

Einfluss genetischer Varianten auf die Bindung von Transkriptionsfaktoren in der kraniofazialen Entwicklung

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Hohen Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Katharina Laura Mirja Ruff

aus Bonn

2023

Angefertigt mit der Genehmigung
der Medizinischen Fakultät der Universität Bonn

1. Gutachterin: Dr. rer. nat. Kerstin U. Ludwig
2. Gutachter: Prof. Dr. Benjamin Odermatt

Tag der Mündlichen Prüfung: 09.08.2023

Aus dem Institut für Humangenetik
Direktor: Prof. Dr. med. Markus Nöthen

Für meine Familie, Daniel, Kristina und Kimberly

Inhaltsverzeichnis

Abkürzungsverzeichnis	5
1. Deutsche Zusammenfassung	8
1.1 Einleitung	8
1.2 Material und Methoden	11
1.3 Ergebnisse	17
1.4 Diskussion	20
1.5 Zusammenfassung	23
1.6 Literaturverzeichnis der deutschen Zusammenfassung	24
2. Veröffentlichung	35
Abstract	35
Introduction	35
Results	36
Discussion	39
Methods	41
References	42
3. Danksagung	45

Abkürzungsverzeichnis

ADCY9	Adenylyl cyclase type 9
ARID3B	AT-rich interactive domain-containing protein 3B
ARHGAP29	Rho GTPase Activating Protein 29
BMP4	Bone morphogenetic protein 4
ChIP	Chromatinimmunopräzipitation
ChIP-seq	Chromatinimmunopräzipitation gefolgt von einer Hochdurchsatzsequenzierung
CL/P	<i>Cleft lip with or without cleft palate</i> , Lippenspalten mit oder ohne Gaumenspalte
CRE	Cis-regulatorisches Element
CS15	Carnegiestadium 15
DNA	<i>Desoxyribonucleic acid</i> , Desoxyribonukleinsäure
EMT	<i>Epithelial mesenchymal transition</i> , epithelial-mesenchymale Transition
eQTL	<i>Expression Quantitative Trait Locus</i>
FDR	<i>False discovery rate</i> , Falscherkennungsrate
FILIP1L	Filamin A interacting protein 1 like
GWAS	<i>Genomewide association study</i> , genomweite Assoziationsstudie
GREM1	Gremlin 1
HEPM	<i>Human embryonal palatal mesenchyme</i> , humanes embryonales Gaumenmesenchym
IRF6	Interferon regulatory factor 6

KLF4	Kruppel-like factor 4
LD	<i>Linkage disequilibrium</i> , Kopplungsungleichgewicht
MAF	<i>Minor allele frequency</i> , Frequenz des selteneren Allels
MAFB	V-maf musculoaponeurotic fibrosarcoma oncogene homolog B
MEE	<i>Medial edge epithelia</i> , Epithel der medialen Kante
MSX1	Homeobox protein MSX-1
mRNA	<i>Messenger RNA</i> , Boten-RNA
NGS	<i>Next generation sequencing</i> , Hochdurchsatzsequenzierung
NOG	Noggin
nsCL/P	Nicht-syndromale CL/P
p.c.	post conceptionem
RNA	<i>Ribonucleic acid</i> , Ribonukleinsäure
SNP	<i>Single nucleotide polymorphism</i> , Einzelnukleotid-Polymorphismus
SPRY2	Sprouty homolog 2
TAD	<i>Topologically associating domain</i> , topologisch assoziierte Domäne
Tbx1	T-box protein 1
Tbx3	T-box protein 3
TF	Transkriptionsfaktor
TFAP2A	Transcription factor AP-2 alpha
TFBS	Transkriptionsfaktorbindestelle
TGFB	Transforming growth factor beta
TPM1	Tropomyosin alpha-1
TP63	Transformation-related protein 63

3'RNA-Seq

3'-mRNA-Sequenzierung

1. Deutsche Zusammenfassung

1.1 Einleitung

Lippenspalten mit oder ohne Gaumenspalte (engl. *cleft lip with or without cleft palate*, CL/P) sind eine angeborene Fehlbildung des Mittelgesichts, mit einer weltweiten Prävalenz von 1 pro 600 Lebendgeborenen (Mossey et al., 2009). Betroffene benötigen häufig eine umfangreiche chirurgische Behandlung sowie eine audiologische, logopädische und zahnärztliche Betreuung, verbunden mit substanziellen Kosten für Gesundheitssysteme (Wehby et al., 2009). Ergänzend leiden Betroffene häufiger unter Mobbing im schulischen Umfeld und einer geringeren Zufriedenheit mit ihrem äußeren Erscheinungsbild, im Vergleich zu nicht-betroffenen Individuen (Al-Namankany und Alhubaishi, 2018; Lorot-Marchand et. al., 2015; Marcusson et. al., 2002; Mullens et. al., 2019; Wahl et. al., 2015). In rund 30 % der Fälle treten CL/P im Rahmen von einem der mehr als 330 beschriebenen Syndrome auf. 70% der Patienten haben eine isolierte, nicht-syndromale CL/P (nsCL/P) (Leslie und Marazita 2015; Tolarova et al., 2015). Letztere ist Gegenstand der hier veröffentlichten Dissertation.

Die Ätiologie von nsCL/P ist multifaktoriell. Sie setzt sich zusammen aus einem komplexen Zusammenspiel von Umweltfaktoren und Genen (Dixon et al., 2011). In Studien wurde ein Effekt von Umwelteinflüssen wie mütterlichem Rauchen, mütterlichem Übergewicht und einem mütterlichen Folsäuremangel beschrieben, diese sind jedoch noch nicht konkludent (Jackson et. al., 2016; Jahanbin et. al., 2018; Little et. al., 2004; Shi et. al., 2007; Stott-Miller et. al., 2010;). Basierend auf Stammbaumanalysen und Zwillingsstudien, wird von einem starken genetischen Einfluss ausgegangen. Dieser wurde auf bis zu 90% beziffert (Grosen et. al., 2010; Grosen et. al., 2011). Mithilfe von genomweiten Assoziationsstudien (engl. *genome-wide association studies*, GWAS) konnten bislang 45 genetische Loci (d.h., chromosomale Abschnitte) identifiziert werden, an denen Einzelnukleotid-Polymorphismen (engl. *single nucleotide polymorphisms*, SNPs) eine Assoziation mit nsCL/P aufwiesen (Beaty et. al., 2010; Birnbaum et. al., 2009; Grant et. al., 2009; Leslie et. al., 2016; Leslie et. al., 2017; Ludwig et. al., 2012; Ludwig et. al., 2016; Ludwig et. al., 2017; Mangold et. al., 2010; Moreno et. al., 2009; Rahimov et. al., 2008; Sun et. al., 2015; Welzenbach et. al., 2021; Yu et. al., 2017). Eine

Schwierigkeit in der Aufarbeitung des biologischen Effektes dieser genetischen Varianten bei multifaktoriellen Erkrankungen ist, dass ein Großteil dieser assoziierten Risiko-Loci in nicht-kodierenden Regionen des Genoms verortet ist (Thieme und Ludwig, 2017). Darüber hinaus gilt es, die jeweils kausalen Varianten von nicht-kausalen Varianten zu unterscheiden, welche im Kopplungsungleichgewicht (engl. *linkage disequilibrium*, LD) stehen, bei dem eine erhöhte Wahrscheinlichkeit einer gemeinsamen Vererbung besteht und die nur statistisch zur selben Assoziation beitragen (Thieme und Ludwig, 2017). Im Rahmen von funktionellen *in-vitro* und *in-vivo* Studien wird versucht, gen-regulatorische Effekte der assoziierten Varianten aufzudecken. Diese Effekte beruhen unter anderem auf einer Beeinflussung der Genexpression durch sogenannte cis-regulatorische Elemente (CRE), wie Enhancer und Silencer. Aufgrund epigenetischer Mechanismen wie z. B. Histonmethylierung und Histonacetylierung und deren Kombination sind in verschiedenen Zellen und Zeitpunkten unterschiedliche CRE aktiv mit entsprechend variablem Transkriptom und Proteom (Birney et. al., 2007). Hierdurch kann eine Differenzierung von Organismen erzielt werden, da aus variablen transkribierten Genen und nachfolgend synthetisierten Proteinen verschiedener Zellen unterschiedliche Phänotypen resultieren (Mack und Nachmann, 2017).

Risikovarianten in nicht-kodierenden Regionen des Genoms können zu einer veränderten Bindung von Transkriptionsfaktoren (TF) führen. TF beeinflussen über ihre Bindung an CRE die Expression von Genen und genregulatorische Netzwerke (Degtyareva et. al., 2021; Abb.1). Dieser Mechanismus wurde auch bereits an einzelnen Loci für nsCL/P gezeigt: So wurden in einer Studie von Liu et. al. (2017) die Effekte von SNPs eines nsCL/P GWAS-Risikolocus am Locus Chr. 1p22 untersucht. Es wurde gezeigt, dass Risikoallele mit nsCL/P assoziierter SNPs an Chr. 1p22, innerhalb von Enhancern des nsCL/P Kandidatengens Rho GTPase Activating Protein 29 (*ARHGAP29*), zu einer veränderten Bindung der Transkriptionsfaktoren V-maf musculoaponeurotic fibrosarcoma oncogene homolog B (MAFB) und Kruppel-like factor 4 (KLF4) führen. Hierdurch wurde die Expression von *ARHGAP29* in humanen embryonalen Gaumenzellen (engl. *human embryonal palatal mesenchyme*, HEPM) beeinflusst. Rahimov et. al. (2008) zeigten für den nsCL/P GWAS Risikolocus an Chr. 1q32.2, dass das A-Allel des SNPs rs642961 an diesem Locus in Individuen mit nsCL/P

überrepräsentiert ist und dieses Allel zu einer veränderten Bindung des TF Transcription factor AP-2 alpha (TFAP2A) an einem mit rs642961 überlappendem Enhancer des nsCL/P Kandidatengens Interferon regulatory factor 6 (*IRF6*) führt.

Ziel der vorliegenden Arbeit war die systematische Analyse der Binderegionen eines TF mit einer Schlüsselrolle in der kraniofazialen Entwicklung. Da diese Effekte spezifisch für Entwicklungszeitpunkte und Zelltypen sind, wurde die Studie mit HEPM-Zellen durchgeführt, die ein adäquates Modell des embryonalen Gaumens vor der Fusion der Gaumenfortsätze repräsentieren (Dhulipala et. al., 2004; Greene et. al., 2015; Yoneda und Pratt 1981). Mittels Literaturrecherche und Genexpressionsanalyse in HEPM-Zellen konnten wir TFAP2A als einen entwicklungsrelevanten, in unserem Zellmodell exprimierten, TF identifizieren. Zur Generierung eines Datensatzes mit TFAP2A Binderegionen in HEPM-Zellen führten wir eine Chromatinimmunopräzipitation gefolgt von einer Hochdurchsatzsequenzierung (ChIP-seq) durch. Eine zentrale Fragestellung dieser Arbeit war, ob TFAP2A in HEPM-Zellen an, oder in der Nähe von, bekannten nsCL/P Risikoloci oder bisher noch nicht signifikant assoziierten Loci bindet und ob dort allelspezifische Bindeeffekte auftreten. Zu diesem Zweck führten wir einen Abgleich der TFAP2A ChIP-seq Daten aus HEPM-Zellen mit den Daten einer GWAS Meta-Analyse (Ludwig et. al., 2017) durch, welche Informationen über ca. 8,01 Millionen SNPs und deren Assoziation mit nsCL/P enthält. Zudem erfolgte eine Integration der TFAP2A ChIP-seq Daten aus HEPM-Zellen mit aktiven regulatorischen Regionen in menschlichem kraniofazialen Gewebe des Carnegie Stadiums 15 (CS15; Wilderman et. al., 2018). Die Carnegie Stadien stellen eine Einteilung der Embryonalperiode in insgesamt 23 Entwicklungsstadien dar (O’Rahilly und Müller, 2010). CS15 beschreibt den Entwicklungsstand um den 36. Tag post conceptionem (p.c.) vor der Bildung des sekundären Gaumens in der 6. bis 8. Woche p.c. und stimmt mit dem gestationalen Alter des weiblichen Abortes, aus dem die verwendeten HEPM-Zellen asserviert wurden, näherungsweise überein (Yoneda und Pratt 1981). Außerdem wurde eine Integration mit Gen-Ontologie-Daten zu biologischen Prozessen, humanen Phänotypen und Maus-Einzelgen-Knockout-Phänotypen durchgeführt.

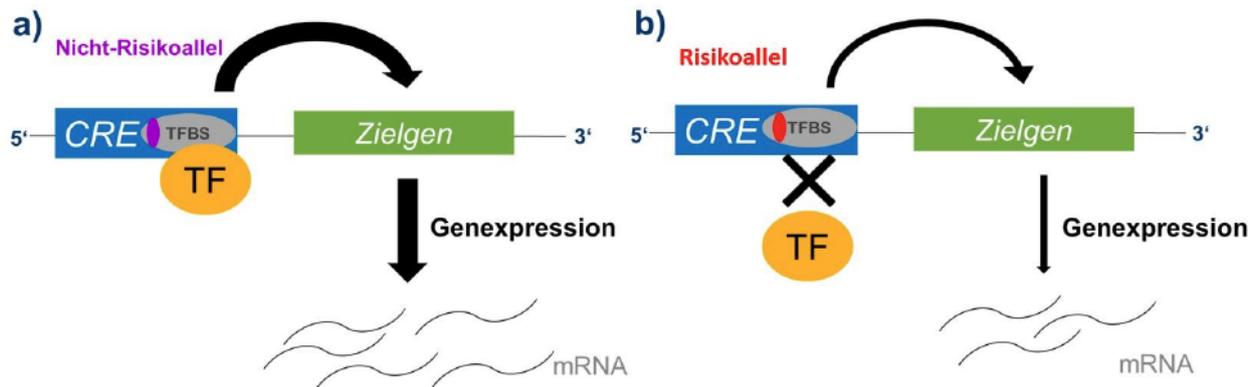


Abb. 1: Veränderte Bindung von Transkriptionsfaktoren (TF) als möglicher Effekt von assoziierten Einzelnukleotid-Polymorphismen (engl. *single nucleotide polymorphisms*, SNPs) in nicht-kodierenden Regionen des Genoms: (a) Die Bindung eines TF an Transkriptionsfaktorbindestellen (TFBS) cis-regulatorischer Elemente (CRE) der DNA kann regulatorische Effekte auf die Expression von Genen und die Herstellung von Boten-RNA (engl. *messenger RNA*, mRNA) und Genprodukt haben. (b) Ein Risikoallel eines SNPs in einem CRE kann im Vergleich zu einem nicht mit einer Krankheit assoziierten Allel eines SNPs (Nicht-Risikoallel) zu einer veränderten Bindungsaffinität eines Transkriptionsfaktors führen.

1.2 Material und Methoden

Im Rahmen der vorliegenden Studie wurden Genexpressionsdaten aus HEPM-Zellen, TFAP2A ChIP-seq Peaks in HEPM-Zellen und HeLaS3 Zellen (Gerstein et. al., 2012), epigenetische Daten aus CS15 embryonalem Gewebe (Wilderman et. al., 2018), Gen-Ontologie-Daten (McLean et. al., 2010), eine nsCLP GWAS Metaanalyse (Ludwig et al. 2017) und nsCLP Risikoloci beinhaltende topologisch assoziierten Domänen (engl. *topologically associating domain*, TAD) (Welzenbach et. al., 2021) miteinander integriert. Eine Übersicht des Studienablaufs findet sich in Abb. 2.

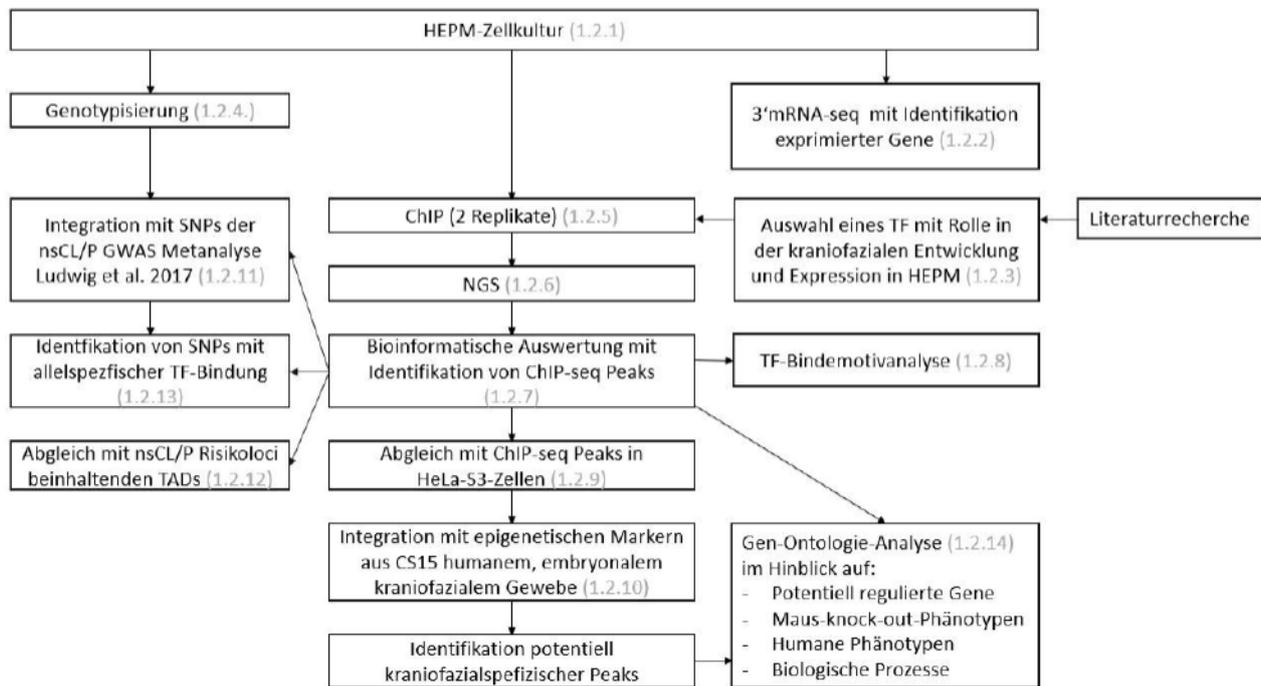


Abb. 2: Vereinfachte Darstellung des Studiendesigns. Verweise zu den korrespondierenden Sektionen des Kapitels „Material und Methoden“ in grauen Klammern. Abkürzungen entsprechend den Verweisen im Text und dem Abkürzungsverzeichnis.

1.2.1 Zellkultur

HEPM-Zellen (ATCC Cat# CRL-1486) wurden über 5 Tage in DMEM-Dulbecco's Modified Eagle Medium (high Glucose) mit 10% hitzeinaktiviertem fetalem Rinderserum und 1 % Penicillin/Streptomycin bei 37°C und 5% CO₂ kultiviert. Anschließend erfolgte die Extraktion der DNA und RNA (QIAGEN DNeasy® Blood & Tissue kit, QIAGEN RNeasy® kit) gemäß Herstellerprotokoll. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.2 Generierung von HEPM Genexpressionsdaten mittels 3'-mRNA-Seq

Nach einem Anhang von Index-Sequenzen für eine parallele Sequenzierung mehrerer RNA-Proben aus HEPM-Zellen (Library preparation, Lexogen QuantSeq 3'mRNA-Seq Library Prep Kit) erfolgte eine Hochdurchsatzsequenzierung mit ca. 20 Millionen Reads

pro Probe (Illumina HiSeq2500). Es erfolgte eine bioinformatische Aufarbeitung der Sequenzierungsdaten, bestehend aus einer Qualitätskontrolle (FastQC v0.11.7, Wingett und Andrews, 2018; MultiQC v1.2, Ewels et. al., 2016), einer Entfernung von Adaptersequenzen (bbduk, BMap v37.44, Bushnell, 2019), einem Abgleich mit dem GRCh37/hg19 Referenzgenom (STAR Aligner v.2.5.2b, Dobin et. al., 2013), und einer quantitativen Auswertung der Reads pro Gen (featureCounts v1.5.1, Liao et. al., 2014; Referenzgenom: GRCh37.p13, Genome Reference Consortium, 2009). Die 3'm-RNA-Seq wurde in Triplikaten durchgeführt und die Zahl der Reads pro Gen wurden über die Replikate gemittelt. Es wurden exprimierte Gene (mindestens 5 3'mRNA-Seq Reads) und stark exprimierte Gene (mindestens 327 3'm-RNA-Seq Reads \geq dem oberen Quartil der exprimierten Gene) unterschieden. Zuletzt erfolgte eine Identifikation von TF unter den exprimierten Genen mittels eines Abgleiches mit der TF-Bindemotive enthaltenden JASPAR CORE vertebrate assembly Datenbank (2020, Fornes et. al., 2020).

1.2.3 Auswahl eines in HEPM exprimierten Transkriptionsfaktors

Es wurde mittels Literaturrecherche in der Pubmed Datenbank ein Transkriptionsfaktor mit Evidenz für eine Rolle in der kraniofazialen Entwicklung oder der Pathogenese orofazialer Spalten für die vorliegende Studie ausgewählt. Wir wählten einen, entsprechend unseren Genexpressionsdaten, in HEPM-Zellen exprimierten Transkriptionsfaktor aus. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.4 Genotypisierung von HEPM-Zellen

Zur Identifikation heterozygoter SNPs mit möglicher allelspezifischer Transkriptionsfaktorbindung in HEPM Zellen wurde eine Genotypisierung durchgeführt. Wir generierten Genotypinformationen aus HEPM-Zell-DNA mittels eines Illumina Infinium GSA v2.0 array und bioinformatischer Imputation (IMPUTE v2.3.2, Howie et. al., 2009; Referenzdatensatz: 1000 genomes phase 3 GRCh37/hg19, Fairley et. al., 2020). Zudem prüften wir die Genotypen einer Auswahl imputierter Varianten mittels Sanger Sequenzierung. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.5 Chromatinimmunopräzipitation (ChIP)

Wir führten zwei Replikate einer ChIP mit je ca. 4 Millionen HEPM-Zellen und einem polyklonalen TFAP2A Antikörper (Abcam Cat# ab52222) durch. Für die

Chromatinimmunopräzipitation wurde das Cell Signaling Technology® SimpleChIP®Enzymatic Chromatin IPKit #9003 genutzt und das Herstellerprotokoll wurde bis auf einzelne im Folgenden beschriebene Anpassungen befolgt. Nach einer DNA-Quervernetzung mittels Formaldehyd und Inkubation mit verschiedenen im Kit inkludierten Reagenzien auf Eis, erfolgte ein Chromatinverdau mit 0,5 µl Mikrokokkennuklease über 15 Min bei 37°C und eine Ultraschallfragmentierung mit einem Diagenode Bioruptor über 50 Zyklen mit je 30 Sekunden Schallexposition und 30 Sekunden Pause. Nach einer Kontrolle der DNA-Fragmentgröße mittels Gelelektrophorese erfolgte eine Inkubation von 25 µg DNA mit 0,005 µg des TFAP2A Antikörpers über Nacht bei 4 °C mit Rotation. Nach Gewinnung von Chromatinfragmenten mit Antikörperbindung mit Protein G Magnetic Beads und Revidierung der DNA-Quervernetzung wurden die Proben in Zentrifugationssäulengefäßen aufgereinigt. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.6 Hochdurchsatzsequenzierung der ChIP-Proben

Zur Vorbereitung der Proben und zum Anhängen von Indexsequenzen wurde das NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® entsprechend dem Herstellerprotokoll genutzt. Durch den Einsatz von Indizes konnte eine gleichzeitige Sequenzierung von HEPM-DNA-Proben nach ChIP und unbehandelten Proben als Kontrollen erfolgen (sog. Multiplex-Sequenzierung). Es erfolgte eine Hochdurchsatzsequenzierung (engl. *next generation sequencing*, NGS) mit einem Illumina MiSeq v2 Sequenzierer in der NGS Core-facility des Uniklinikums Bonn. Jedes Replikat wurde zweifach sequenziert. Es resultierten 20 Millionen Reads mit einer Länge von je 250 Basenpaaren in beiden DNA-Ableserichtungen pro Probe (zwei Replikate jeweils mit ChIP-Probe und Kontrollprobe). Für weitere Informationen siehe Ruff et. al. (2022).

1.2.7 Bioinformatische Auswertung der ChIP-seq

Die bioinformatische Auswertung der ChIP-seq umfasste eine Qualitätskontrolle der NGS Reads (FastQC, v0.11.7, Wingett und Andrews, 2018), eine Entfernung von Adaptersequenzen und eine Kürzung der Einträge auf 200 Basenpaare Länge

(Cutadapt v1.15. Martin, 2011). Die Reads wurden mit dem GRCh37/hg19 Referenzgenom abgeglichen und den jeweiligen genomischen Regionen zugeordnet (Bowtie 2 v2.3.4, Langmead et. al., 2012). Es erfolgte ein Abgleich der chromosomalen Regionen der ChIP-seq und Kontrollproben, um in der ChIP-Probe überrepräsentierte DNA-Regionen zu identifizieren (MACS2 v2.1.1, Zhang et. al., 2008). Regionen mit 5- bis 50-fach höherer Anreicherung in der ChIP-Probe als in der Kontrolle und einer Falscherkennungsrate (engl. *false discovery rate*, FDR) kleiner als 0,05 wurden als Regionen mit Transkriptionsfaktorbindung angesehen, im Folgenden als TFAP2A ChIP-seq Peaks bezeichnet. Für die weitere Analyse selektierten wir einen Datensatz aus TFAP2A ChIP-seq Peaks, welche sich in beiden Replikaten wiederfanden. Regionen, welche gehäuft als Artefakte in ChIP-seq Datensätzen auftauchen (Encode blacklist regions, Carroll et. al., 2014), wurden aus unserem Datensatz entfernt (BEDtools v2.27.0, Quinlan et. al., 2010). Für weitere Informationen siehe Ruff et. al. (2022).

1.2.8 Bindemotivanalyse

Zur Analyse des prädominanten Transkriptionsfaktorbindemotivs unseres Datensatzes werteten wir die Basensequenz +/- 50 Basenpaare um die Mittelpunkte der TFAP2A ChIP-seq Peaks aus (MEME ChIP v.5.1.0, Machanick et. al., 2011). Um zu validieren, dass dieses Motiv mit bisher bekannten Bindemotiven des TF übereinstimmt, führten wir einen Abgleich mit der TFBS Datenbank JASPAR core 2018 assembly (Khan et. al., 2018) durch (Tomtom motif comparison tool v.5.1.0, Gupta et. al., 2007).

1.2.9 Abgleich mit TFAP2A ChIP-seq Peaks in HeLaS3-Zellen

Um zu zeigen, dass die TFBS unseres Datensatzes tatsächliche Binderegionen von TFAP2A sind, führten wir einen Abgleich mit TFAP2A Binderegionen in einer humanen Zervixkarzinomzellreihe (HeLaS3-Zellen, Gerstein et. al., 2012) durch. Außerdem identifizierten wir im Rahmen dieser Analyse potenziell kraniofazial-spezifische TFAP2A Binderegionen, welche entsprechend nur in dem in dieser Studie generierten Datensatz und nicht dem HeLaS3-Datensatz präsent waren. Wenn mehr als 50% des TFAP2A ChIP-seq Peaks in HeLaS3 mit einem TFAP2A ChIP-seq Peak in HEPM-Zellen übereinstimmte, wurde dies als Überlappung gewertet. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.10 Abgleich mit epigenetischen Markern aus humanem, embryonalem, kraniofazialem Gewebe

Um potenziell aktive regulatorische Regionen zu identifizieren, integrierten wir unsere TFAP2A ChIP-seq Peaks in HEPM-Zellen mit Histonmarkierungen aktiver Enhancer und DNase1 Hypersensitivitätsregionen in humanem, embryonalem kraniofazialem Gewebe des CS15 (Wilderman et. al., 2018). Eine Überlappung wurde ab einem gemeinsamen Basenpaar angenommen. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.11 Abgleich mit nsCL/P GWAS SNPs

Die TFAP2A ChIP-seq Peaks wurden mit einer GWAS-Metaanalyse von Ludwig et. al. (2017) abgeglichen. Diese enthielt Assoziationsdaten für ca. 8,01 Millionen SNPs aus asiatischen und europäischen Erkrankten und Kontrollen mit einer Frequenz des selteneren Allels (engl. *minor allele frequency*, MAF) von $\geq 1\%$. SNPs mit einem Infoscore von mindestens 0,8 wurden in die Analyse eingeschlossen. Ein mögliches Kopplungsungleichgewicht, wurde mithilfe des Tools LD link (v.5, Einstellung: Alle Populationen, Machiela et. al., 2015) ermittelt. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.12 Analyse von topologisch assoziierten Domänen der GWAS loci

Wir untersuchten, ob SNPs, welche in TFAP2A ChIP-seq Peaks in HEPM-Zellen lagen, in TADs der 45 nsCL/P GWAS Risikoloci verortet waren. Topologisch assoziierte Domänen sind abgrenzbare, genetische Bereiche innerhalb derer, als funktionelle Einheiten, eine hohe regulatorische Aktivität stattfindet (Dixon et. al., 2012). Eine entsprechende Zusammenstellung von nsCL/P GWAS Risikoloci und TADs entnahmen wir einer vorausgegangenen Arbeit unserer Gruppe (Welzenbach et. al., 2021).

1.2.13 Identifikation von SNPs mit allelspezifischer TF Bindung

Mithilfe des Tools ABC (v.1.3., Bailey et. al., 2015) untersuchten wir die TFAP2A ChIP-seq Peaks in HEPM-Zellen auf eine Überrepräsentation bestimmter Allele von in HEPM-Zellen heterozygoten Varianten. Eine allelspezifische Bindung wurde bei einem binominalen p-Wert $< 0,05$ angenommen. Für weitere Informationen siehe Ruff et. al. (2022).

1.2.14 Gen-Ontologie

Wir nutzten das Genomic Regions Enrichment of Annotations Tool (GREAT, v4.0.4, McLean et al., 2010) um zu untersuchen, ob es eine Anreicherung von TFAP2A ChIP-seq Peaks an Regionen mit bestimmten, in den Gen-Ontologie-Daten gelisteten, humanen Phänotypen, Einzelgen-Knockout-Phänotypen der Maus oder biologischen Prozessen gab. Außerdem erfolgte mithilfe des Tools die Identifikation möglicher durch die Bindung von TFAP2A in ihrer Expression beeinflusster Gene. Hierzu wurden die Standardeinstellungen des Tools verwendet. Für weitere Informationen siehe Ruff et. al. (2022).

1.3 Ergebnisse

1.3.1 TFAP2A als ein in HEPM-Zellen exprimierter, kraniofazialer Transkriptionsfaktor

Nach bioinformatischer Auswertung einer 3'-mRNA-seq in HEPM-Zellen zeigten sich 14.508 exprimierte Gene und hierunter 350 für TF kodierende Gene (Siehe Supplementary Table S1, Fig. 1 in Ruff et. al., 2022). Zur weiteren Eingrenzung für die vorliegende Fragestellung relevanter TF erfolgte eine Literaturrecherche in der Pubmed-Datenbank. Wir priorisierten 22 TF durch Evidenz für i) eine kraniofaziale Malformation bei Mutation des entsprechenden Gens, ii) eine Schlüsselrolle in der Entwicklung des Gaumens basierend auf Tiermodellen oder bekannten genregulatorischen Netzwerken, oder iii) den Status eines Kandidatengens durch Lokalisation an einem nsCL/P Risikolocus. Darüber hinaus zeigten 6 von diesen 22 TFs eine starke Expression in HEPM-Zellen. In Zusammenschau von Literatur und Expressionsdaten wurde TFAP2A als in HEPM-Zellen stark exprimiertes Kandidatengen des nsCL/P Risikolocus Chr. 6p24 (Yu et. al., 2017) für die weiteren Analysen der vorliegenden Arbeit ausgewählt.

1.3.2 TFAP2A ChIP-seq Peaks nahe bekannter nsCL/P Kandidatengene

Die Daten der TFAP2A-ChIP-seq zeigten 3.989 und 5.820 TFAP2A ChIP-seq Peaks in den zwei durchgeführten Replikaten, und 2.845 in beiden Replikaten präsente TFAP2A ChIP-seq Peaks (FDR <5%, Supplementary Tables S3-S5, Fig, 2a aus Ruff et al.,

2022). Die weiteren Analysen wurden mit diesen 2.845 TFAP2A ChIP-seq Peaks durchgeführt. Eine Bindemotivanalyse dieser Regionen zeigte ein überrepräsentiertes, 15 Basenpaare umfassendes Motiv mit hoher Übereinstimmung mit bekannten TFAP2A Bindesequenzen der JASPAR core vertebrate assembly (Fig. 2b aus Ruff et. al., 2022, Khan et. al., 2018). Ergänzend resultierte eine Gen-Ontologie-Analyse, basierend auf der räumlichen Nähe von TFAP2A ChIP-seq Peaks zu Transkriptionsstartpunkten von Genen, in 3.470 möglicherweise durch die Bindung von TFAP2A in ihrer Expression beeinflussten Genen (Supplementary Table S5 aus Ruff et. al., 2022). Hierunter waren 10 nsCL/P Kandidatengene und unter diesen 8 Kandidatengene mit Expression in HEPM-Zellen (Table 1 aus Ruff et. al., 2022).

1.3.3 Identifikation spezifischer kraniofazialer TFAP2A ChIP-seq Peaks

Im Abgleich mit einem Datensatz aus humanen Zervixkarzinomzellen (HeLa-S3-Zellen, Gerstein et. al., 2012) erfolgte die Identifikation von 1.333 übereinstimmenden TFAP2A ChIP-seq Peaks in HEPM-Zellen und HeLa-S3-Zellen ($P_{\text{Chi2}} < 0,0001$, Supplementary Table S5, Fig.3 aus Ruff et. al., 2022). Diese Übereinstimmung unterstützte die Hypothese, dass es sich um tatsächliche Binderegionen von TFAP2A handelte. Zudem erfolgte die Integration mit Histonmarkern aus embryonalem kraniofazialem Gewebe der CS 15 (Wilderman et. al., 2018). Es zeigte sich eine Überlappung von 2.678 TFAP2A ChIP-seq Peaks in HEPM-Zellen mit DNase1-Hypersensitivitätsregionen und hierunter von 1.634 Regionen mit H3-K27ac-Histonstatus jeweils in embryonalem kraniofazialem Gewebe (jeweils $P_{\text{Chi2}} < 0,0001$, Supplementary Table S5, Fig. 3 aus Ruff et. al., 2022). Diese Marker wurden als Indiz für aktive Enhancer gewertet (Birney et. al., 2007). Von diesen 1.634 TFAP2A ChIP-seq Peaks zeigten 802 keine Überlappung mit TFAP2A ChIP-seq Peaks in HeLaS3-Zellen, sodass es sich um spezifische, kraniofaziale TFAP2A ChIP-seq Peaks handeln könnte.

1.3.4 Gen-Ontologie-Analysen weisen auf Relevanz von TFAP2A-Bindestellen in embryonalen Prozessen und Phänotypen im Kontext der kraniofazialen Entwicklung hin

Die Gen-Ontologie-Analyse aller 2.845 und der 802 potenziell für die kraniofaziale Entwicklung spezifischen TFAP2A ChIP-seq Peaks zeigte diverse relevante

Assoziationen (FDR q-Wert $<0,05$; Supplementary Tables S6-S11) im Hinblick auf murine Einzelgen-Knockout-Phänotypen (z.B. „abnormale Elevation der Gaumenanlagen“), humane Phänotypen (z.B. „abnormales Weichgewebe des Gesichts“) und biologische Prozesse (z.B. „Regulation der zellulären Reaktion auf einen transforming growth factor Stimulus“). Einige der überrepräsentierten Prozesse mit Bezug zu nsCL/P oder der kraniofazialen Entwicklung und zeigten entweder keine oder eine weniger signifikante Assoziation zum Gesamtset der 2.845 TFAP2A ChIP-seq Peaks, verglichen mit der Subgruppe der potenziell spezifischen kraniofazialen Peaks.

1.3.5 Keine signifikant assoziierten nsCL/P Risikovarianten in TFAP2A ChIP-seq Peaks in HEPM

Im Abgleich der 2.845 TFAP2A ChIP-seq Peaks in HEPM-Zellen mit SNPs einer GWAS Metaanalyse von Ludwig et. al. (2017) konnten wir 1.564 häufige Varianten innerhalb der Peaks identifizieren (Supplementary Table S13 aus Ruff et. al., 2022). Hierunter waren vier SNPs mit einem GWAS p-Wert $<0,001$ (rs463271 (22q11.21, p-Wert= $2,72 \times 10^{-4}$); rs12882215, rs7144455 (14q24.3, p-Wert= $3,30 \times 10^{-4}$ und p-Wert= $3,77 \times 10^{-4}$, $D'=1,0$; $r^2=0,85$); rs4774822 (15q21.3, p-Wert= $6,27 \times 10^{-4}$, Supplementary Table S13 aus Ruff et. al., 2022). Zudem wurden die TADs der 45 aktuell bekannten nsCL/P GWAS Risikoloci mit den 2.845 TFAP2A ChIP-seq Peaks abgeglichen (Welzenbach et. al., 2021). Innerhalb von 20 TADs waren 70 TFAP2A-ChIP-seq Peaks und 46 häufige Varianten lokalisiert. Keiner dieser SNPs zeigte eine signifikante oder nominale Assoziation mit nsCL/P. Es lagen mit rs672819 und rs447476 zwei Varianten innerhalb der 70 Peaks, die ein LD mit signifikant mit nsCL/P assoziierten Varianten an bekannten Risikoloci aufweisen. So lag ein Kopplungsungleichgewicht zwischen rs672819 und rs3753517 auf 1q32.1 ($D'=0,51$; $r^2=0,18$) und eines zwischen rs447476 und rs2303914 auf 2p25.1 ($D'=0,91$; $r^2=0,28$) vor.

1.3.6 Allelspezifische Bindung von TFAP2A in HEPM-Zellen

Durch die Genotypisierung der HEPM-Zellen mittels eines Arrays und genomischer Imputation identifizierten wir 461 heterozygote SNPs der GWAS Metaanalyse von Ludwig et. al. (2017) innerhalb der TFAP2A ChIP-seq Peaks in HEPM-Zellen. Die Analyse der TFAP2A ChIP-seq Peaks in HEPM-Zellen auf allelspezifische TFAP2A-Bindung an

diesen SNPs zeigte 18 SNPs mit Überrepräsentation eines Allels in den ChIP-seq Reads (binominaler p-Wert $<0,05$, Supplementary S14 aus Ruff et. al., 2022). Keiner dieser SNPs zeigte eine nominale oder statistisch signifikante Assoziation mit nsCL/P. Von den vier SNPs mit einem p-Wert $<0,001$ in der nsCL/P GWAS Metaanalyse von Ludwig et. al. (2017) war nur rs447476 heterozygot in HEPM-Zellen. Jedoch zeigte sich hier keine allelspezifische Bindung von TFAP2A. In der Gen-Ontologie-Analyse der 18 TFAP2A-ChIP-seq Peaks mit Varianten mit allelspezifischer Bindung zeigten sich keine für nsCL/P oder die kraniofaziale Entwicklung relevanten Gene, Phänotypen oder biologischen Prozesse überrepräsentiert.

1.4 Diskussion

Basierend auf bisheriger Evidenz kommt TFAP2A eine Schlüsselrolle in der kraniofazialen Entwicklung zu. Ein Knockout von *Tfap2a* in Mausmodellen führte zu einem Phänotyp mit Gesichtsspalten neben anderen Fehlbildungen (Green et. al., 2015; Schorle et. al., 1996). Das Branchio-Oculo-Faziale Syndrom, welches durch Deletionen im *TFAP2A*-Gen verursacht wird, führt zu Lippenspalten mit oder ohne Gaumenspalten und anderen Symptomen wie z.B. Hautdefekten (Milunsky et. al., 2008). Eine reduzierte Verfügbarkeit der Mikro-RNA miR-17-92, deren Transkription durch TFAP2A reguliert wird, führt zu einer gesteigerten Expression von T-box protein 1 (*Tbx1*) und von T-box protein 3 (*Tbx3*) im murinen embryonalen Mittelgesicht und zu Gaumenspalten mit vollständiger Penetranz (Wang et. al., 2013). Regulatorische Netzwerke in der kraniofazialen Entwicklung mit TFAP2A im Zentrum wurden im Rahmen von Multi-Omics-Ansätzen charakterisiert (Razaghi-Moghadam et. al., 2019).

In unserer Analyse konnten wir 2.845 TFAP2A ChIP-seq Peaks, und darunter 802 potenziell für die kraniofaziale Entwicklung spezifische TFAP2A ChIP-seq Peaks, identifizieren. Diese überlappten mit DNase1-Hypersensitivitätsregionen und H3K27ac Histonmarkierung in humanem embryonalem kraniofazialen Gewebe des CS15 und könnten somit aktive Enhancer während des relevanten Entwicklungszeitpunktes darstellen (Wilderman et. al., 2018; Birney et. al., 2007). TFAP2A ChIP-seq Peaks lagen in der Nähe der nsCL/P GWAS Kandidatengene Homeobox protein MSX-1 (*MSX1*, Yu

et. al., 2017), Sprouty homolog 2 (*SPRY2*, Ludwig et. al., 2012), ARHGAP29 (Beaty et. al., 2012), Filamin A interacting protein 1 like (*FILIP1L*, Beaty et. al., 2013), transformation-related protein 63 (*TP63*, Leslie et. al., 2017), Gremlin 1 (*GREM1*, Ludwig et. al., 2016), Tropomyosin alpha-1 (*TPM1*, Ludwig et. al., 2012), AT-rich interactive domain-containing protein 3B (*ARID3B*, Ludwig et. al., 2017), Adenylyl cyclase type 9 (*ADCY9*, Sun et. al., 2015) und *NOG* (Mangold et. al., 2010). Interessant waren hierunter insbesondere *MSX1* und *SPRY2*. Diese Gene waren in HEPM-Zellen stark exprimiert und es wurde in Studien bereits eine Expression und funktionelle Relevanz der beiden Gene im Gaumenmesenchym während der Gesichtsentwicklung gezeigt (Carroll et. al., 2020; Goodnough et. al., 2007; Matsumura et. al., 2011; Zhang et. al., 2002). Eine Bindung nahe *IRF6* und rs642961 ließ sich nicht replizieren und ist möglicherweise zelltypspezifisch. Regulatorische Effekte der TF Bindung könnten, vermittelt über eine Veränderung der Genexpression, einen Effekt auf Entwicklungsvorgänge haben. Ebenso zeigten sich einige mit der Entwicklung des Gesichtes verknüpfte biologische Prozesse und relevante humane und murine Phänotypen in der Gene-Ontologie-Analyse der TFAP2A ChIP-seq Peaks überrepräsentiert. Hervorzuheben ist hier eine mögliche Regulation des, für die kraniofaziale Entwicklung notwendigen, Transforming growth factor beta (TGFB) Signalwegs durch TFAP2A und eine Rolle in der epithelial-mesenchymalen Transition (engl. *epithelial mesenchymal transition*, EMT), in die auch TGFB involviert ist (Nakajima et. al., 2018). Die Fusionsregionen der Gaumenfortsätze bestehen aus mesenchymalen Zellen und einem Saum aus epithelialen Zellen (engl. *medial edge epithelia*, MEE), welcher ein zu frühes Zusammenwachsen verhindert. Vor der Bildung eines kontinuierlichen Gaumens kommt es zu einer Umwandlung der MEE-Zellen zu Mesenchymzellen im Rahmen der EMT und zu einer Apoptose und Migration der MEE-Zellen (Nakajima et. al., 2018). Da die Gen-Ontologie-Assoziationen auf der räumlichen Nähe der TFAP2A ChIP-seq Peaks und zu den Transkriptionsstartpunkten der Gene beruhen und dadurch räumlich weiter entfernte Effekte ausgeschlossen werden, sind funktionelle Studien, zur Validierung dieser Effekte notwendig.

Eine Integration der TFAP2A ChIP-seq Peaks mit SNPs der nsCL/P GWAS-Metaanalyse von Ludwig et. al. (2017) zeigte keine genomweit signifikant assoziierten

SNPs innerhalb der TFAP2A CHIP-seq Peaks in HEPM Zellen. Insgesamt zeigten nur vier SNPs eine relevante Assoziation. In der Genotypisierung der HEPM-Zellen zeigte sich ein heterozygoter Genotyp nur an einem dieser SNPs, jedoch keine allelspezifische Transkriptionsfaktorbindung. 18 heterozygote Varianten mit allelspezifischer TFAP2A-Bindung waren nicht signifikant mit nsCL/P assoziiert. Da nur ein Abgleich mit häufigen Varianten mit einer $MAF \geq 1\%$ der GWAS Meta-Analyse von Ludwig et. al. (2017) erfolgte, sind mögliche Effekte seltenerer Varianten in unserer Studie nicht abgebildet. Ein weiterer Ansatz, Varianten mit einem Effekt auf die Genexpressionen zu identifizieren sind sogenannte expression Quantitative Trait Locus (eQTL) Analysen, bei welchen Genexpressions- und Genotyp-Daten multipler Proben auf SNPs mit signifikanter Assoziation von Genotypen mit Genexpressionsleveln untersucht werden (Nica et. al., 2013). Durch die begrenzte Verfügbarkeit humaner embryonaler Gewebe des entsprechenden Entwicklungsstadiums ist eine Durchführung von eQTL-Studien bei der Erforschung von nsCL/P derzeit nicht möglich.

Zusammenfassend konnten wir mögliche regulatorische Netzwerke, Gene und biologische Prozesse, welche durch die Bindung von TFAP2A beeinflusst werden, identifizieren. Der Versuchsaufbau kann auf weitere Phänotypen, Transkriptionsfaktoren und Zellmodelle übertragen werden.

Limitationen unserer Studie waren, dass die HEPM Zellen nur ein zweidimensionales Modell darstellen, sodass einerseits räumliche Effekte und andererseits Interaktionen mit anderen Zelltypen nicht abgebildet sind. Insbesondere die Interaktion von epithelialen und mesenchymalen Zellen ist jedoch für den Ablauf der Fusion der Gaumenfortsätze relevant (Nakajima et. al., 2018). Organoid-Modelle könnten hier einen nächsten Schritt darstellen (Wolf et. al., 2018). Außerdem konnten allelspezifische Bindungseffekte von TFAP2A nur an SNPs mit heterozygotem Genotyp untersucht werden. Ansätze wie Genome-editing mit CRISPR-Cas9 könnten hier eingesetzt oder Zellkulturen anderer Donoren untersucht werden (Doudna und Charpentier, 2012).

1.5 Zusammenfassung

Die Identifizierung genetischer Risikovarianten für multifaktorielle Erkrankungen ist in den letzten Jahren rasant vorangeschritten. Dies gilt auch für nicht-syndromale Lippenpalten mit oder ohne Gaumenspalten (nsCL/P), eine der häufigsten Fehlbildungen des Menschen. Die Translation der genetischen Befunde in biologische Erkenntnisse stellt jedoch eine Herausforderung dar, da viele mit der Krankheit assoziierte genetische Varianten in nicht-kodierenden Regionen des Genoms liegen. Ein möglicher funktioneller Effekt könnte eine veränderte Bindung von Transkriptionsfaktoren an ihre Bindestellen sein. In der vorliegenden Studie wurden Transkriptionsfaktorbindestellen von Transcription factor AP-2 alpha (TFAP2A), einem nsCL/P Kandidatengen, in Zellen des humanen Gaumenmesenchyms (HEPM-Zellen) untersucht. Im Rahmen einer Chromatinimmunpräzipitation (ChIP) und anschließender Sequenzierung identifizierten wir 2.845 Regionen mit TFAP2A Bindung (TFAP2A ChIP-seq Peaks). Im Abgleich mit Chromatinmarkern aus humanem embryonalem kraniofazialen Gewebe des Carnegie Stadiums 15 und TFAP2A ChIP-seq Peaks aus humanen Zervixkarzinomzellen konnten 802 potentiell für kraniofaziales Gewebe spezifische TFAP2A ChIP-seq Peaks mit möglicher Enhanceraktivität identifiziert werden. Basierend auf einer Gen-Ontologie-Analyse zeigte sich eine mögliche Funktion von TFAP2A in relevanten Prozessen wie der epithelial-mesenchymalen Transition und in regulatorischen Netzwerken mit anderen Genen der kraniofazialen Entwicklung, wie *MSX1* und *SPRY2*. Wir konnten 18 Regionen mit allelspezifischer Bindung von TFAP2A erfassen. Eine Integration mit einer Metaanalyse von genomweiten Assoziationsstudien (GWAS) zu nsCL/P von Ludwig et al. (2017) zeigte jedoch keine Überlappung der TFAP2A ChIP-seq Peaks mit genomweit signifikanten nsCL/P Risikoloci. Limitationen der Studie waren, dass allelspezifische Effekte nur an in unserem Zellmodell heterozygoten SNPs untersucht werden und dass dieses nur begrenzt die biologischen Gegebenheiten im embryonalen Gaumen simulieren konnten. In zukünftigen Studien können diese Einschränkungen mittels genomischer Editierung und komplexerer *in-vitro* und *in-vivo* Modelle überwunden werden. Unser Versuchsansatz kann auf andere Phänotypen und Transkriptionsfaktoren übertragen werden.

1.6 Literaturverzeichnis der deutschen Zusammenfassung

Al-Namankany A, Alhubaishi A. Effects of cleft lip and palate on children's psychological health: A systematic review. *J Taibah Univ Med Sci* 2018; 13: 311–318

Bailey SD, Virtanen C, Haibe-Kains B, Lupien M. ABC: a tool to identify SNVs causing allele-specific transcription factor binding from ChIP-Seq experiments. *Bioinformatics* 2015; 31: 3057–3059

Beaty TH, Murray JC, Marazita ML, Munger RG, Ruczinski I, Hetmanski JB, Liang KY, Wu T, Murray T, Fallin MD, Redett RA, Raymond G, Schwender H, Jin S-C, Cooper ME, Dunnwald M, Mansilla MA, Leslie E, Bullard S, Lidral AC, Moreno LM, Menezes R, Vieira AR, Petrin A, Wilcox AJ, Lie RT, Jabs EW, Wu-Chou YH, Chen PK, Wang H, Ye X, Huang S, Yeow V, Chong SS, Jee SH, Shi B, Christensen K, Melbye M, Doheny KF, Pugh EW, Ling H, Castilla EE, Czeizel AE, Ma L, Field LL, Brody L, Pangilinan F, Mills JL, Molloy AM, Kirke PN, Scott JM, Arcos-Burgos M, Scott AF. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat Genet* 2010; 42: 525–529

Beaty TH, Taub MA, Scott AF, Murray JC, Marazita ML, Schwender H, Parker MM, Hetmanski JB, Balakrishnan P, Mansilla MA, Mangold E, Ludwig KU, Noethen MM, Rubini M, Elcioglu N, Ruczinski I. Confirming genes influencing risk to cleft lip with/without cleft palate in a case-parent trio study. *Hum Genet* 2013; 132: 771–781

Birnbaum S, Ludwig KU, Reutter H, Herms S, Steffens M, Rubini M, Baluardo C, Ferrian M, Almeida de Assis N, Alblas MA, Barth S, Freudenberg J, Lauster C, Schmidt G, Scheer M, Braumann B, Bergé SJ, Reich RH, Schiefke F, Hemprich A, Pötzsch S, Steegers-Theunissen RP, Pötzsch B, Moebus S, Horsthemke B, Kramer F-J, Wienker TF, Mossey PA, Propping P, Cichon S, Hoffmann P, Knapp M, Nöthen MM, Mangold E. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet* 2009; 41: 473 EP -

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P, Boyle

PJ, Cao H, Carter NP, Clelland GK, Davis S, Day N, Dhimi P, Dillon SC, Dorschner MO, Fiegler H, Giresi PG, Goldy J, Hawrylycz M, Haydock A, Humbert R, James KD, Johnson BE, Johnson EM, Frum TT, Rosenzweig ER, Karnani N, Lee K, Lefebvre GC, Navas PA, Neri F, Parker SCJ, Sabo PJ, Sandstrom R, Shafer A, Vetrie D, Weaver M, Wilcox S, Yu M, Collins FS, Dekker J, Lieb JD, Tullius TD, Crawford GE, Sunyaev S, Noble WS, Dunham I, Denoeud F, Reymond A, Kapranov P, Rozowsky J, Zheng D, Castelo R, Frankish A, Harrow J, Ghosh S, Sandelin A, Hofacker IL, Baertsch R, Keefe D, Dike S, Cheng J, Hirsch HA, Sekinger EA, Lagarde J, Abril JF, Shahab A, Flamm C, Fried C, Hackermüller J, Hertel J, Lindemeyer M, Missal K, Tanzer A, Washietl S, Korbel J, Emanuelsson O, Pedersen JS, Holroyd N, Taylor R, Swarbreck D, Matthews N, Dickson MC, Thomas DJ, Weirauch MT, Gilbert J, Drenkow J, Bell I, Zhao X, Srinivasan KG, Sung W-K, Ooi HS, Chiu KP, Foissac S, Alioto T, Brent M, Pachter L, Tress ML, Valencia A, Choo SW, Choo CY, Ucla C, Manzano C, Wyss C, Cheung E, Clark TG, Brown JB, Ganesh M, Patel S, Tammana H, Chrast J, Henrichsen CN, Kai C, Kawai J, Nagalakshmi U, Wu J, Lian Z, Lian J, Newburger P, Zhang X, Bickel P, Mattick JS, Carninci P, Hayashizaki Y, Weissman S, Hubbard T, Myers RM, Rogers J, Stadler PF, Lowe TM, Wei C-L, Ruan Y, Struhl K, Gerstein M, Antonarakis SE, Fu Y, Green ED, Karaöz U, Siepel A, Taylor J, Liefer LA, Wetterstrand KA, Good PJ, Feingold EA, Guyer MS, Cooper GM, Asimenos G, Dewey CN, Hou M, Nikolaev S, Montoya-Burgos JI, Löytynoja A, Whelan S, Pardi F, Massingham T, Huang H, Zhang NR, Holmes I, Mullikin JC, Ureta-Vidal A, Paten B, Seringhaus M, Church D, Rosenbloom K, Kent WJ, Stone EA, Batzoglou S, Goldman N, Hardison RC, Haussler D, Miller W, Sidow A, Trinklein ND, Zhang ZD, Barrera L, Stuart R, King DC, Ameur A, Enroth S, Bieda MC, Kim J, Bhinge AA, Jiang N, Liu J, Yao F, Vega VB, Lee CWH, Ng P, Yang A, Moqtaderi Z, Zhu Z, Xu X, Squazzo S, Oberley MJ, Inman D, Singer MA, Richmond TA, Munn KJ, Rada-Iglesias A, Wallerman O, Komorowski J, Fowler JC, Couttet P, Bruce AW, Dovey OM, Ellis PD, Langford CF, Nix DA, Euskirchen G, Hartman S, Urban AE, Kraus P, van Calcar S, Heintzman N, Kim TH, Wang K, Qu C, Hon G, Luna R, Glass CK, Rosenfeld MG, Aldred SF, Cooper SJ, Halees A, Lin JM, Shulha HP, Zhang X, Xu M, Haidar JNS, Yu Y, Iyer VR, Green RD, Wadelius C, Farnham PJ, Ren B, Harte RA, Hinrichs AS, Trumbower H, Clawson H, Hillman-Jackson J, Zweig AS, Smith K, Thakkapallayil A, Barber G, Kuhn RM, Karolchik D, Armengol L, Bird CP, Bakker PIW de, Kern AD,

Lopez-Bigas N, Martin JD, Stranger BE, Woodroffe A, Davydov E, Dimas A, Eyras E, Hallgrímsson IB, Huppert J, Zody MC, Abecasis GR, Estivill X, Bouffard GG, Guan X, Hansen NF, Idol JR, Maduro VVB, Maskeri B, McDowell JC, Park M, Thomas PJ, Young AC, Blakesley RW, Muzny DM, Sodergren E, Wheeler DA, Worley KC, Jiang H, Weinstock GM, Gibbs RA, Graves T, Fulton R, Mardis ER, Wilson RK, Clamp M, Cuff J, Gnerre S, Jaffe DB, Chang JL, Lindblad-Toh K, Lander ES, Koriabine M, Nefedov M, Osoegawa K, Yoshinaga Y, Zhu B, Jong PJ de. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007; 447: 799–816

Bushnell B., 2019: BBMap BBMap short read aligner, and other bioinformatic tools. <https://sourceforge.net/projects/bbmap/> (Zugriffsdatum: 28.03.2019)

Carroll SH, Macias Trevino C, Li EB, Kawasaki K, Myers N, Hallett SA, Alhazmi N, Cotney J, Carstens RP, Liao EC. An *Irf6-Esrp1/2* regulatory axis controls midface morphogenesis in vertebrates. *Development* 2020; 147

Carroll TS, Liang Z, Salama R, Stark R, Santiago I de. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front Genet* 2014; 5: 75

Degtyareva AO, Antontseva EV, Merkulova TI. Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases. *Int J Mol Sci* 2021; 22

Dhulipala VC, Welshons WV, Reddy CS. Inhibition of human embryonic palatal mesenchymal cell cycle by secalonic acid D: a probable mechanism of its cleft palate induction. *Orthod Craniofac Res* 2004; 7: 227–236

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; 485: 376–380

Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* 2011; 12: 167–178

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15–21

Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014; 346: 1258096

Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016; 32: 3047–3048

Fairley S, Lowy-Gallego E, Perry E, Flicek P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res* 2020; 48: D941-D947

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2020; 48: D87-D92

Genome Reference Consortium, 2009: GRCh37.
https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/ (Zugriffsdatum: 10.07.2022)

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, Min R, Alves P, Abyzov A, Addleman N, Bhardwaj N, Boyle AP, Cayting P, Charos A, Chen DZ, Cheng Y, Clarke D, Eastman C, Euskirchen G, Fietze S, Fu Y, Gertz J, Grubert F, Harmanci A, Jain P, Kasowski M, Lacroute P, Leng JJ, Lian J, Monahan H, O'Geen H, Ouyang Z, Partridge EC, Patacsil D, Pauli F, Raha D, Ramirez L, Reddy TE, Reed B, Shi M, Slifer T, Wang J, Wu L, Yang X, Yip KY, Zilberman-Schapira G, Batzoglou S, Sidow A, Farnham PJ, Myers RM, Weissman SM, Snyder M. Architecture of the human regulatory network derived from ENCODE data. *Nature* 2012; 489: 91–100

Goodnough LH, Brugmann SA, Hu D, Helms JA. Stage-dependent craniofacial defects resulting from *Sprouty2* overexpression. *Dev Dyn* 2007; 236: 1918–1928

Grant SFA, Wang K, Zhang H, Glaberson W, Annaiah K, Kim CE, Bradfield JP, Glessner JT, Thomas KA, Garris M, Frackelton EC, Otieno FG, Chiavacci RM, Nah H-D, Kirschner RE, Hakonarson H. A Genome-Wide Association Study Identifies a Locus for Nonsyndromic Cleft Lip with or without Cleft Palate on 8q24. *J Pediatr* 2009; 155: 909–913

Green RM, Feng W, Phang T, Fish JL, Li H, Spritz RA, Marcucio RS, Hooper J, Jamniczky H, Hallgrímsson B, Williams T. *Tfap2a*-dependent changes in mouse facial morphology result in clefting that can be ameliorated by a reduction in *Fgf8* gene dosage. *Dis Model Mech* 2015; 8: 31–43

Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, Zhang R, Hartmann BM, Zaslavsky E, Sealfon SC, Chasman DI, FitzGerald GA, Dolinski K, Grosser T, Troyanskaya OG. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet* 2015; 47: 569–576

Grosen D, Chevrier C, Skytthe A, Bille C, Mølsted K, Sivertsen A, Murray JC, Christensen K. A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *J Med Genet* 2010; 47: 162–168

Grosen D, Bille C, Petersen I, Skytthe A, Hjelmberg JvB, Pedersen JK, Murray JC, Christensen K. Risk of oral clefts in twins. *Epidemiology (Cambridge, Mass.)* 2011; 22: 313–319

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. *Genome Biol* 2007; 8: R24

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; 5: e1000529

Jackson A, Bromley R, Morrow J, Irwin B, Clayton-Smith J. In utero exposure to valproate increases the risk of isolated cleft palate. *Arch Dis Child Fetal Neonatal Ed* 2016; 101: F207

Jahanbin A, Shadkam E, Miri HH, Shirazi AS, Abtahi M. Maternal Folic Acid Supplementation and the Risk of Oral Clefts in Offspring. *J Craniofac Surg* 2018; 29

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, Sandelin A, Vandepoele K, Lenhard B, Ballester B, Wasserman WW, Parcy F, Mathelier A. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 2018; 46: D260-D266

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012; 9: 357–359

Leslie EJ, Carlson JC, Shaffer JR, Butali A, Buxó CJ, Castilla EE, Christensen K, Deleyiannis FWB, Leigh Field L, Hecht JT, Moreno L, Orioli IM, Padilla C, Vieira AR, Wehby GL, Feingold E, Weinberg SM, Murray JC, Beaty TH, Marazita ML. Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum Genet* 2017; 136: 275–286

Leslie EJ, Carlson JC, Shaffer JR, Feingold E, Wehby G, Laurie CA, Jain D, Laurie CC, Doheny KF, McHenry T, Resick J, Sanchez C, Jacobs J, Emanuele B, Vieira AR, Neiswanger K, Lidral AC, Valencia-Ramirez LC, Lopez-Palacio AM, Valencia DR, Arcos-Burgos M, Czeizel AE, Field LL, Padilla CD, Cutiongco-de la Paz, Eva Maria C, Deleyiannis F, Christensen K, Munger RG, Lie RT, Wilcox A, Romitti PA, Castilla EE, Mereb JC, Poletta FA, Orioli IM, Carvalho FM, Hecht JT, Blanton SH, Buxó CJ, Butali A, Mossey PA, Adeyemo WL, James O, Braimah RO, Aregbesola BS, Eshete MA, Abate F, Koruyucu M, Seymen F, Ma L, Salamanca JE de, Weinberg SM, Moreno L, Murray JC, Marazita ML. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Hum Mol Genet* 2016; 25: 2862–2872

Leslie EJ, Marazita ML. Genetics of Orofacial Cleft Birth Defects. *Current Genetic Medicine Reports* 2015; 3: 118–126

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; 30: 923–930

Little J, Cardy A, Munger RG. Tobacco smoking and oral clefts: a meta-analysis. *Bull World Health Organ* 2004; 82: 213–218

Liu H, Leslie EJ, Carlson JC, Beaty TH, Marazita ML, Lidral AC, Cornell RA. Identification of common non-coding variants at 1p22 that are functional for non-syndromic orofacial clefting. *Nat Commun* 2017; 8: 14759

Lorot-Marchand A, Guerreschi P, Pellerin P, Martinot V, Gbaguidi CC, Neiva C, Devauchelle B, Frochisse C, Poli-Merol M-L, François-Fiquet, Fiquet, Fiquet-François C. Frequency and socio-psychological impact of taunting in school-age patients with cleft lip-palate surgical repair. *Int J Pediatr Otorhinolaryngol* 2015; 79

Ludwig KU, Ahmed ST, Böhmer AC, Sangani NB, Varghese S, Klamt J, Schuenke H, Gültepe P, Hofmann A, Rubini M, Aldhorae KA, Steegers-Theunissen RP, Rojas-Martinez A, Reiter R, Borck G, Knapp M, Nakatomi M, Graf D, Mangold E, Peters H. Meta-analysis Reveals Genome-Wide Significance at 15q13 for Nonsyndromic Clefting of Both the Lip and the Palate, and Functional Analyses Implicate *GREM1* As a Plausible Causative Gene. *PLoS Genet* 2016; 12: e1005914-e1005914

Ludwig KU, Böhmer AC, Bowes J, Nikolic M, Ishorst N, Wyatt N, Hammond NL, Gözl L, Thieme F, Barth S, Schuenke H, Klamt J, Spielmann M, Aldhorae K, Rojas-Martinez A, Nöthen MM, Rada-Iglesias A, Dixon MJ, Knapp M, Mangold E. Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip ± cleft palate and cleft palate only. *Hum Mol Genet* 2017; 26: 829–842

Ludwig KU, Mangold E, Herms S, Nowak S, Reutter H, Paul A, Becker J, Herberz R, AlChawa T, Nasser E, Böhmer AC, Mattheisen M, Alblas MA, Barth S, Kluck N, Lauster C, Braumann B, Reich RH, Hemprich A, Pötzsch S, Blaumeiser B, Daratsianos N, Kreusch T, Murray JC, Marazita ML, Ruczinski I, Scott AF, Beaty TH, Kramer F-J,

Wienker TF, Steegers-Theunissen RP, Rubini M, Mossey PA, Hoffmann P, Lange C, Cichon S, Propping P, Knapp M, Nöthen MM. Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat Genet* 2012; 44: 968–971

Machanick P, Bailey TL. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 2011; 27: 1696–1697

Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015; 31: 3555–3557

Mack KL, Nachman MW. Gene Regulation and Speciation. *Trends Genet : TIG* 2017; 33: 68–80

Mangold E, Ludwig KU, Birnbaum S, Baluardo C, Ferrian M, Herms S, Reutter H, Assis NA de, Chawa TA, Mattheisen M, Steffens M, Barth S, Kluck N, Paul A, Becker J, Lauster C, Schmidt G, Braumann B, Scheer M, Reich RH, Hemprich A, Pötzsch S, Blaumeiser B, Moebus S, Krawczak M, Schreiber S, Meitinger T, Wichmann H-E, Steegers-Theunissen RP, Kramer F-J, Cichon S, Propping P, Wienker TF, Knapp M, Rubini M, Mossey PA, Hoffmann P, Nöthen MM. Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat Genet* 2010; 42: 24–26

Marcusson A, Paulin G, Östrup L. Facial appearance in adults who had cleft lip and palate treated in childhood. *Scand J Plast Reconstr Surg Hand Surg* 2002; 36: 16–23

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011, 17(1): 10-12

Matsumura K, Taketomi T, Yoshizaki K, Arai S, Sanui T, Yoshiga D, Yoshimura A, Nakamura S. Sprouty2 controls proliferation of palate mesenchymal cells via fibroblast growth factor signaling. *Biochem Biophys Res Commun* 2011; 404: 1076–1082

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010; 28: 495–501

Milunsky JM, Maher TA, Zhao G, Roberts AE, Stalker HJ, Zori RT, Burch MN, Clemens M, Mulliken JB, Smith R, Lin AE. TFAP2A mutations result in branchio-oculo-facial syndrome. *Am J Hum Genet* 2008; 82: 1171–1177

Moreno LM, Mansilla MA, Bullard SA, Cooper ME, Busch TD, Machida J, Johnson MK, Brauer D, Krahn K, Daack-Hirsch S, L'heureux J, Valencia-Ramirez C, Rivera D, López AM, Moreno MA, Hing A, Lammer EJ, Jones M, Christensen K, Lie RT, Jugessur A, Wilcox AJ, Chines P, Pugh E, Doheny K, Arcos-Burgos M, Marazita ML, Murray JC, Lidral AC. FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum Mol Genet* 2009; 18: 4879–4896

Mossey PA, Little J, Munger RG, Dixon MJ, Shaw WC. Cleft lip and palate. *Lancet* 2009; 374: 1773–1785

Mullens CL, McCulloch IL, Hardy KM, Mathews RE, Mason AC. *Plast Reconstr Surg Glob Open*. *Plast Reconstr Surg Glob Open* 2019; 7: e2095-e2095

Nakajima A, F Shuler C, Gulka AOD, Hanai J-I. TGF- β Signaling and the Epithelial-Mesenchymal Transition during Palatal Fusion. *Int J Mol Sci* 2018; 19

Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* 2013; 368: 20120362

O'Rahilly R, Müller F. Developmental stages in human embryos: revised and new measurements. *Cells Tissues Organs* 2010; 192: 73–84

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; 26: 841–842

Rahimov F, Marazita ML, Visel A, Cooper ME, Hitchler MJ, Rubini M, Domann FE, Govil M, Christensen K, Bille C, Melbye M, Jugessur A, Lie RT, Wilcox AJ, Fitzpatrick DR, Green ED, NISC Comparative Sequencing Program, Mossey PA, Little J, Steegers-

Theunissen RP, Pennacchio LA, Schutte BC, Murray JC. Disruption of an AP-2 α binding site in an IRF6 enhancer is associated with cleft lip. *Nat Genet* 2008; 40: 1341 EP -

Razaghi-Moghadam Z, Namipashaki A, Farahmand S, Ansari-Pour N. Systems genetics of nonsyndromic orofacial clefting provides insights into its complex aetiology. *Eur J Hum Genet* 2019; 27: 226–234

Ruff KLM, Hollstein R, Fazaal J, Thieme F, Gehlen J, Mangold E, Knapp M, Welzenbach J, Ludwig KU. Allele-specific transcription factor binding in a cellular model of orofacial clefting. *Scientific Reports* 2022; 12: 1807

Schorle H, Meier P, Buchert M, Jaenisch R, Mitchell PJ. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* 1996; 381: 235–238

Shi M, Christensen K, Weinberg CR, Romitti P, Bathum L, Lozada A, Morris RW, Lovett M, Murray JC. Orofacial cleft risk is increased with maternal smoking and specific detoxification-gene variants. *Am J Hum Genet* 2007; 80: 76–90

Stott-Miller M, Heike CL, Kratz M, Starr JR. Increased risk of orofacial clefts associated with maternal obesity: case-control study and Monte Carlo-based bias analysis. *Paediatr Perinat Epidemiol* 2010; 24: 502–512

Sun Y, Huang Y, Yin A, Pan Y, Wang Y, Wang C, Du Y, Wang M, Lan F, Hu Z, Wang G, Jiang M, Ma J, Zhang X, Ma H, Ma J, Zhang W, Huang Q, Zhou Z, Ma L, Li Y, Jiang H, Xie L, Jiang Y, Shi B, Cheng J, Shen H, Wang L, Yang Y. Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat Commun* 2015; 6: 6414 EP -

Thieme F, Ludwig KU. The Role of Noncoding Genetic Variation in Isolated Orofacial Clefts *J Dent Res* 2017; 96: 1238–1247

Wahl SE, Kennedy AE, Wyatt BH, Moore AD, Pridgen DE, Cherry AM, Mavila CB, Dickinson AJG. The role of folate metabolism in orofacial development and clefting. *Dev Biol* 2015; 405: 108–122

Wang J, Bai Y, Li H, Greene SB, Klysiak E, Yu W, Schwartz RJ, Williams TJ, Martin JF. MicroRNA-17-92, a direct Ap-2 α transcriptional target, modulates T-box factor activity in orofacial clefting. *PLoS Genet* 2013; 9: e1003785-e1003785

Welzenbach J, Hammond NL, Nikolić M, Thieme F, Ishorst N, Leslie EJ, Weinberg SM, Beaty TH, Marazita ML, Mangold E, Knapp M, Cotney J, Rada-Iglesias A, Dixon MJ, Ludwig KU. Integrative approaches generate insights into the architecture of non-syndromic cleft lip with or without cleft palate. *HGG Adv* 2021; 2: 100038

Wilderman A, VanOudenhove J, Kron J, Noonan JP, Cotney J. High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* 2018; 23: 1581–1597

Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* 2018; 7: 1338

Wolf CJ, Belair DG, Becker CM, Das KP, Schmid JE, Abbott BD. Development of an organotypic stem cell model for the study of human embryonic palatal fusion. *Birth Defects Res* 2018; 110: 1322–1334

Yoneda T, Pratt RM. Mesenchymal cells from the human embryonic palate are highly responsive to epidermal growth factor. *Science* 1981; 213: 563

Yu Y, Zuo X, He M, Gao J, Fu Y, Qin C, Meng L, Wang W, Song Y, Cheng Y, Zhou F, Chen G, Zheng X, Wang X, Liang B, Zhu Z, Fu X, Sheng Y, Hao J, Liu Z, Yan H, Mangold E, Ruczinski I, Liu J, Marazita ML, Ludwig KU, Beaty TH, Zhang X, Sun L, Bian Z. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat Commun* 2017; 8: 14364

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008; 9: R137

Zhang Z, Song Y, Zhao X, Zhang X, Fermin C, Chen Y. Rescue of cleft palate in *Msx1*-deficient mice by transgenic *Bmp4* reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. *Development* 2002; 129: 4135–4146



OPEN

Allele-specific transcription factor binding in a cellular model of orofacial clefting

Katharina L. M. Ruff¹, Ronja Hollstein¹, Julia Fazaal¹, Frederic Thieme¹, Jan Gehlen², Elisabeth Mangold¹, Michael Knapp³, Julia Welzenbach¹ & Kerstin U. Ludwig¹✉

Non-syndromic cleft lip with/without cleft palate (nsCL/P) is a frequent congenital malformation with multifactorial etiology. While recent genome-wide association studies (GWAS) have identified several nsCL/P risk loci, the functional effects of the associated non-coding variants are largely unknown. Furthermore, additional risk loci remain undetected due to lack of power. As genetic variants might alter binding of transcription factors (TF), we here hypothesized that the integration of data from TF binding sites, expression analyses and nsCL/P GWAS might help to (i) identify functionally relevant variants at GWAS loci, and (ii) highlight novel risk variants that have been previously undetected. Analysing the craniofacial TF TFAP2A in human embryonic palatal mesenchyme (HEPM) cells, we identified 2845 TFAP2A ChIP-seq peaks, several of which were located near nsCL/P candidate genes (e.g. *MSX1* and *SPRY2*). Comparison with independent data suggest that 802 of them might be specific to craniofacial development, and genes near these peaks are enriched in processes relevant to nsCL/P. Integration with nsCL/P GWAS data, however, did not show robust evidence for co-localization of common nsCL/P risk variants with TFAP2A ChIP-seq peaks. This data set represents a new resource for the analyses of craniofacial processes, and similar approaches with additional cell lines and TFs could be applied to generate further insights into nsCL/P etiology.

Cleft lip with or without cleft palate (CL/P) is a frequent form of human orofacial clefting, and ranks among the most common of all congenital malformations¹. In approximately one third of patients, CL/P occurs within the context of a known genetic syndrome^{2,3}. However, in the majority of cases, CL/P arises as an isolated malformation, and is referred to as non-syndromic CL/P (nsCL/P)³. nsCL/P has a multifactorial etiology, which is characterized by the contribution of both genetic and environmental risk factors⁴. Heritability estimates from twin studies and multiplex pedigrees are high, with reported values of up to 90%^{5,6}. This suggests that genetic factors make a substantial contribution to nsCL/P etiology.

Over the past decade, multiple investigations have been performed to identify the causal variants, risk genes, and functional mechanisms that contribute to nsCL/P⁷. These have included several genome-wide association studies (GWAS) and meta-analyses in diverse populations. Together, these genetic studies have identified 45 nsCL/P risk loci, which explain around 30% of the heritability^{8–21}. Still, identifying causal variants at these loci remains challenging, since most of the associated single nucleotide polymorphisms (SNPs) are (i) located in non-coding regions, and (ii) their biological effect is difficult to dissect due to the presence of linkage disequilibrium²².

One of the mechanisms through which risk variants in non-coding regions can contribute to disease phenotypes is altered transcription factor (TF) binding to *cis*-regulatory elements²³. Differential TF binding can modify the expression pattern of direct target genes, and also trigger downstream effects at the gene network level. Since the majority of gene regulation networks are highly cell-type and cell-state-specific, identifying effects of this nature requires analyses of cellular systems that are relevant to the specific disease in question²⁴.

The major goal of the present study was to develop a framework for investigation of transcription factor binding events in nsCL/P, through integration of molecular data from human embryonic palatal mesenchyme (HEPM) cells^{25–27} and nsCL/P GWAS data. We first identified candidate TFs in HEPM through expression profiling. Among those, we prioritized TFAP2A for further analyses, for several reasons. First, deleterious mutations in *TFAP2A* cause Branchio-Oculo-Facial Syndrome²⁸. This syndrome is characterized by dysmorphic anomalies and characteristic facial phenotypes including cleft palate²⁸. Second, TFAP2A has been shown to bind to the nsCL/P risk variant rs642961, located within an enhancer of the nsCL/P candidate gene *IRF6*²⁰. Furthermore, the

¹School of Medicine and University Hospital Bonn, Institute of Human Genetics, University of Bonn, Bonn, Germany. ²Centre for Human Genetics, University of Marburg, Marburg, Germany. ³Institute for Medical Biometry, Informatics and Epidemiology IMBIE, University of Bonn, Bonn, Germany. ✉email: kerstin.ludwig@uni-bonn.de

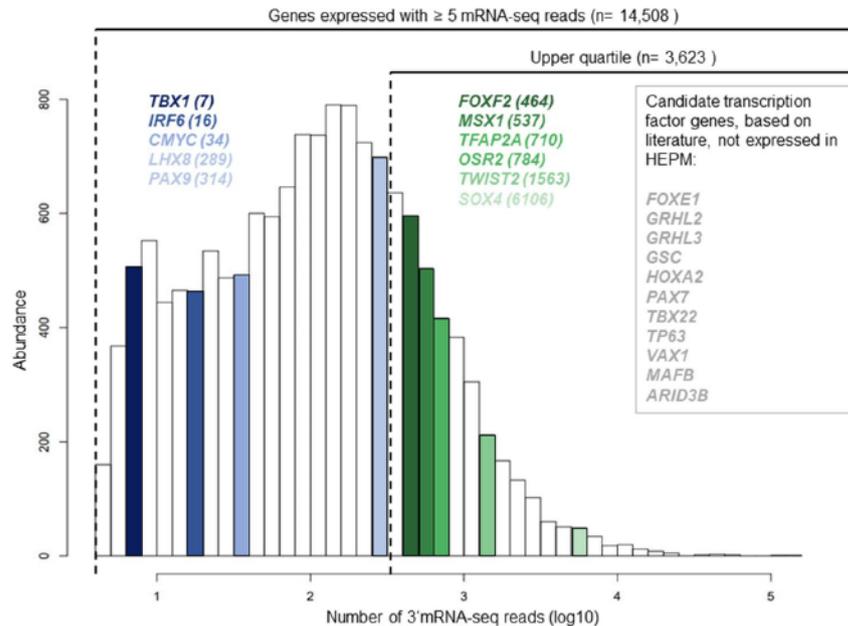


Figure 1. Expression analysis in human embryonal palatal mesenchyme (HEPM) cells. Density plot of 3'mRNA sequence reads, provided at log10 scale for the average of three replicates. In total, 14,508 genes were observed with ≥ 5 reads, and 3,623 genes were in the upper quartile (≥ 327 reads) of all genes with ≥ 5 reads. Out of 22 candidate transcription factors for craniofacial development, 11 were expressed in HEPM.

TFAP2A-IRF6 pathway is a well-established pathway involved in orofacial clefting²⁹. Finally, *TFAP2A* is located at an nsCL/P risk locus (6p24) previously identified by GWAS⁸.

Performing chromatin immunoprecipitation followed by sequencing (ChIP-seq) in HEPM, we then identified ChIP-seq peaks indicative of TFAP2A binding and confirmed plausibility of these regions through comparisons with external data sets. This map of binding regions was next integrated with (i) genotype information from HEPM cells, and (ii) summary statistics of a previous nsCL/P GWAS meta-analysis⁹ with two aims. First, we wanted to identify potential causal variants at known GWAS risk loci. Therefore, we looked whether TFAP2A ChIP-seq peaks co-localize with associated risk variants, and whether they exhibit allele-specific TFAP2A binding effects at heterozygous positions. Second, we hypothesized that binding of TFAP2A in HEPM might contribute to nsCL/P at loci that are not yet genome-wide significant, due to limited power of current GWAS studies. While the results of our analyses provide only limited evidence for a role of genetically mediated effect of nsCL/P risk variants at TFAP2A ChIP-seq peaks in HEPM, the map of binding sites as well as the framework described here can be used as blueprint for further integrative analyses of epigenetic and genetic data in nsCL/P.

Results

TFAP2A is a candidate TF expressed in HEPM cells. RNA-seq in HEPM revealed 14,508 expressed genes, defined by an expression with ≥ 5 aligned m-RNA-seq reads in average. This included 350 TFs that were represented by motifs in the JASPAR core vertebrate assembly 2020³⁰ (Supplementary Table S1, Fig. 1). In addition, literature research revealed 22 TFs that have been previously reported with a role in craniofacial development and/or orofacial clefting (Supplementary Table S2). Integration both data sets revealed an overlap of 11 TFs, with 6 TFs being considered "strongly expressed" as defined by the largest quartile (Supplementary Table S2, Fig. 1). For 2 out of these 11 TFs (*TFAP2A* and *MSX1*), additional support for an involvement in nsCL/P etiology was available through their location at previously identified nsCL/P risk loci (i.e., 6p24 and 4p16⁸). In the present study we prioritized *TFAP2A*, as the GWAS signal at this locus was mainly driven by the European population, thus matching the ethnicity of the HEPM donor (as confirmed by principal component analysis of array-based genotypes, *data not shown*).

TFAP2A ChIP-seq peaks are located near nsCL/P candidate genes. Peak calling of the TFAP2A ChIP-seq reads in two replicates revealed 5,820 and 3,989 unique peaks, respectively (FDR < 5%, fold enrichment (FE) 5–50, Supplementary Tables S3, S4). The intersection encompassed 2845 regions, which were considered high confidence peaks (hc-peaks, average size of 324 bp, Supplementary Table S5, Fig. 2a). Retrieving the sequence from the hc-peak summit regions identified a highly enriched 15 bp motif (e-value = 2.9×10^{-333} ; present in 1535 summit regions Fig. 2b), which matched the three TFAP2A binding motifs of the JASPAR core 2018 assembly³¹ (Fig. 2c). We also compared the TFAP2A ChIP-seq peaks from HEPM cells to a set of TFAP2A

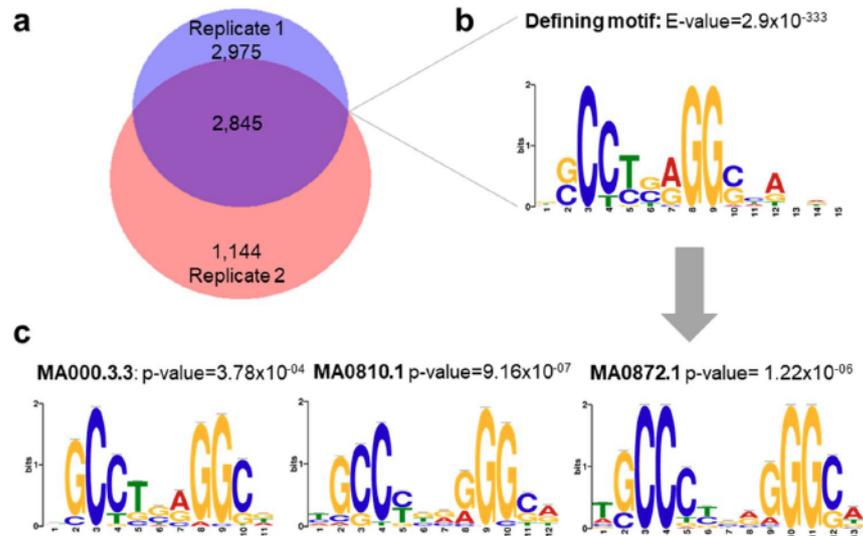


Figure 2. TFAP2A high confidence peaks (hc-peaks) in human embryonal palatal mesenchyme (HEPM) cells (a) Venn diagram showing the total number of TFAP2A ChIP-seq peaks across two replicates (replicate 1 $n=5820$; replicate 2 $n=3989$). The overlap of 2845 regions was denoted as „hc-peaks” for the subsequent analyses. Plotted with BioVenn[®] (2020). (b) Identification of the most abundant binding motif within hc-peaks (present in $n=1535$ hc-peaks) with MEME-ChIP. (c) Comparative analysis using Tomtom and JASPAR core 2018 assembly identified the significant alignment of the enriched motif with three distinct TFAP2A binding motifs.

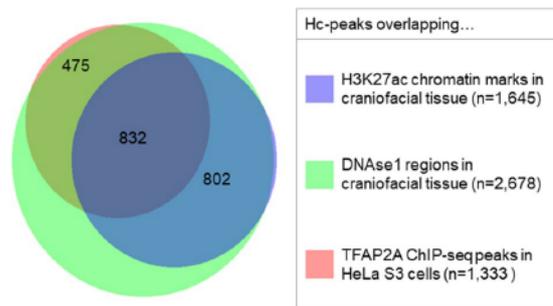


Figure 3. Intersection of TFAP2A hc-peaks overlapping with TFAP2A ChIP-seq peaks in HeLa S3 and/or chromatin marks in craniofacial tissue. Venn diagram displaying the intersection of 2,713 hc-peaks overlapping with TFAP2A ChIP-seq peaks in HeLa S3/ DNase1 hypersensitivity regions/ H3K27ac chromatin marks in CS15 craniofacial tissue. Hc-peaks that did not overlap with any of these are not displayed in the diagram ($n=132$). A total of 802 hc-peaks overlapped with DNase1 hypersensitivity regions and H3K27ac marks in CS15 craniofacial tissue, but did not overlap with a TFAP2A ChIP-seq peak in HeLa S3 cells, thus indicating craniofacial-specific hc-peaks. Overlaps were defined as described in the Methods. Plotted with BioVenn[®] (2020).

ChIP-seq peaks obtained from HeLa S3 cells³², which is a non-craniofacial cell line of human cervical cancer cells³³. The purpose of using data from HeLa S3 cells in our study was to use them as background to identify TFAP2A ChIP-seq peaks that might be specific to facial mesenchyme based on their absence in HeLa S3 cells. We observed a highly significant co-localization, with 1333 of the 2845 hc-peaks overlapping at a minimum of one site ($P\chi_2 < 0.0001$, Supplementary Table S5, Fig. 3), indicative of plausibility of the ChIP-seq peaks in HEPM. The assignment of genes adjacent to hc-peaks using GREAT yielded a total of 3470 genes which can be considered candidates for TFAP2A-mediated regulation (Supplementary Table S5). These genes included a set of 10 established nsCL/P candidate genes, such as *MSX1*⁸, *TP63*³², and *SPRY2*¹¹ (Table 1), eight of which were also expressed in HEPM cells (average of ≥ 5 reads, Table 1).

Hc-peak information			NsCL/P candidate gene ^a			
Chromosome	Start	End	Gene symbol	Distance peak to TSS (bp)	Reference (PubMed-ID)	Expression in HEPM cells (average no. of 3'RNA-seq reads) ^b
1	94,787,207	94,787,445	ARHGAP29	-84,205	Beaty et al. 2010 (PMID: 20,436,469)	1436
1	94,791,145	94,791,437		-88,170		
3	99,844,546	99,844,795	FILIP1L	-11,314	Beaty et al. 2013 (PMID: 23,512,105)	523
3	99,878,493	99,878,721		45,250		
3	189,281,805	189,282,039	TP63	-67,294	Leslie et al. 2017 (PMID: 28,054,174)	Not expressed
3	189,655,227	189,655,963		306,379		
4	4,860,996	4,861,363	MSX1	-213	Yu et al. 2017 (PMID: 28,232,668)	537
13	80,205,476	80,205,754	SPRY2	708,179	Ludwig et al. 2012 (PMID: 22,863,734)	125
13	80,604,734	80,605,110		308,872		
13	80,788,600	80,788,933		125,027		
13	80,915,560	80,915,801		-1887		
15	32,962,254	32,962,778	GREM1	-47,659	Ludwig et al. 2016 (PMID: 26,968,009)	5471
15	62,898,345	62,898,659	TPM1	-436,382	Ludwig et al. 2012 (PMID: 22,863,734)	7613
15	63,233,187	63,233,480		-101,550		
15	74,838,564	74,838,823	ARID3B	5176	Ludwig et al. 2017 (PMID: 28,087,736)	4
16	4,166,662	4,166,962	ADCY9	-626	Sun et al. 2015 (PMID: 25,775,280)	126
17	54,240,701	54,240,761	NOG	-430,329	Mangold et al. 2010 (PMID: 20,023,658); Leslie et al. 2015 (PMID: 25,704,602)	60
17	54,672,130	54,672,266		1138		

Table 1. Candidate genes for nsCL/P located near TFAP2A binding sites in HEPM cells. ^aNsCL/P candidate genes as putative downstream target genes of hc-peaks assigned with GREAT (association rules described in methods). ^bAverage number of 3'-mRNA-seq reads (Supplementary Table S1). Methods provided in the text.

Developmental processes are enriched at sites of TFAP2A ChIP-seq peaks in HEPM. We next investigated the relevance of the HEPM-based TFAP2A ChIP-seq peaks for craniofacial development. Using previously published DNase I hypersensitivity regions from embryonic craniofacial tissue of Carnegie Stage 15³⁴, we identified co-localization of 2678 out of 2845 hc-peaks ($P_{\chi^2} < 0.0001$, Supplementary Table S5, Fig. 3). Of these, 1,634 (61%) also overlapped with at least 1 signal for H3K27ac in CS15 craniofacial tissue, which is suggestive of enhancer activity of this region during human craniofacial development ($P_{\chi^2} < 0.0001$, Supplementary Table S5, Fig. 3). Interestingly, 802 of the 2845 hc-peaks overlapped with both DNase I and H3K27ac marks in CS15 craniofacial tissue, but did not overlap with the TFAP2A ChIP-seq peaks in HeLa S3 cells (Supplementary Table S5, Fig. 3). Results of a GO analysis for genes located at these 802 hc-peaks yielded significant results for 28 biological processes, 10 human phenotypes, and 15 mouse single knockout phenotypes with an FDR q -value < 0.05 (Supplementary Tables S6–S8, Fig. 4). These included processes such as “regulation of transforming growth factor beta receptor signaling pathway” or “regulation of cell–matrix adhesion”, but also “abnormality of facial soft tissue” and “abnormal palatine bone morphology”. When comparing these GO enrichment results with those obtained for all 2845 hc-peaks, we found that several terms were either more significantly enriched, or did only show an association in the analysis of craniofacial-specific peaks (Supplementary Tables S6–S11). Together, these findings suggest that TFAP2A-binding in human palatal mesenchymal cells might play a role in craniofacial processes.

Limited evidence for GWAS risk variants in TFAP2A peaks. Subsequent positional integration of the 2845 hc-peaks with the topological associating domains (TAD) comprising the 45 nsCL/P GWAS risk loci^{21,35} revealed the presence of 70 TFAP2A hc-peaks within 20 of these TADs. Among the 46 common SNPs located within these hc-peaks, we observed rs672819 at 1q32.1 to be in moderate LD with rs3753517 ($D' = 0.51$; $r^2 = 0.18$), which is the lead variant at this locus, and of rs447476 with rs2303914, the lead SNP at 2p25.1 ($D' = 0.91$; $r^2 = 0.28$)²¹ (Supplementary Table S12). At genome-wide scale, we observed a total of 1,564 common SNPs being located within hc-peaks (Supplementary Table S13). Data integration with a previously published nsCL/P GWAS meta-analysis⁹ revealed 29 SNPs which had $P_{\text{GWAS}} < 0.01$ (including four SNPs at $P_{\text{GWAS}} < 0.001$: rs463271 (22q11.21, $P_{\text{GWAS}} = 2.72 \times 10^{-4}$); rs12882215 and rs7144455 (14q24.3, $P_{\text{GWAS}} = 3.30 \times 10^{-4}$ and $P_{\text{GWAS}} = 3.77 \times 10^{-4}$, respectively, $D' = 1.0$; $r^2 = 0.85$); and rs4774822 (15q21.3; $P_{\text{GWAS}} = 6.27 \times 10^{-4}$, Supplementary Table S13). Together, our data provide only very limited support for the hypothesis that the association of common risk SNPs at nsCL/P GWAS loci can be attributed to TFAP2A binding in HEPM.

Identification of candidate variants for nsCL/P with allele-specific effects on binding. We finally investigated whether there is evidence for a genetically-mediated effect of TFAP2A-bound regions (and variants therein) outside of the known GWAS loci, for instance, at suggestive loci. For the analysis of allele-specific effects of TFAP2A binding, we first retrieved array-based genotypes for HEPM and found that 461/1,564 SNPs were heterozygous, including rs463271, rs672819 and rs447476 (see above). The other 3 SNPs at $P_{\text{GWAS}} < 0.001$ were found to be homozygous. At 18 positions we found statistical evidence for allele-specific

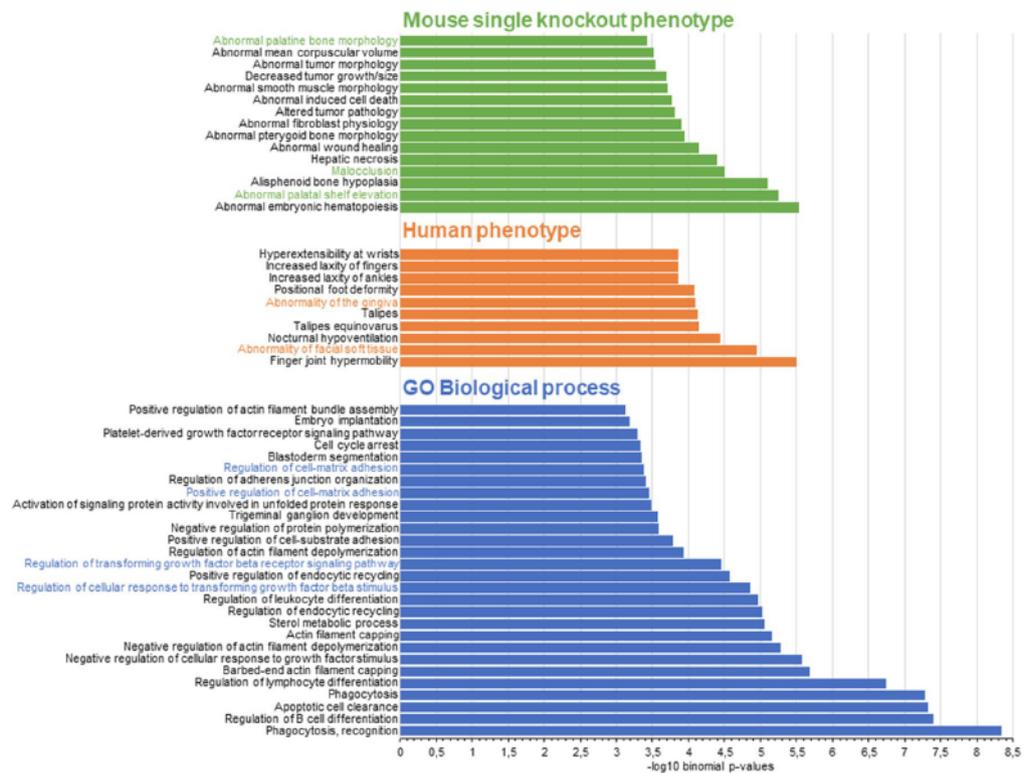


Figure 4. Gene Ontology (GO) analysis of 802 craniofacial-specific TFAP2A high confidence peaks (hc-peaks). Bar charts displaying $-\log_{10}$ binomial p -values of enriched mouse single knockout phenotypes with false discovery rate (FDR) < 0.05 , human phenotypes with FDR < 0.05 and GO biological processes with FDR < 0.05 . Terms relating to orofacial clefting are highlighted in color. Genes associated with these terms in assigned to TFAP2A hc-peaks, based on their proximity, as described in the Methods, are displayed in the boxes.

reads (binomial $p < 0.05$, with consistent directions of effect across both replicates, Supplementary Table S14), but this did neither include rs463271 nor any other SNP with statistically significant association results from the nsCL/P GWAS meta-analysis (lowest P -value observed: $P_{\text{GWAS}} = 0.091$ for rs616822 at 18q21.33). Analysis of potential downstream genes at these 18 sites revealed that 8 and 10 SNPs were associated with 1 or 2 possible target genes, respectively (Supplementary Table S14). However, none of these genes has been reported as candidate gene for nsCL/P, and also GO enrichment for these gene sets did not yield any significant process at an FDR q -value < 0.05 .

Discussion

A major challenge to our understanding of the biological role of nsCL/P-associated risk variants is their predominant mapping to non-coding regions of the genome. Although in vivo and in vitro approaches exist to identify functional variants among the statistically associated passenger variants, the major prerequisite is the analysis of disease-relevant tissue. At mechanistic level, altered TF binding to variant alleles has been suggested as one biological process through which non-coding risk variants might contribute to complex traits and diseases, including nsCL/P^{22,25}. Examples include a recent study by Huo et al.³⁶, who analyzed TFBS from ChIP-seq data of 30 TFs in brain tissues. The authors identified 132 schizophrenia risk variants that exert a functional effect on gene expression by disrupting the binding sites of 21 TFs. Also, Benaglio et al.³⁷ investigated the regulatory landscape and gene expression profiles of human induced pluripotent stem cell derived cardiomyocytes and identified differential binding of the cardiac TF NKX2-5 to risk SNPs at GWAS loci for electrocardiographic traits such as atrial fibrillation.

In the present study we examined binding of TFAP2A in human embryonal palatal mesenchyme cells which serve as in vitro model of human palate development, since they have been retrieved from the palatal shelves of an embryo at a gestational age when elevation and fusion processes take place²⁷. TFAP2A was chosen since it is (i) encoded by an nsCL/P candidate gene with robust evidence for an involvement in craniofacial development, and (ii) is strongly expressed in HEPM. Further evidence for a role of TFAP2A has been gained in mice, where knock-down of *Tfap2a* leads to severe malformations including facial clefting^{24,38}, and in a multi-omics study of human dental pulp stem cells³⁹. Importantly, while the orchestrational role of TFAP2A in structures derived from neural crest or ectodermal cells is well documented, few data are available concerning its role in the palatal

mesenchyme^{40,41}. For instance, it has been shown that in mice, *Tfap2a* is expressed in the facial mesenchyme during palate development⁴², and knock down of *Tfap2a* leads to an increase in the expression of *Fgf8*, which is implicated in the differentiation of mesenchymal cells into cartilage in the anterior plate^{43,44}. While this suggests a regulatory connection of *Tfap2a* and *Fgf8*, in our study, we did not observe any TFAP2A ChIP-seq peak in the proximity of *FGF8*. However, we observed TFAP2A binding sites near other nsCL/P candidate genes, such as *MSX1* on chr4p16⁸, and *SPRY2* on chr13q31¹¹. At the *MSX1* locus, the TFAP2A ChIP-seq peak is located ~200 bp upstream of the *MSX1* transcription start site (TSS). In concordance with prior evidence of *Msx1* being expressed in the palatal mesenchyme at various stages of development, we also observed strong expression of *MSX1* in HEPM. Additionally, *Msx1* has been suggested to be required for normal outgrowth of palatal shelves and mesenchymal proliferation^{45,46}, and research has shown that *Tfap2a* regulates *Msx1* expression in murine neural crest cells⁴⁷. At the *SPRY2*-locus, several TFAP2A ChIP-seq peaks have been identified, the closest of which is located ~1.9 kb upstream of the TSS of *SPRY2*¹¹. *Spry2* is expressed in mouse palatal mesenchyme, and both knock out and overexpression of the gene lead to cleft palate^{48,49}. *Spry2* knock out mice display an increased proliferation of palatal mesenchyme and a higher expression of other clefting genes, such as *Msx1*, *Etf5* and *Ptx1*⁴⁸. We also observed strong expression of *SPRY2* in HEPM. Together, these lines of evidence suggest that these three nsCL/P risk genes (*MSX1*, *SPRY2*, and *TFAP2A*) might form an nsCL/P regulatory network in HEPM cells. In addition, TFAP2A binding sites have been observed near other nsCL/P candidate genes from GWAS, i.e., *ARHGAP29*, *FILIP1L*, *TP63*, *GREM1*, *TPM1*, *ARID3B*, *ADCY9*, and *NOG*. With the exception of *TP63*, all of these were also expressed in HEPM. We also observed TFAP2A ChIP-seq peaks downstream of *CD58* (~16 kb), and *PTGS2* (two peaks; located ~261 bp and ~75 kb upstream), both of which have been implicated in the TFAP2A-gene regulatory network described by Razaghi-Moghadam et al.³⁹. Together, these genes represent interesting candidates for further studies of nsCL/P in HEPM cells. Based on the well-established interaction between *TFAP2A* and *IRF6*, we specifically looked at TFAP2A binding sites near *IRF6*. We could not detect any TFAP2A ChIP-seq peaks within 1 Mb of the TSS of *IRF6*, and also *IRF6* expression was only merely detected above background in HEPM. Thus, while there is robust interaction of *TFAP2A* and *IRF6* in epithelial cells²⁹, our data indicate that such effects might not be present in palatal mesenchymal cells. Importantly, in the present study, the genes were assigned to the hc-peaks on the basis of proximity. Therefore, long-distance regulatory effects of TFAP2A binding sites would have been missed by design.

In the comparison of TFAP2A ChIP-seq peaks in HEPM cells and HeLa S3 cells, almost half of the ChIP-seq peaks in HEPM cells were found to overlap between both cell lines. Besides indicating that these peaks are true positives, this finding also suggests that some TFBS are specific to HEPM cells (and, presumably, to craniofacial processes). In support of this we observed a strong overlap of TFAP2A sites within DNase1 hypersensitivity sites (94.1% overlap), and H3K27ac signals (57.8% overlap) from human embryonic craniofacial tissue. This indicates that these TFAP2A ChIP-seq peaks reside at chromatin sites that are accessible to TF binding during human facial development. The output of our GO analysis provides further support for a role of TFAP2A binding sites in facial development, as craniofacial-specific terms such as “abnormality of facial soft tissue”, “abnormal palatine bone morphology”, or “abnormal palatal shelf elevation” were significantly enriched. We also found evidence for a contribution of the transforming growth factor beta pathway in HEPM, which is implicated in the epithelial-mesenchymal-transition processes that occur during secondary palate formation^{50,51}.

Finally, our data set was used to analyze whether common nsCL/P risk alleles located at the TFAP2A ChIP-seq peaks in HEPM might contribute to disease pathomechanism. We did not observe any risk SNP at test-wide significance within the TFAP2A ChIP-seq peaks. Four SNPs were detected at suggestive significance, which does not exceed the number expected by chance. In addition, no allele-specific effect was observed for the one variant of those four that was heterozygous. Further analyses outside of established GWAS risk loci identified 18 SNPs with allele-specific TFAP2A binding, but again, none of them showed a nominally significant association with nsCL/P. Thus, our data do not provide evidence that the biological effect at any of the common risk variants is genetically mediated through differential TFAP2A binding in mesenchymal cells.

Our study is influenced by some limitations. First, we investigated a two-dimensional in vitro model, which lacks the complexity of cell-to-cell interactions. Particularly, we might have missed effects that are driven by the interaction between epithelial and mesenchymal cells, and/or environmental clues, such as those present in three-dimensional embryonic palate^{50–52}. This could be overcome by future investigations of three-dimensional organoid systems, such as recently established by Wolf et al.⁵³ and Hughes et al.⁵⁴. Second, the integration of TFAP2A binding peaks and genetic risk variants from GWAS meta-analyses only informs about common risk variants, but does not provide information on potential effects of rare variants on TFAP2A binding. Therefore, the integration of whole genome sequencing data would be required, e.g. those recently described by Bishop et al.⁵⁵. Also, the analyses of allele-specific effects were limited by the fact that only one HEPM-donor was available, resulting in a limited number of heterozygous sites usable for the analysis. Finally, our approach does not cover effects of (i) TFAP2A in other cell types (e.g. oral facial epithelium), (ii) other TFs in HEPM (e.g. *MSX1*, which we also observed at high expression in HEPM), or (iii) other types of gene regulation (e.g. miRNA^{56,57}). While potential TFs can be identified as presented in this study, alternative approaches also exist—for instance, integrating GWAS risk SNPs and TF databases such as JASPAR³⁰. This, however, would then require subsequent identification of the relevant cell type for in vitro investigation, which is still a challenge for embryonic human phenotypes.

Taken together, the present data suggest that TFAP2A binding in HEPM might play a role in normal craniofacial development, and indicate a set of presumably craniofacial-specific TFAP2A ChIP-seq peaks near nsCL/P candidate genes, which might be functionally followed up. No strong evidence was obtained for the hypothesis that genetic variability at these sites contributes to nsCL/P etiology. Despite this, similar analyses in other cell types of relevance to craniofacial development might provide novel insights into our understanding of genetically-mediated nsCL/P risk.

Methods

Cell culture. A human embryonic palatal mesenchyme cell line was commercially available at ATCC (ATCC Cat# CRL-1486). Upon purchase these cells were cultured in DMEM-Dulbecco's Modified Eagle Medium (high Glucose) with 10% heat inactivated fetal bovine serum (FBS) and Penicillin/Streptomycin at a final concentration of 1%. Cells were stored in an incubator (37 °C, 5% CO₂), with a change of medium every two days. After five days of cultivation, the cells were split using 0.25% Trypsin/EDTA. DNA and RNA were extracted from the cells using the DNeasy[®] Blood & Tissue kit and the RNeasy[®] kit (QIAGEN, Germany), respectively, in accordance with the manufacturer's protocols.

RNA-Seq. To capture the expression profile of HEPM cells, 3'mRNA-Seq was performed in triplicate. For library preparation, the QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen, Austria) was used, in accordance with the manufacturer's instructions. Sequencing was performed at 1 × 50 bp on an Illumina HiSeq2500, with a minimum depth of ~ 20 mio reads per sample. Reads were quality checked using FastQC (v0.11.7), adapters were trimmed using bbdduk (BBMap v37.44), and reads were aligned to the GRCh37/hg19 reference genome using STAR Aligner (v2.5.2b). Gene expression was quantified using featureCounts (v1.5.1), and the Ensembl Human GRCh37.p13 annotation as a reference. Quality control was carried out using MultiQC (v1.2). Genes were classified as "expressed" if the average number of aligned mRNA reads was ≥ 5, and "strongly expressed" if the average number of aligned mRNA reads was within the upper quartile of all genes expressed with ≥ 5 mRNA-seq reads in HEPM (i.e., ≥ 327 reads). The JASPAR CORE vertebrate assembly (2020)³⁰ was used to identify TFs among genes that are expressed in HEPM.

Selection of candidate transcription factors. To identify TFs with an involvement in craniofacial development and/or orofacial clefting, a systematic search was performed in the Pubmed database. A TF was considered to be a candidate TF if the respective gene had been reported previously: (i) as a candidate gene at a nsCL/P risk locus; (ii) in a mutated state in patients with craniofacial malformations; (iii) to result in disturbed craniofacial development when modified in animal models; or (iv) to be part of a gene regulatory network involved in facial development. To identify TFs whose potential role in nsCL/P etiology involved a change of TF binding in HEPM, this list of candidate TFs was cross-referenced with the HEPM expression data.

Genotyping. To determine genotypes for common variants in HEPM, DNA was extracted from HEPM cells. The DNA was then genotyped on an Illumina Infinium GSAv2.0 array (Illumina, USA), comprising ~700.000 SNPs with a major allele frequency >0.1%. After stringent quality control, genotypes of SNPs that were not represented on the array were imputed using IMPUTE2 (v2.3.2), the 1000 genomes phase 3 GRCh37/hg19 variants as a reference panel, and an info metric threshold of 0.5. For imputed variants with mono-allelic binding in the subsequent ChIP-seq analysis, genotypes were validated by Sanger sequencing in order to exclude imputation artifacts.

Chromatin Immunoprecipitation (ChIP-seq). The SimpleChIP[®]Enzymatic Chromatin IP Kit (#9003, Cell Signaling Technology[®], USA) was applied with minor modifications and using two replicates. Briefly, for each immunoprecipitation (IP), around 4 × 10⁶ HEPM cells were crosslinked for 10 min at room temperature using 37% formaldehyde at a final concentration of 1%. The reaction was stopped with 0.125 mM Glycine. Cells were washed twice with ice-cold PBS and scraped into a tube. Cells lysis was performed by sequential cold incubation (4 °C, on ice) with two buffers provided in the reagent kit. Chromatin was digested by 0.5 μl micrococcal nuclease per IP for 15 min at 37 °C on a constantly shaking heating block. After stopping digestion by the addition of 0.5 M EDTA, nuclei were suspended in ChIP buffer and sonicated with a Diagenode Bioruptor (settings: 50 cycles, 30 s sonication/30 s break). A fraction of the sample was used for measuring DNA-concentration using Nano Drop and to check fragmentation size via electrophoresis. A total of 0.5 μg DNA was used as the input control. A total of 0.005 μg of polyclonal ChIP Grade TFAP2A antibody (Abcam Cat# ab52222) was added to 25 μg of DNA and incubated overnight with rotation at 4 °C. Fragments were pulled down using Protein G magnetic beads, and then removed from the beads via incubation with ChIP elution buffer on a shaking thermomixer at 65 °C for 30 min. The supernatant was treated overnight with NaCl and Proteinase K in order to reverse DNA-crosslinking, and then purified using spin column tubes, as provided in the reagent kit.

Library preparation and next generation sequencing. Library preparation for sequencing was performed using the NEBNext[®] Ultra[™] II DNA Library Prep Kit for Illumina[®] (New England Biolabs[®] GmbH, USA), in accordance with the manufacturer's protocol. AmPure XP beads (Beckman Coulter[™], USA) were used for the cleanup steps. Since the amount of input DNA was < 50 ng, no size selection was performed. Equimolar pooling of the samples was then performed, and the quality of the library was controlled on an Agilent High Sensitivity D1000 system. The samples were diluted to a final concentration of 2 nM. To achieve sufficient coverage, each replicate was sequenced twice on an Illumina MiSeq v2. This yielded ~ 20 million 2 × 250 bp paired end reads for each sample (ChIP and input control), in accordance with the ChIP-seq guidelines of the ENCODE consortium⁵⁸.

Bioinformatic processing of ChIP-Seq data. Quality control of the fastq-files was performed using FastQC (v0.11.7). Adapter sequences were cut-off using Cutadapt (v1.15), and reads were trimmed to a maximum length of 200 bp. For each sample and replicate, fastq files of both sequencing runs were merged and aligned to the GRCh37/hg19 reference genome using Bowtie2 (v2.3.4). Peak calling was performed using

MACS2 (v2.1.1). Quality control included the retention of peaks with a fold enrichment (FE) of 5–50, and a false discovery rate (FDR) < 0.05. Peaks mapping to irregular chromosomes and ENCODE blacklist regions were removed²⁹ using BEDtools (v2.27.0). Peaks were then visually inspected in the Integrative Genomics Viewer (IGV, v2.4.6). To extract a high-confidence set of peaks (termed “hc-peaks”), only those regions included in both replicates were included in the analysis.

Motif discovery. Motif discovery was performed using the genomic sequence around the summit of each hc-peak (\pm 50 bp) and MEME ChIP (v5.1.0). The similarity between TFAP2A motifs from the JASPAR Core 2018 assembly³¹ and the most significant motif from ChIP-seq was further quantified and displayed by presenting the optimal alignments with the Tomtom motif comparison tool (v5.1.0).

Comparison with other data sets. First, hc-peak positions were compared to TFBS in another TFAP2A ChIP-seq data set that had been obtained in HeLa S3 cells by ENCODE³². Here, concordant regions were defined as those for which HEPM TFAP2A ChIP-seq peaks showed a \geq 50% overlap with the base pairs of HeLa S3 TFAP2A sites. Second, HEPM TFAP2A ChIP-seq peaks were compared with H3K27ac histone marks and DNase1 hypersensitivity sites derived from craniofacial tissues (Epigenomic Atlas of Human Craniofacial Development) at Carnegie stage 15 (CS15)³⁴, indicating active regulatory elements in a human developmental stage that equated with the time point of the HEPM cells during craniofacial development. Since the ChIP-seq peak distribution for TFs and histone modifications differs, overlaps were defined using separate cut-offs (TF: overlap defined as 50% overlap; histone modifications: overlap defined as 1 bp). For both data sets, enrichment of the identified hc-peaks was determined using Chi2-test (1df).

Allele-specific binding. nsCL/P associated SNPs that were located within the hc-peaks and predicted to be in a heterozygous state in the HEPM cells were extracted, and corresponding allele counts from ChIP-seq were analyzed using ABC (v.1.3). SNPs were considered allele-specific variants if one of the two alleles was overrepresented at a statistically significant level (P binomial < 0.05). The subset of hc-peaks that contained SNPs with allele-specific TF binding was compared to TFAP2A ChIP-seq peaks in HeLa S3 cells³² and chromatin marks in CS15 craniofacial tissue³⁴.

Gene ontology analysis. Gene Ontology (GO) enrichment analysis was performed using the Genomic Regions Enrichment of Annotations Tool (GREAT, v4.0.4) with default parameters (5 kb upstream, 1 kb downstream, and 1 Mb maximum extension with inclusion of curated regulatory domains). GO biological processes, human phenotypes, and mouse single knockout phenotypes were considered. For these analyses, three subsets were defined: (i) all hc-peaks; (ii) hc-peaks that contained SNPs with allele-specific effects; and (iii) hc-peaks that overlapped with H3K27ac markers and DNase1 hypersensitivity regions in CS15 craniofacial tissue but not with the TFAP2A ChIP-seq peaks in HeLa S3.

Integration of nsCL/P GWAS data. Positional data of the hc-peaks were integrated with results from our recent nsCL/P GWAS meta-analysis⁹. This imputed dataset contains nsCL/P association data for \sim 8.01 million variants (meta_{all} as described in Ludwig et al.⁹). Briefly, this study included individuals of European (Bonn GWAS¹⁷: 399 cases and 1318 controls; Baltimore study¹⁵: 666 European case-parent trios) and of Asian ancestry (Baltimore study¹⁵: 795 Asian case-parent trios). The association P -values of the GWAS meta-analysis were not corrected for multiple testing. Only SNPs with an info score > 0.8 in the GWAS meta-analysis were retrieved, and pairs of SNPs in high linkage disequilibrium were identified using LD link (v.5; all populations). This selection of SNPs was intersected with topological associated domains in embryonal stem cells, as identified by Dixon et al.³⁵, and information on the 45 nsCL/P risk loci as described in Welzenbach et al.²¹.

Data availability

The ChIP-seq datasets generated in the present study are available at the Gene Expression Omnibus (GEO) repository (accession numbers GSE169341 and GSE169342). The original GWAS datasets on which the GWAS meta-analysis of Ludwig et al.⁹ is based are available at Zenodo (<https://doi.org/10.5281/zenodo.3724148>; Bonn GWAS; Mangold et al.¹⁷) and dbGap (dbGaP: phs000094; Baltimore study; Beaty et al.¹⁶). References and online availability of datasets and tools employed in the project workflow are provided in Supplementary Table S15.

Received: 3 August 2021; Accepted: 17 January 2022

Published online: 02 February 2022

References

- Mossey, P. A., Little, J., Munger, R. G., Dixon, M. J. & Shaw, W. C. Cleft lip and palate. *Lancet* 374, 1773–1785 (2009).
- Leslie, E. J. & Marazita, M. L. Genetics of Orofacial Cleft Birth Defects. *Curr. Genet. Med. Rep.* 3, 118–126 (2015).
- Tolarova, M. & Cervenka, J. Classification and birth prevalence of orofacial cleft. *Am. J. Med. Genet.* 75, 126–137 (1998).
- Dixon, M. J., Marazita, M. L., Beaty, T. H. & Murray, J. C. Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* 12, 167–178 (2011).
- Grosen, D. et al. Risk of oral clefts in twins. *Epidemiology* 22, 313–319 (2011).
- Grosen, D. et al. A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *J. Med. Genet.* 47, 162–168 (2010).
- Beaty, T. H., Marazita, M. L. & Leslie, E. J. Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *F1000 Res* 5, 2800 (2016).

8. Yu, Y. *et al.* Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nat. Commun.* **8**, 14364 (2017).
9. Ludwig, K. U. *et al.* Imputation of orofacial clefting data identifies novel risk loci and sheds light on the genetic background of cleft lip \pm cleft palate and cleft palate only. *Hum. Mol. Genet.* **26**, 829–842 (2017).
10. Ludwig, K. U. *et al.* Meta-analysis reveals genome-wide significance at 15q13 for nonsyndromic clefting of both the lip and the palate, and functional analyses implicate GREM1 as a plausible causative gene. *PLOS Genet.* **12**, e1005914–e1005914 (2016).
11. Ludwig, K. U. *et al.* Genome-wide meta-analyses of nonsyndromic cleft lip with or without cleft palate identify six new risk loci. *Nat. Genet.* **44**, 968–971 (2012).
12. Leslie, E. J. *et al.* Genome-wide meta-analyses of nonsyndromic orofacial clefts identify novel associations between FOXE1 and all orofacial clefts, and TP63 and cleft lip with or without cleft palate. *Hum. Genet.* **136**, 275–286 (2017).
13. Leslie, E. J. *et al.* A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Hum. Mol. Genet.* **25**, 2862–2872 (2016).
14. Sun, Y. *et al.* Genome-wide association study identifies a new susceptibility locus for cleft lip with or without a cleft palate. *Nat. Commun.* **6**, 6414 (2015).
15. Beaty, T. H. *et al.* A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nat. Genet.* **42**, 525–529 (2010).
16. Moreno, L. M. *et al.* FOXE1 association with both isolated cleft lip with or without cleft palate, and isolated cleft palate. *Hum. Mol. Genet.* **18**, 4879–4896 (2009).
17. Mangold, E. *et al.* Genome-wide association study identifies two susceptibility loci for nonsyndromic cleft lip with or without cleft palate. *Nat. Genet.* **42**, 24–26 (2010).
18. Grant, S. F. A. *et al.* A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J. Pediatr.* **155**, 909–913 (2009).
19. Birnbaum, S. *et al.* Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat. Genet.* **41**, 473–477 (2009).
20. Rahimov, F. *et al.* Disruption of an AP-2 α binding site in an IRF6 enhancer is associated with cleft lip. *Nat. Genet.* **40**, 1341–1347 (2008).
21. Welzenbach, J. *et al.* Integrative approaches generate insights into the architecture of non-syndromic cleft lip with or without cleft palate. *HGGADVANCE* **2**, 100038 (2021).
22. Thieme, F. & Ludwig, K. U. The role of noncoding genetic variation in isolated orofacial clefts. *J. Dent. Res.* **96**, 1238–1247 (2017).
23. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015).
24. Greene, C. S. *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).
25. Liu, H. *et al.* Identification of common non-coding variants at 1p22 that are functional for non-syndromic orofacial clefting. *Nat. Commun.* **8**, 14759 (2017).
26. Dhulipala, V. C., Welshons, W. V. & Reddy, C. S. Inhibition of human embryonic palatal mesenchymal cell cycle by secalonic acid D: A probable mechanism of its cleft palate induction. *Orthod. Craniofac. Res.* **7**, 227–236 (2004).
27. Yoneda, T. & Pratt, R. M. Mesenchymal cells from the human embryonic palate are highly responsive to epidermal growth factor. *Science* **213**, 563 (1981).
28. Milunsky, J. M. *et al.* TFAP2A mutations result in branchio-oculo-facial syndrome. *Am. J. Hum. Genet.* **82**, 1171–1177 (2008).
29. Koussa, Y. A. & Schutte, B. C. Toward an orofacial gene regulatory network. *Dev. Dyn.* **245**, 220–232 (2016).
30. Fornes, O. *et al.* JASPAR 2020: Update of the open-access database of transcription factor binding profiles. *Nucl. Acids Res.* **48**, D87–D92 (2020).
31. Khan, A. *et al.* JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucl. Acids Res.* **46**, D260–D266 (2018).
32. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 208 (2012).
33. Gey, G. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res* **12**, 264–265 (1952).
34. Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J. & Cotney, J. High-resolution epigenomic atlas of human embryonic craniofacial development. *Cell Rep.* **23**, 1581–1597 (2018).
35. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
36. Huo, Y., Li, S., Liu, J., Li, X. & Luo, X.-J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat. Commun.* **10**, 670 (2019).
37. Benaglio, P. *et al.* Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. *Nat. Genet.* **51**, 1506–1517 (2019).
38. Schorle, H., Meier, P., Buchert, M., Jaenisch, R. & Mitchell, P. J. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **381**, 235–238 (1996).
39. Razaghi-Moghadam, Z., Namipashaki, A., Farahmand, S. & Ansari-Pour, N. Systems genetics of nonsyndromic orofacial clefting provides insights into its complex aetiology. *Eur. J. Hum. Genet.* **27**, 226–234 (2019).
40. Brewer, S., Feng, W., Huang, J., Sullivan, S. & Williams, T. Wnt1-Cre-mediated deletion of AP-2 α causes multiple neural crest-related defects. *Dev. Biol.* **267**, 135–152 (2004).
41. de Croz , N., Maczkowiak, F. & Monsoro-Burq, A. H. Reiterative AP2 α activity controls sequential steps in the neural crest gene regulatory network. *Proc. Natl. Acad. Sci. USA* **108**, 155–160 (2011).
42. Mitchell, P. J., Timmons, P. M., H bert, J. M., Rigby, P. W. & Tjian, R. Transcription factor AP-2 is expressed in neural crest cell lineages during mouse embryogenesis. *Genes Dev.* **5**, 105–119 (1991).
43. Xu, J. *et al.* FGF8 signaling alters the osteogenic cell fate in the hard palate. *J. Dent. Res.* **97**, 589–596 (2018).
44. Green, R. M. *et al.* Tfp2 α -dependent changes in mouse facial morphology result in clefting that can be ameliorated by a reduction in Fgf8 gene dosage. *Dis. Model. Mech.* **8**, 31–43 (2015).
45. Medio, M. *et al.* Wnt/ β -catenin signaling and Msx1 promote outgrowth of the maxillary prominences. *Front. Physiol.* **3**, 375 (2012).
46. Zhang, Z. *et al.* Rescue of cleft palate in Msx1-deficient mice by transgenic Bmp4 reveals a network of BMP and Shh signaling in the regulation of mammalian palatogenesis. *Development (Camb, Engl)* **129**, 4135–4146 (2002).
47. Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Res.* **30**, 35–48 (2020).
48. Matsumura, K. *et al.* Sprout2 controls proliferation of palate mesenchymal cells via fibroblast growth factor signaling. *Biochem. Biophys. Res. Commun.* **404**, 1076–1082 (2011).
49. Goodnough, L. H., Brugmann, S. A., Hu, D. & Helms, J. A. Stage-dependent craniofacial defects resulting from Sprout2 overexpression. *Dev. Dyn.* **236**, 1918–1928 (2007).
50. Nakajima, A., Shuler, F. C., Gulka, A. O. D., Hanai, J.-I., TGF- β Signaling and the Epithelial-Mesenchymal Transition during Palatal Fusion. *Int. J. Mol. Sci.* **19**, 3638 (2018).
51. Chiquet, M., Blumer, S., Angellini, M., Mitsiadis, T. A. & Katsaros, C. Mesenchymal remodeling during palatal shelf elevation revealed by extracellular matrix and F-actin expression patterns. *Front. Physiol.* **7**, 392 (2016).

52. Antoni, D., Burckel, H., Josset, E. & Noel, G. Three-dimensional cell culture: A breakthrough in vivo. *Int. J. Mol. Sci.* **16**, 5517–5527 (2015).
53. Wolf, C. J. *et al.* Development of an organotypic stem cell model for the study of human embryonic palatal fusion. *Birth Defects Res.* **110**, 1322–1334 (2018).
54. Hughes, A. J. *et al.* Engineered tissue folding by mechanical compaction of the mesenchyme. *Dev. Cell* **44**, 165–178.e6 (2018).
55. Bishop, M. R. *et al.* Genome-wide enrichment of De Novo coding mutations in orofacial cleft trios. *Am. J. Hum. Genet.* **107**, 124–136 (2020).
56. Garland, M. A. *et al.* Role of epigenetics and miRNAs in orofacial clefts. *Birth Defects Res.* **112**, 1635–1659 (2020).
57. Stüssel, L. G. *et al.* miRNA-149 as a candidate for facial clefting and neural crest cell migration. *J. Dental Res.* 220345211038203 (2021).
58. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
59. Carroll, T. S., Liang, Z., Salama, R., Stark, R. & de Santiago, I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front. Genet.* **5**, 75 (2014).

Acknowledgements

We thank Nadine Siebert for support with cell culture, Christine Schmäler for manuscript editing, and the laboratory staff of the Institute of Human Genetics Bonn for their technical assistance and insightful discussions. This research was funded by the German Research Foundation (DFG, LU1944-3/1 to Kerstin U. Ludwig), and the BONFOR program of the Medical Faculty of the University of Bonn (O-149.0125 to Katharina L.M. Ruff).

Author contributions

K.U.L. conceptualized the project and provided funding. K.L.M.R., J.F., F.T. and R.H. performed the experiments. K.L.M.R., J.W. and J.G. processed the data and performed bioinformatic analysis. E.M. and M.K. provided access to samples and resources. K.L.M.R. prepared Tables and Figures. K.L.M.R. and K.U.L. drafted the manuscript. All authors approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-05876-7>.

Correspondence and requests for materials should be addressed to K.U.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

2. Danksagung

Zunächst möchte ich dem BONFOR-Förderungsprogramm der Universität Bonn für die mir zugekommene fachliche und finanzielle Förderung im Rahmen des SciMed-Promotions-Stipendiums danken.

Es gibt viele Menschen, ohne die ich diese Promotion nicht hätte anfertigen können.

Ich danke meiner Doktormutter Frau Dr. rer. nat. Kerstin U. Ludwig für Ihre hervorragende Betreuung während des Projektes und ihre zahlreichen konstruktiven Anmerkungen und Vorschläge. Ich habe während dieser Zeit viel von ihr gelernt, wovon ich sicher in meinem späteren Leben weiterhin profitieren werde.

Außerdem möchte ich mich bei meiner Arbeitsgruppe bedanken: Bei Dr. rer. agr. Julia Heggemann, Dr. rer. nat. Ronja Hollstein und PhD Frederic Thieme die mir jederzeit mit Rat und Tat zur Seite standen. Insbesondere hat mir die Arbeit mit Julia Fazaal sehr viel Spaß gemacht und ich schätze sie als Freundin. Sehr positiv wird mir auch die Kooperation innerhalb der verschiedenen Arbeitsgruppen des Instituts für Humangenetik der Universität Bonn in Erinnerung bleiben.

Meine Familie, meine Freund*innen und mein Partner Daniel Adams haben mich immer unterstützt und ein offenes Ohr für mich gehabt. Sie haben mich auch in stressigen oder schwierigen Phasen motiviert.