiation Protocol for Human Embryonic Bipotential Gonad and Testis Cell Development*. *Stem Cell Reports*, 2020. **15**(6): p. 1377-1391.
237. Garieri, M., et al., *Extensive cellular heterogeneity of X inactivation revealed by single-cell allele-specific expression in human fibroblasts*. *Proc Natl Acad Sci U S A*, 2018. **115**(51): p. 13015-13020.
238. Heinz, L.X., et al., *TASL is the SLC15A4-associated adaptor for IRF5 activation by TLR7-9*. *Nature*, 2020. **581**(7808): p. 316-322.

7. Anhang

7.1. Ethikvotum

Die in dieser Arbeit durchgeführten WGS-Analysen erfolgten mit den Zustimmungen der Ethikkommissionen der Universität zu Lübeck (AZ 17-219; 09.08.2017) und der Universität Campinas, Sao Paulo (CAAE 24972513.5.0000.5404).

Das qPCR-Screening erfolgt im Rahmen des EuroDSD-Projekts und den Zustimmungen der Ethikkommissionen der Universität zu Lübeck (AZ 08-081; 28.05.2008) und der Christian-Albrechts-Universität zu Kiel (D 410/08, 05.03.2008).

7.1. Abbildungen

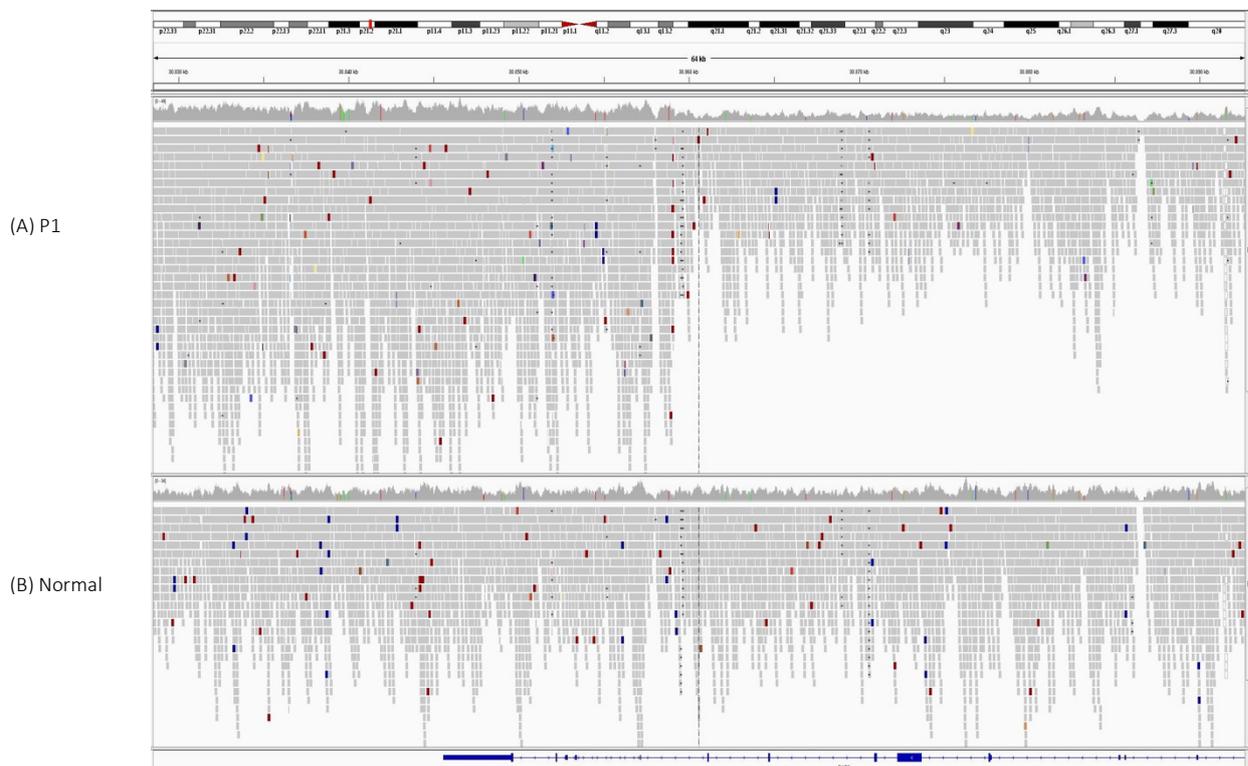


Abb. I. Im IGV dargestellte WGS-Daten der Patient:in 1 (A) und einer Kontrollprobe (B)

Zu sehen ist der Chromosomenabschnitt ChrX:30,828,511-30,892,682. (A) Die Daten der Patient:in 1 (P1) zeigen bei ca. 30,859kb eine deutliche Veränderung der Anzahl an Sequenzierungs-Reads (graue Streifen). Darüber hinaus ist an dem Übergang von hoher zu niedriger Read-Anzahl eine hohe Dichte an chimären Reads (rote Streifen) zu erkennen. Die chimären Reads und die im Verhältnis 2:1 erhöhte Read-Anzahl deuten darauf hin, dass es sich hierbei um das Ende einer Duplikation handelt. (B) Zum Vergleich eine normale Vergleichsprobe, ohne Kopiezahlveränderung im entsprechenden Chromosomenabschnitt.

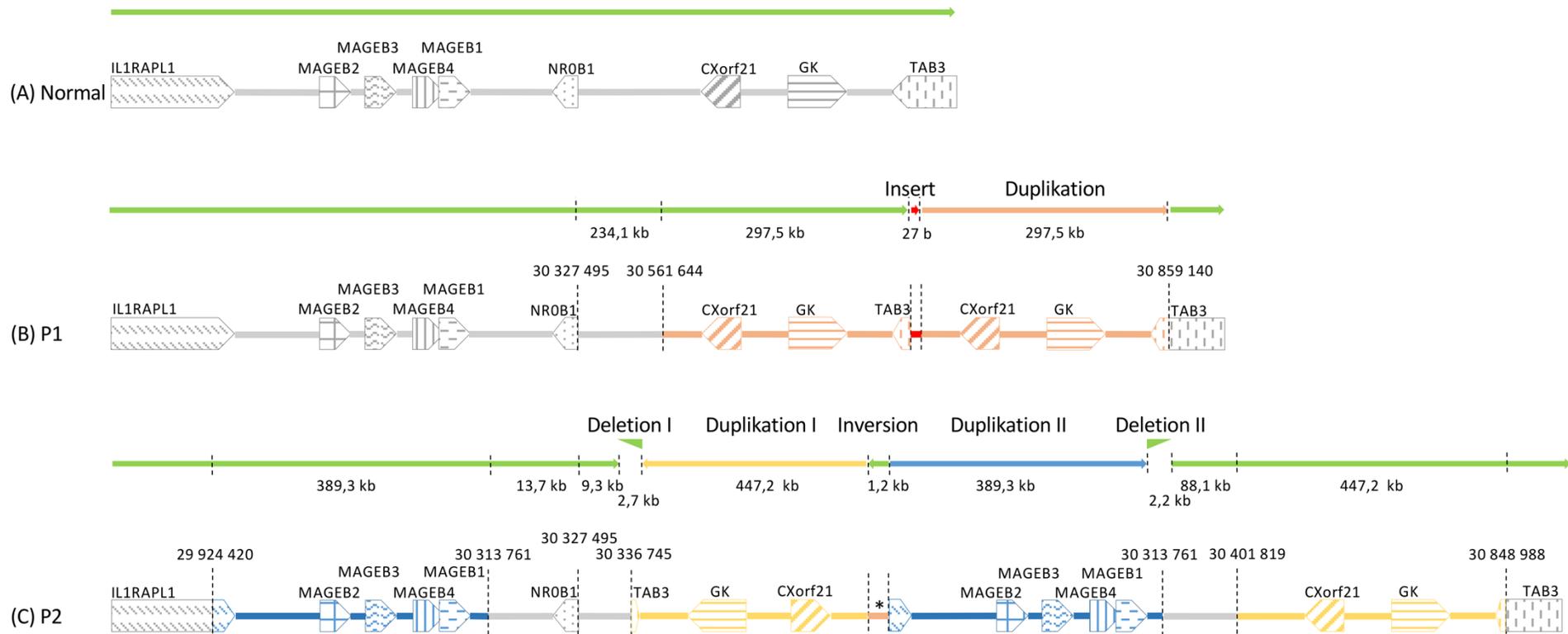


Abb. II. Darstellung der strukturellen Variationen der Patient:innen 1 (B) und 2 (C) im Vergleich zur Normalsequenz (A) inklusive Positionsangaben der chromosomalen Bruchpunkte.

* = invertiert

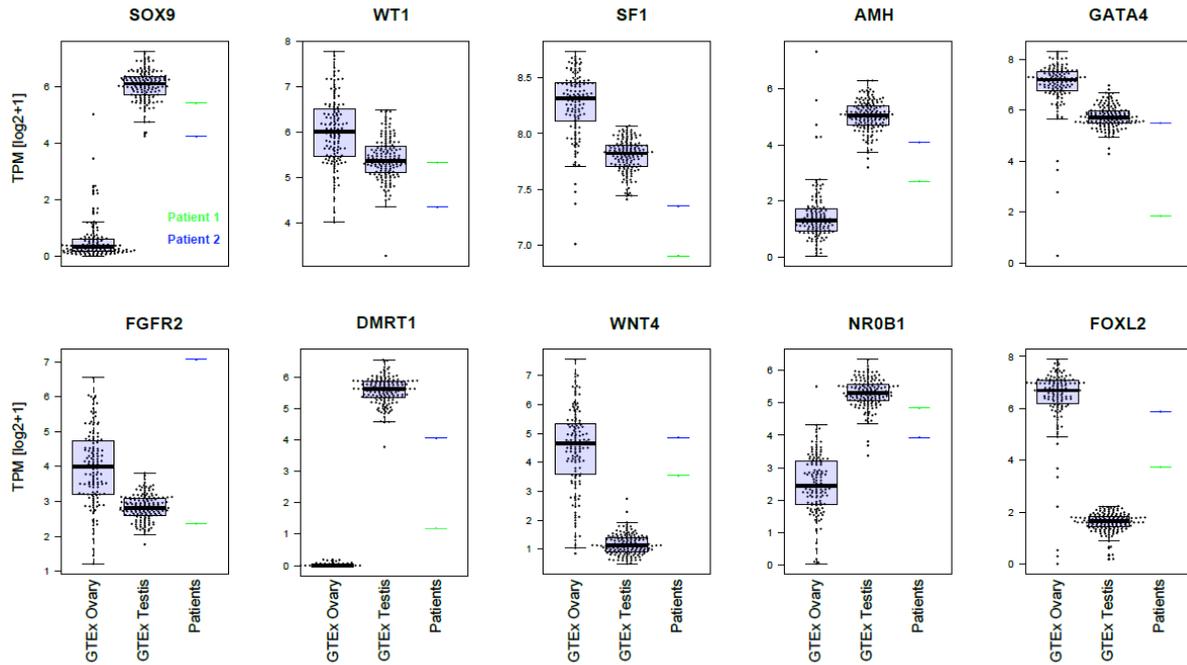


Abb. III. Expressionsvergleich ausgewählter Gene der Geschlechtsentwicklung

Die Expressionsdaten von ovariellen (links) und testikulären (Mitte) Kontrollproben der GTEx Datenbank sind denen der Patient:innen 1 und 2 (rechts) gegenübergestellt. Die aufgetragenen GTEx-Proben entsprechen denen, die für die t-SNE Berechnung (Abb. 10) genutzt wurden. Der Beeswarm und Box-Whisker-Plot zeigen die log₂-transformierte Zahl der Transkripte pro Million des entsprechenden Gens. Die Box umspannt den Bereich von der 1. zur 3. Quartile und die Antennen (Whisker) den 1,5fachen Interquartilenabstand.

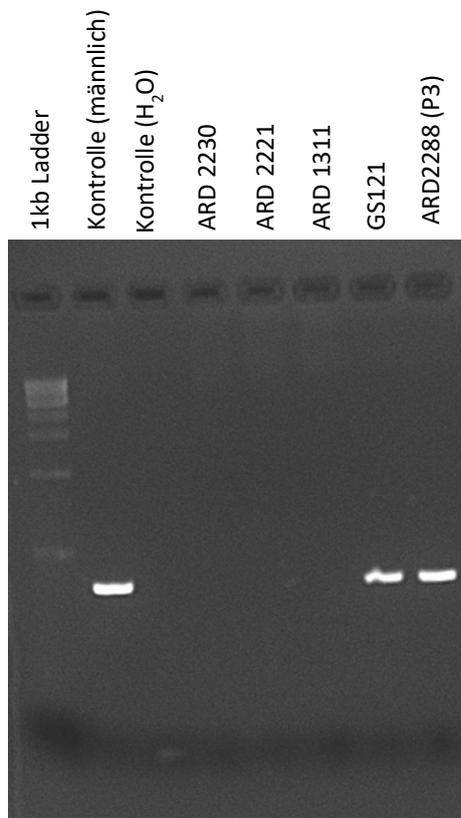


Abb. V. SRY-PCR von fünf im qPCR-Screening auffälligen Patient:innen

Darstellung des Agarose-Gels nach Amplifikation eines Segments im Exon 1 des *SRY*-Gens. Die Patient:innen ARD2230, ARD2221, ARD1311, ARD2288, GS121 zeigten in der qPCR eine auffällige Erhöhung am Xp21.2-Locus. Bei ARD2230 und ARD2221 kann kein *SRY*-Amplifikationsprodukt nachgewiesen werden. In Zusammenschau mit dem qPCR-Ergebnis ist anzunehmen, dass es sich bei diesen Fällen um 46,XX-Patient:innen handelt. Zum Karyotyp dieser beiden Patient:innen liegen jedoch keine zytogenetischen Informationen vor.

Bei der ebenfalls *SRY*-negativen Patient:in ARD1311, ließ sich nach Rücksprache mit einem behandelnden Arzt verifizieren, dass die Patient:in eine syndromale DSD-Form mit 46,XX-Karyotyp trägt, die in den Forschungsakten nicht vermerkt war.

Für die Patient:in GS121 ließ sich in einer Wiederholungsläufen der qPCR die Xp21.2-Duplikation nicht bestätigen. Es ist bei ihr von einer anderen DSD-Ursache auszugehen. Die *SRY*-PCR von Patient:in ARD2288 konnte hingegen ein *SRY*-Amplifikat nachweisen und die Folge-qPCRs eine Kopiezahlerhöhungen am Xp21.2-Lokus bestätigen. ARD2288 wurde daraufhin in die vorliegende Arbeit als Patient:in 3 (P3) eingeschlossen.

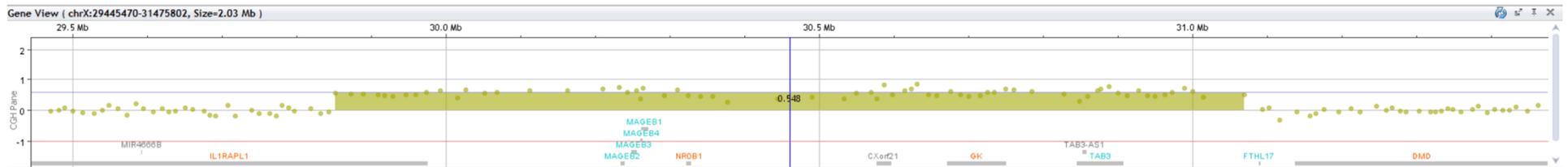


Abb. VI. ArrayCGH-Ergebnis der Mutter von P3 in der Region von *NROB1*

Im Bereich von ChrX:29,851,537-31,069,736 zeigt sich ein um x0,5 erhöhtes Hybridisierungsverhältnis, als Zeichen einer Duplikation. Der Duplikationsbereich endet telomerisch in *IL1RAPL1* und zentromerisch kurz vor *FTHL17* und entspricht damit dem triplizierten Bereich von P3.

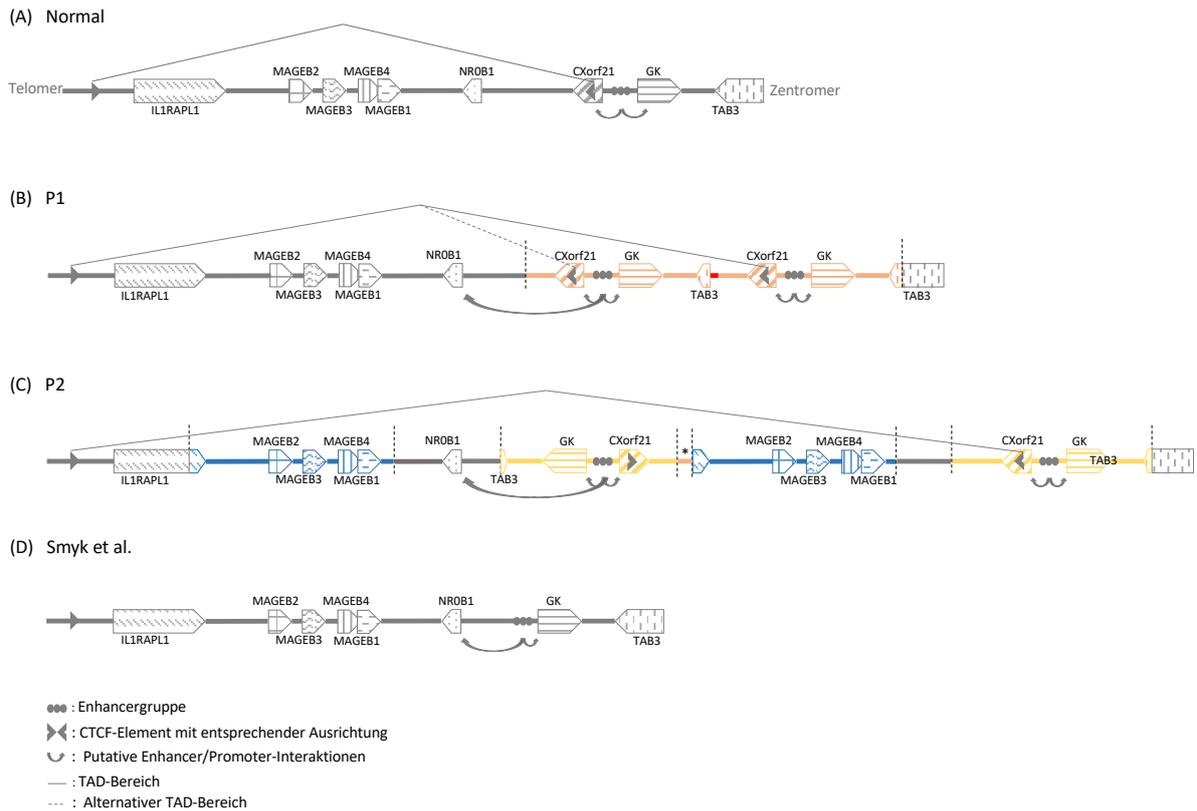


Abb. VII. Mögliche TAD-Konformationen im Bereich des *NROB1*-Gens

Normale Anordnung am Xp21.2-Locus (A) im Vergleich zu Patient:innen mit Kopiezahlvariation an dem Locus (B-D). Die grauen Pfeilspitzen stellen Bindungsstellen des CCCTC-binding factor (CTCF) dar, welcher als Isolator wirken kann. Wenn CTCFs konvergent angeordnet sind, kann zwischen Ihnen eine TAD (graue Dreiecke) ausgebildet werden. (B) Bei Patient:in 1 ergeben sich zwei mögliche TAD-Konformationen, dargestellt mit Hilfe durchgehender und/oder gestrichelter Linie/n des grauen Dreiecks. Die von den durchgehenden grauen Linien dargestellte TAD würde eine Interaktion der zwischen *GK* und *CXorf21* gelegenen Enhancern (graue Punkte) mit *NROB1* erlauben. (C) Die bei Patient:in 2 nachgewiesene Inversion bringt die Enhancer in die Nähe von *NROB1* und umgeht zeitgleich die normalerweise isolierende Wirkung des zwischen *NROB1* und den Enhancern gelegenen CTCFs. (D) Die von Smyk et al. [114] beschriebene Deletion verändert die physiologische TAD-Konformation, indem das CTCF-Bindungsmotiv entfernt wird. Die Deletion erlaubt Interaktionen über die sonst physiologische TAD-Grenze hinaus.

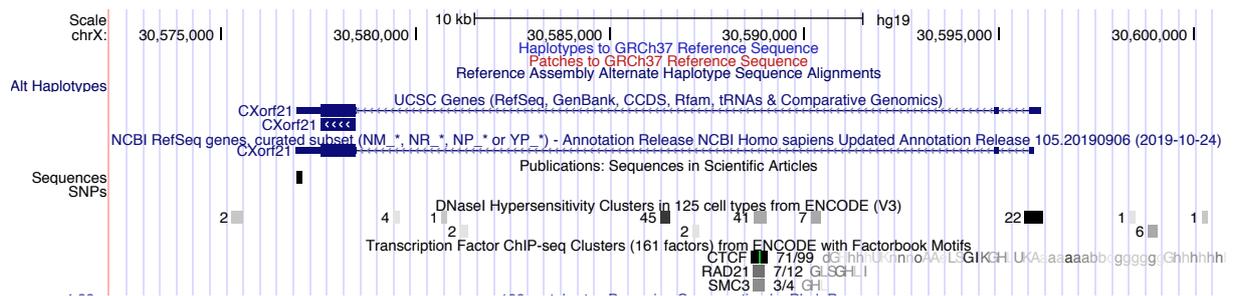


Abb. VIII. Dnase I-Hypersensitivität im Bereich des *CXorf21*-Gen

Daten der ENCODE-Datenbank [116] zu Dnase I-Hypersensitivität und den Bindungsstellen von CTCF, RAD21 und SMC3 wurden im UCSC-Browser [107] visualisiert. Es zeigt sich eine genaue Übereinstimmung der Dnase I Hypersensitivität und der Isolator-Bindungsstelle.

7.2 Geräte

4-Kapillarsequenzierer 3130 Genetic Analyzers	Applied Biosystems, ThermoFisher Scientific, USA
BioPhotometer Plus	Eppendorf, Deutschland
Centrifuge 5804R	Eppendorf, Deutschland
Precision Tower 5810	Dell, USA
Fusion SL	Vilber Lourmat, Deutschland
GPS 200/400 Powersupply	Pharmacia, Schweden
IKAMAG RCT	IKA, Deutschland
LightCycler 96	Hoffmann-La Roche, Schweiz
Mastercycler Gradient	Eppendorf, Deutschland
Mastercycler Nexus X2e	Eppendorf, Deutschland
Micro Centrifuge II	Labtech, Vereinigtes Königreich
MS2 Minishaker	IKA, Deutschland
Owl B1A EasyCast	ThermoFisher Scientific, USA
PCR Workstation Pro	VWR Life Science Competence Center, Deutschland
Qubit 3.0 Fluorometer	ThermoFisher Scientific, USA
Refrigerated Condensation Trap	ThermoFisher Scientific, USA
Sigma 1-15 Zentrifuge	Sigma, Deutschland
SpeedVac Concentrator	ThermoFisher Scientific, USA
Thermomixer 5436	Eppendorf, Deutschland

7.3 Computerprogramme

Affinity Designer	Serif, Vereinigtes Königreich
Chromas	Griffith University, Australien
CytoGenomics	Agilent Technologies, USA
Excel	Microsoft, USA
Firefox	Mozilla Foundation, USA
Fusion.Capt Advance	Vilber Lourmat, Germany
Integrative Genome Viewer (IGV)	Broad Institute, USA
Lasergene 14 & 15 (SeqMan, EditSeq, SeqBuilder, MegAlign, PrimerSelect)	DNASTAR, USA
LightCycler 96 SW 1.1	Roche, Schweiz
Powerpoint	Microsoft, USA
SeqScape 3	Applied Biosystems, ThermoFisher Scientific, USA
Sequence Analysis 6	Applied Biosystems, ThermoFisher Scientific, USA
Xmind	Xmind Ltd, China

7.4 Reagenzien und Kits

AmpliTaq DNA Polymerase Kit	Applied Biosystems, ThermoFisher Scientific, USA
Ampuwa	Fresenius Kabi, Deutschland
BigDye Terminator v3.1 Sequencing kit	ThermoFisher Scientific, USA
Borsäure	Sigma-Aldrich, Merck, Deutschland
Bromophenolblau	Sigma-Aldrich, Merck, Deutschland
dNTPs (dATP, dCTP, dGTP, dTTP)	Carl Roth, Deutschland
Ethanol	Merck, Deutschland
Exonuklease I	ThermoFisher Scientific, USA
Glycerin	Merck, Deutschland
HiDi-Formamid	ThermoFisher Scientific, USA
Midori green direct	Nippon Genetics Europe, Deutschland)
Na ₂ -EDTA	Sigma-Aldrich, Merck, Deutschland
NEEO Agarose	Carl Roth, Deutschland
Endotoxin-free TE Buffer	Qiagen, Deutschland
Qubit dsDNA BR Assay Kit	ThermoFisher Scientific, USA
Shrimp Alkaline Phosphatase	ThermoFisher Scientific, USA
Takyon No ROX SYBR MasterMix dTTP Blue	Eurogentec, Belgien
Trishydroxymethylaminomethan (Tris)	Merck, Deutschland
100bp ladder	New England BioLabs, USA

7.5 Online-Tools & Anwendungen

Chorogenome Navigator	http://chorogenome.ie-freiburg.mpg.de
Database for CTCF binding sites and genome organization	http://insulatordb.uthsc.edu
dbSNP	https://www.ncbi.nlm.nih.gov/snp
EnhancerAtlas	http://www.enhanceratlas.org/index.php
ExAC Browser	http://exac.broadinstitute.org
Fathmm	http://fathmm.biocompute.org.uk
Genotype-Tissue Expression (GTEx) project	https://gtexportal.org/home/
gnomAD Browser	https://gnomad.broadinstitute.org
Human BLAT Search	http://mgc.ucsc.edu/cgi-bin/hgBlat
JASPAR	http://jaspar.genereg.net
MutationAssessor	http://mutationassessor.org/r3/
MutationTaster	http://www.mutationtaster.org
NCBI BLAST Global Align	https://blast.ncbi.nlm.nih.gov/Blast.cgi
NCBI Gene	https://www.ncbi.nlm.nih.gov/gene
PennState 3D Genome Browser	http://promoter.bx.psu.edu/hi-c/
Polyphen	http://genetics.bwh.harvard.edu/pph2/
RTPrimer Database	http://www.rtpimerdb.org
SankeyMATIC	http://sankeymatic.com
SIFT	https://sift.bii.a-star.edu.sg
SNPCheck V3	https://genetools.org/SNPCheck/snpcheck.htm
The Human Protein Atlas	https://www.proteinatlas.org
The International Genome Sample Resource	http://www.internationalgenome.org
UCSC Genome Browser	https://genome-euro.ucsc.edu/cgi-bin/hgGateway?redirect=manual&source=genome.ucsc.edu
UCSC In-Silico PCR	http://rohsdb.cmb.usc.edu/Gbshape/cgi-bin/hgPcr
Universal Probelibrary Assay Design Center	https://lifescience.roche.com/en_de/brands/universal-probe-library.html#assay-design-center

7.6 PCR-Primer

Primername	Sequenz (5'→3')	Annealing-temperatur (°C)	Verwendung
AP_4AJ_CXorf21_fwd	CAA ATG CTG GTG GGG TTA TG	55,5	Sequenzierung P1
AP_4AJ_CXorf21_rev	TCA AAT TGG CTT CCT TCT CTT AGT	55,5	Sequenzierung P1
AP_4AJ_TAB3_fwd	GTT TGT CTG GCG ATA CTT CTT	55,5	Sequenzierung P1
AP_4AJ_TAB3_rev	GTT TCA TCT GAT CTA CAT TTA	55,5	Sequenzierung P1
AP_4AJ_bp_fwd	TTT CCA ATG CAC TCC CTT AGA CAG	55,5	Sequenzierung P1
AP_4AJ_bp_rev	CCT TTT CAA ATT GGC TTC CTT C	55,5	Sequenzierung P1
RW013_Loc1_forw	ACC TAT ATG TTC CCT CTG TTA CC	57,0	Sequenzierung P2
Rwo13_Loc1_rev	GGC ATC TTA CCA CCT CTG A	57,0	Sequenzierung P2
RW013_Loc2_Forw	TTC CCA ATT TCA ATC CTC ATA	50,3	Sequenzierung P2
RW013_Loc2_rev	GAA ATC CAC TCT GCC AAC AC	50,3	Sequenzierung P2
RW013_Loc3_forw	TAG TAC ATG GCA TTA TTT ATT GAC	57,0	Sequenzierung P2
RW013_Loc3_rev	ATG GCA GTT GTA GGT AGA GAA GTA	57,0	Sequenzierung P2
RW013_Loc4_forw	CAC CTT GCA TAT GGA AAA ACC	50,3	Sequenzierung P2
RW013_Loc4_rev	TGA CTA GAC CAA TAA CAG GAT	50,3	Sequenzierung P2
FTHL17_bp_fwd	GAT TCC CAG TCT ACC CAC AC	55,8	Sequenzierung P3
FTHL17_wt_rev	AGT CCA AGT CCT TTT ATT TAC ATT	55,8	Sequenzierung P3
IL1RAPL1_bp_rev	AAG CAA TAC CTA AAA CTG AAC TAC	55,8	Sequenzierung P3
IL1RAPL1_wt_fwd	ATG AGA AAT GAC ACT GGG GAA ATA	55,8	Sequenzierung P3
SRY_1s	GTT GAG GGG GTG TTG AG	65,0	SRY-Amplifikation
SRY_1a	CTG AGT TTC GCA TTC TGG GAT TCT	65,0	SRY-Amplifikation

7.7 qPCR-Primer

Primername	Sequenz (5'→3')	Effizienz
NR0B1_UPL#22_Left	TCA CAA GCC ATG AAT CTA GAC C	2,07
NR0B1_UPL#22_Right	AGT AGG AAG AGG CTT GCA TGA	
Insulator_UPL74_lef	AAA GTG AGC ATT AAT CAG AAT CAA AAT	2,04
Insulator_UPL74_righ	TTT TAA AAG CCT CCT TAG TGT GC	
ZNF80Ref_SYBR_Fwd	CTG TGA CCT GCA GCT CAT CCT	1,98
ZNF80Ref_SYBR_Rev	TAA GTT CTC TGA CGT TGA CTG ATG TG	
GPR15Ref_SYBR_Fwd	GGT CCC TGG TGG CCT TAA TT	2,08
GPR15_SYBR_Rev	TTG CTG GTA ATG GGC ACA CA	
IL1RAPL1_UPL47_Fwd	CTG TAT GAA AGT ATC CAT CTC ACT GAC	2,00
IL1RAPL1_UPL47_Rev	GCT TTT TCA AAA TAC TTC ATC TTG G	
MAGEB1_UPL14_Fwd	AAA TGC CAA GGT GAG GAA AA	2,07
MAGEB1_UPL14_Rev	CTG AGC TCT CAG TGG ATG TTG T	
GK_UPL75_fwd	TGG TAA AGG CAA AAG TGT GGA	2,03
GK_UPL75_rev	CCA CCA TCC TCT CTC AGG AC	
TAB3_UPL48_Fwd	TCC ACA GCA TCA AGT GCA A	1,98
TAB3_UPL48_Rev	TGG GGT GGA GTA GTT GAA GG	
CXorf21_UPL11_fwd	GTG AAG GGG AAT TAA GGG CTA	2,04
CXorf21_UPL11_rev	GAG ATT TCT CCA TGA TTC ACA CA	

7.8 Cycler-Programme

Tabelle I. Programm: Amplifikations-PCR auf Eppendorf Master Cycler Gradient

	Temperatur (°C)	Zeit (s)	Zyklen
Aktivierung	95	180	1
Denaturierung	95	30	34
Primerhybridisierung	*	30	
Elongation	72	60	
Abschließende Elongation	72	300	1
Kühlung	8	Bis Entnahme	

* Temperatur abhängig von verwendeten Primern (siehe Annealingtemperaturen S. 110). Das Amplifikations-PCR-Programm wurde auch für die Gradienten-PCR benutzt, welche mit zehn Temperaturen im Bereich von 60 ± 10 °C erfolgte.

Tabelle II. Programm: Primerverdau auf Eppendorf Mastercycler Nexus X2e

	Temperatur (°C)	Zeit (min)
Primer- & Nukleotidverdau	37	15
Enzymdeaktivierung	85	15
	10	Bis Entnahme

Tabelle III. Programm: Sequenzierungsreaktions-PCR („HotDye“) auf Eppendorf Mastercycler Nexus X2e

	Temperatur (°C)	Zeit (s)	Zyklen
Aktivierung	96	300	1
Denaturierung	98	10	25
Primerhybridisierung	60	90	
Elongation	50	90	
	50	60	1
Kühlung	10	Bis Entnahme	

Tabelle IV. Programm: qPCR auf Roche LightCycler 96

	Temperatur (°C)	Zeit (s)	Zyklen
Preinkubation	96	180	1
3 Schritt -Amplifikation	95	10	40
	60	20	
	72	30	
Schmelzen	95	5	1
	65	60	
	97	1	
Abkühlen	37	30	1

7.9 Abbildungsverzeichnis

Abb. 1: Entwicklung der männlichen und weiblichen inneren Geschlechtsorgane aus gemeinsamer indifferenten Anlage	3
Abb. 2: DSD-Klassifikation.....	5
Abb. 3: Vereinfachte Darstellung der Geschlechtsdetermination im Menschen/Maus	9
Abb. 4: Sox9-Aktivierung im Rahmen der Hodenentwicklung	10
Abb. 5: Erstellung einer DNA-Fragmentbibliothek für das Whole Genome Sequencing.....	18
Abb. 6: Brückenamplifikation in der Flusszelle	19
Abb. 7: Kopizahlanalyse von P2 mittels aCGH und WGS.....	39
Abb. 8: Ermittlung der Strukturellen Variation in P1	41
Abb. 9: Ermittlung der strukturellen Variation in P2.....	42
Abb. 10: t-SNE zum Vergleich der Expressionsdaten von P1 und P2 mit gesunden Kontrollgeweben.....	45
Abb. 11: Sankey-Diagramm zur Veranschaulichung der Screening-Ergebnisse.....	47
Abb. 12: qPCR Ergebnis der Kopiezahl verschiedener Gene am Xp21.2 locus	47
Abb. 13: ArrayCGH Ergebnis von P3 in der Region von <i>NROB1</i>	51
Abb. 14: Ermittlung der Strukturellen Variarion in P3	51
Abb. 15: Isolator-Bindungsstellen in minimaler Überlappungsregion.....	53
Abb. 16: TAD-Konformation in verschiedenen menschlichen Zelllinien am <i>NROB1</i> -Locus	54
Abb. 17: GeneHancer Promotoren und Enhancer stromaufwärts von <i>NROB1</i>	56
Abb. 18: Auswirkung struktureller Variationen auf mögliche DNA-Faltung in P1 bis P3 im Vergleich zur regulären Anordnung.....	59
Abb. 19: Homologievergleich zwischen Maus und Menschen im Intron von <i>5430427019Rik</i> beziehungsweise <i>CXorf21</i>	69
Abb. 20: Gegenüberstellung bisher bekannter Xp21.2-Kopiezahlvariationen	76
Abb. I. Im IGV dargestellte WGS-Daten der Patient:in 1 (A) und einer Kontrollprobe (B)	92
Abb. II. Darstellung der strukturellen Variationen der Patient:innen 1 (B) und 2 (C) im Vergleich zur Normalsequenz (A) inklusive Positionsangaben der chromosomalen Bruchpunkte.....	93
Abb. III. Expressionsvergleich ausgewählter Gene der Geschlechtsentwicklung.....	94
Abb. IV. Gene Set Variation Analysis aller ‚MsigDB hallmark gene sets‘	95
Abb. V. SRY-PCR von fünf im qPCR-Screening auffälligen Patient:innen.....	96
Abb. VI. ArrayCGH-Ergebnis der Mutter von P3 in der Region von <i>NROB1</i>	97
Abb. VII. Mögliche TAD-Konformationen im Bereich des <i>NROB1</i> -Gens.....	98
Abb. VIII. Dnase I-Hypersensitivität im Bereich des <i>CXorf21</i> -Gen	99

7.10 Tabellenverzeichnis

Tab. 1. Heterozygote missense SNVs von Patient:in 1 in bekannten DSD-Genen	35
Tab. 2. Heterozygote missense SNV von Patient:in 2 in bekanntem DSD-Gen <i>ZFPM</i>	37
Tab. 3. Heterozygote missense SNV von Patient:in 3 in bekanntem DSD-Gen <i>ESR2</i>	50

8. Danksagungen

An dieser Stelle möchte ich mich bei allen bedanken, die wesentlich zum Gelingen dieser Arbeit beigetragen haben. Zunächst gilt mein Dank Prof. Dr. Olaf Hiort, der dieses spannende Forschungsvorhaben angeregt und intensiv sowie kritisch begleitet hat. Zahlreiche Gespräche und Diskussionen mit ihm haben diese Arbeit mitgeprägt. Mein besonderer Dank gilt Dr. Ralf Werner, der mir mit seinem Wissen, seiner Erfahrung und Engagement unermüdlich zur Seite stand. Er hat mich hervorragend unterstützt, immer wieder bei der Arbeit bestärkt und dabei viel Zeit und Herzblut in das Projekt gesteckt.

Außerdem bedanke ich mich bei Prof. Dr. Hauke Busch und Dr. Axel Künstner vom Lübecker Institut für Experimentelle Dermatologie, sowie Prof. Dr. Frank Kaiser vom Institut für Human-genetik der Universität Essen für den angeregten wissenschaftlichen Austausch, sowie ihre Unterstützung durch Expertise und Ausstattung. Von entscheidender Bedeutung für diese Arbeit war außerdem die Zusammenarbeit mit anderen Standorten der DSD-Versorgung. Daher gilt mein Dank einerseits den am Forschungsprojekt teilnehmenden Patient:innen und Angehörigen, andererseits den Kolleg:innen der Christian-Albrechts-Universität/Kiel, der Uni-versität Campinas/Sao Paulo und dem All India Institute Of Medical Sciences/New Delhi durch die die Patient:innen zum Teil betreut werden.

Mein Dank gilt darüber hinaus Karen Rieck, für ihre erfahrene und geduldige Unterstützung im Labor, sowie ihr stets offenes Ohr. Die Zeit im Labor war außerdem geprägt durch die angenehme Atmosphäre und wertschätzende Zusammenarbeit mit den Mitarbeiter:innen der AG Zechel und meiner Mitdoktorandin Anna aus der AG Hiort. Bei allen möchte ich mich für die schönen gemeinsamen Monate bedanken.

Drei Jahre sind eine lange Zeit, in der viel passieren kann – in der Forschung, aber auch im persönlichen Leben. Daher gilt der größte Dank meiner Familie und meinen Freunden, die mich in diesen drei Jahren unterstützt und begleitet haben. Sie haben die schönen und an-strengenden Augenblicke mit mir geteilt, mir aber auch ganz praktisch geholfen z.B. durch kritisches Lektorat.

9. Lebenslauf

Persönliche Angaben

Vor- und Mittelname Jakob Arnold
Nachname Meinel
Geburtsdatum 09.11.1995
Geburtsort Flensburg



Studium und Forschung

Studiums der Humanmedizin an der Universität zu Lübeck	10/2015 bis voraussichtlich 12/2022
Studentische Hilfskraft zur Auswertung von Studien im Hormonzentrum für Kinder und Jugendliche, MVZ Dr. Eberhard & Partner, Dortmund (3 Poster, 1 Peer-Review Artikel)	Seit 01/2016
Erster Abschnitt der Ärztlichen Prüfung (Abschlussnote 1,0)	09/2017
Doktorand in der Sektion für Pädiatrische Endokrinologie und Diabetologie an der Universität zu Lübeck (2 Vorträge, 3 Poster, 2 Peer-Review Artikel)	08/2018-08/2021
Versuche und Datenauswertung	08/2018 – 10/2019
Verfassen der Dissertation	03/2020 – 10/2020
Anmeldung der Dissertation	02/2021
Studentisches Mitglied im Senatsausschuss Medizin (SAM)	10/2018 – 9/2021
Zweiter Abschnitt der Ärztlichen Prüfung (Abschlussnote 2,0)	04/2021
Dritter Abschnitt der Ärztlichen Prüfung (Abschlussnote 1,0)	11/2022

Vorträge und Publikationen

Vorträge

Meinel J., R. Werner, O. Hiort. Von Dreiecken und Schleifen oder Warum die Sequenz nicht alles ist. [†]

3. Kiel-Lübeck-Treffen der Kinderkliniken des UKSH, Lübeck (2019)

Meinel, J., G. Dwivedi, P. M. Holterhus, O. Hiort, R. Werner. qPCR screening for Xp21.2 copy number variations in patients with elusive aetiology of 46,XY DSD. [†]

ESPE 2019, Wien, Österreich

Poster

Meinel, J., U. Finckh, A. Schuster, T. Haverkamp and A. Richter-Unruh. Establishment of Clinical and Lab Algorithms for the Identification Carriers of Mutations in CYP21A2 – A Study of 768 Children and Adolescents. Horm Res Paediatr 86(suppl 1): 127.

ESPE 2016, Paris, Frankreich

Meinel, J., K. Kühle, E. Jürgens and A. Richter-Unruh. Communication is the Key! Findings from transition of Adolescents with Growth Hormone Deficiency (GHD) in Germany. Horm Res Paediatr 88(suppl 1): 462.

IMPE 2017, Washington D.C., USA

Meinel, J., N. Grossmüller and A. Richter-Unruh. Awareness is the Key: Heavy Delay in Diagnosis of 17-β-Hydroxysteroid-Dehydrogenase III Deficiency (17bHSD3D) and other Insights and Conclusions from a Cohort of ten 17bHSD3D Patients in Germany. Horm Res Paediatr 90(suppl 1): 558.

ESPE 2018, Athen, Griechenland

dos Santos A. P.* , **J. Meinel,***, C. d. S. C. Piveta, J. G. R. de Andrade, H. Fabbri-Scallet, V. L. Gil-da-Silva-Lopes, G. Guerra-Junior, A. Kuentner, F. J. Kaiser, P. M. Holterhus, O. Hiort, H. Busch, A. T. Maciel-Guerra, M. P. de Mello and R. Werner. An approach for identifying and evaluating Xp21.2 copy number variations in Patients with elusive aetiology of 46,XY gonadal dysgenesis. [†]

7th I-DSD Symposium 2019, Sao Paulo, Brasilien

dos Santos A. P.* , **J. Meinel,***, C. d. S. C. Piveta, J. G. R. de Andrade, H. Fabbri-Scallet, V. L. Gil-da-Silva-Lopes, G. Guerra-Junior, A. Kuentner, F. J. Kaiser, P. M. Holterhus, O. Hiort, H. Busch, A. T. Maciel-Guerra, M. P. de Mello and R. Werner. Context is the Key: A New Look at Dosage Sensitive Sex Locus Xp21.2. †

7th I-DSD Symposium 2019, Sao Paulo, Brasilien

Meinel, J., G. Dwivedi, P. M. Holterhus, O. Hiort and R. Werner. qPCR screening for Xp21.2 copy number variations in patients with elusive aetiology of 46,XY DSD. Horm Res Paediatr 91(suppl 1): 98. †

ESPE 2019, Wien, Österreich

Peer-Review Artikel

Meinel, J., T. Haverkamp, F. Wunsche and A. Richter-Unruh (2021). Establishment of Clinical and Lab Algorithms for the Identification of Carriers of Mutations in CYP21A2 - A Study of 365 Children and Adolescents. *Exp Clin Endocrinol Diabetes* 129(7): 492-499.

Meinel, J.*, V. Yumiceba*, A. Künstner, K. Schulz, N. Kruse, F. J. Kaiser, P. M. Holterhus, A. Claviez, O. Hiort, H. Busch, M. Spielmann, R. Werner (2022). Disruption of the topologically associated domain at Xp21.2 is related to 46,XY gonadal dysgenesis. *Journal of Medical Genetics*. †

Francese-Santos, A.P., **J. Meinel**, C. d. S. C. Piveta, J. G. R. de Andrade, B. A. Barros, H. Fabbri-Scallet, V. L. Gil-da-Silva-Lopes, G. Guerra-Junior, A. Künstner, H. Busch, O. Hiort, A. T. Maciel-Guerra, R. Werner and M. P. de Mello (2022) A Novel Look at Dosage-Sensitive Sex Locus Xp21.2 in a Case of 46,XY Partial Gonadal Dysgenesis without NR0B1 Duplication. *Int J Mol Sci*, 24(1). †

Pre-Print Artikel

dos Santos A. P.* , **J. Meinel,***, C. d. S. C. Piveta, J. G. R. de Andrade, H. Fabbri-Scallet, V. L. Gil-da-Silva-Lopes, G. Guerra-Junior, A. Kuentner, F. J. Kaiser, P. M. Holterhus, O. Hiort, H. Busch, A. T. Maciel-Guerra, M. P. de Mello and R. Werner (2020). Disruption of the topological associated domain at Xp21.2 is related to gonadal dysgenesis: A general mechanism of pathogenesis. medRxiv: 2020.2003.2025.20041418. †

*geteilte Erstautorschaft

† im Rahmen der Doktorarbeit