

Aus dem Lehrstuhl
für Humangenetik
Prof. Dr. Bernhard Weber
der Fakultät für Medizin
der Universität Regensburg

**Identifikation und Charakterisierung von koregulierten Genen und
regulatorischen Genclustern**

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Martin Kellner

2023

Aus dem Lehrstuhl
für Humangenetik
Prof. Dr. Bernhard Weber
der Fakultät für Medizin
der Universität Regensburg

**Identifikation und Charakterisierung von koregulierten Genen und
regulatorischen Genclustern**

Inaugural – Dissertation
zur Erlangung des Doktorgrades
der Medizin

der
Fakultät für Medizin
der Universität Regensburg

vorgelegt von
Martin Kellner

2023

Dekan: Prof. Dr. Dirk Hellwig

1. Berichterstatter: Prof. Dr. Bernhard Weber

2. Berichterstatter: Prof. Dr. Birte Kehr

Tag der mündlichen Prüfung: 05.06.2024

Inhaltsverzeichnis

Inhaltsverzeichnis	3
Zusammenfassung	4
1 Einleitung	6
1.1 Das nicht-kodierende Genom	6
1.2 Regulation von Genexpression und koregulierte Gene	7
1.3 Ansätze zur Identifikation koregulatorischer Beziehungen	9
1.4 Modifikation von Genexpression in therapeutischen Ansätzen	12
1.5 Ziel der Arbeit	15
2 Material und Methoden	16
2.1 Berechnung der Expression Quantitative Trait Loci (eQTL)	16
2.2 Identifikation von genregulatorischen Clustern	17
2.3 Integration von Datensätzen mit haploinsuffizienten Genen	18
2.4 Programme und Tools	19
3 Ergebnisse	20
3.1 Identifikation von regulatorischen Clustern	20
3.2 Genzentrierte Analyse	21
3.2.1 Zugrundeliegende Genexpressionsdaten	21
3.2.2 Einfluss der Stichprobengröße auf die Anzahl an eGenen sowie auf die Anzahl koregulierter eGene	23
3.2.3 Kollokalisationsanalyse genetischer Signale	25
3.2.4 Analyse koregulierter Genpaarungen	26
3.3 Clusterzentrierte Analyse	27
3.4 Gewebezentrierte Analyse	30
3.5 Klinischer Bezug von regulatorischen Clustern	33
4 Diskussion	36
Abkürzungsverzeichnis	I
Abbildungsverzeichnis	II
Anhang	III
Literaturverzeichnis	V

Zusammenfassung

Seit der vollständigen Sequenzierung des menschlichen Genoms stehen nicht-kodierende DNA-Abschnitte und die Regulation von Genexpression zunehmend im Mittelpunkt der Forschung. Trotz der Erfassung und Analyse von mehr und mehr regulatorischen Elementen bleiben einige regulatorische Mechanismen und Beziehungen weiterhin verborgen. Gleichzeitig können besonders im Bereich genommodifizierender Methoden basierend beispielsweise auf CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats)-Systemen große Fortschritte verzeichnet werden. Bei der Evaluation dieser letztgenannten Methode zur gezielten Veränderung der Genregulation müssen genaue Kenntnisse zu den entsprechenden genregulatorischen Beziehungen und koregulierten Genen vorliegen, um unerwünschte Effekte, sogenannte Off-Target-Effekte, vorab erkennen und bewerten zu können.

Bisher wurden in genomweiten Ansätzen vor allem *Chromosome Conformation Capture* Methoden angewandt, die es erlauben, die dreidimensionale Architektur des Genoms zu analysieren. Anhand der Messung physischer Chromatininteraktionen können Rückschlüsse auf regulatorische Elemente und deren Beziehung zueinander gezogen werden. Trotz verbesserter Methoden und zunehmender Auflösungsvermögen können in komplexen Regionen regulatorische Beziehungen einzelner Gene nicht sicher zugeordnet werden, sodass die Identifikation regulatorischer Beziehungen auch auf alternative Ansätze gestützt werden muss.

In dieser Arbeit wurden koregulierte Gene basierend auf *expression Quantitative Trait Locus* (eQTL) Studien durch Kolokalisationsanalyse der zugrundeliegenden genetischen Signale identifiziert. Die Untersuchung erfolgte in 49 verschiedenen Geweben anhand von Genotyp- und Genexpressionsdaten des *Genotype Tissue-Expression* (GTEx) Projekts. Die Resultate wurden unter verschiedenen Gesichtspunkten analysiert und in einer umfassenden Dantebank leicht zugänglich gemacht, um spezifische Fragestellungen, beispielsweise zu einzelnen Genen, gezielt beantworten zu können.

Über alle Gewebe hinweg zeigten sich insgesamt 14.636 (36,7 %) der 39.832 exprimierten Gene mit mindestens einem anderen Gen koreguliert, sodass eine Zahl von 49.637 koregulierten Genpaarungen identifiziert werden konnte. Dabei waren die koregulierten Gene z. T. in großen Abständen zueinander lokalisiert, was sich gut mit

der Bildung von dreidimensionalen Chromatinstrukturen vereinbaren lässt. Die koregulierten Genpaarungen wurden zu 14.727 einzigartigen regulatorischen Clustern mit Größen von zwei bis zu neun koordiniert regulierten Genen zusammengeführt. Entsprechend den Ergebnissen bereits veröffentlichter Studien wurden Effekte, beispielsweise von bidirektionalen Promotern, die eine Koregulation von Genen bedingen können, in einigen regulatorischen Clustern unmittelbar abgebildet. In einer gewebezentrierten Analyse stellte sich heraus, dass die regulatorischen Effekte eines jeden Gewebes in mindestens 80 regulatorischen Clustern mit anderen Geweben übereinstimmten. Schließlich wurde am Beispiel haploinsuffizienter Gene gezeigt, wie sich im Rahmen von therapeutischen Ansätzen die entsprechende Hochregulation der Expression eines Gens auch auf koregulierte Gene auswirken kann. Von 713 bekannten haploinsuffizienten Genen wurden 231 (32,4 %) in verschiedenen regulatorischen Clustern gefunden und zeigten sich mit weiteren Genen koreguliert. Somit konnten aufgrund der Ergebnisse der hier präsentierten Studie allein im Kontext der Haploinsuffizienz bereits 414 Hinweise auf potenzielle Off-Target-Effekte gegeben werden, die sonst möglicherweise nicht berücksichtigt worden wären.

1 Einleitung

1.1 Das nicht-kodierende Genom

Die Entschlüsselung des menschlichen Genoms durch das Humangenomprojekt ist eine der großen naturwissenschaftlichen Errungenschaften des letzten Jahrhunderts (1). Das Vorliegen der vollständigen DNA-Sequenz durch den Abschluss des Projekts eröffnet eine neue Untersuchungsdimension in der genetischen Forschung. Rückblickend konnte das Projekt einerseits einige hypothetische Annahmen über das Genom bestätigen und beispielsweise Erwartungen an die Identifizierung bekannter sowie unbekannter Gene oder den Nachweis hoher Anteile repetitiver DNA-Sequenzen erfüllen (1,2). Andererseits stellt sich als eines der überraschendsten Resultate des Humangenomprojekts und seiner Folgeprojekte heraus, dass mehr als 98 % des menschlichen Erbguts nicht für Proteine kodiert (3,4).

Diese nicht-kodierenden Bereiche rücken in den verschiedensten Feldern der Forschung immer mehr in den Fokus, nachdem anfangs vor allem die nur etwa 2 % ausmachenden Protein-kodierenden Regionen des Genoms mittels Sequenzierungsmethoden wie *whole exome sequencing* (WES) gezielt untersucht wurden. Zunächst konnten so durch Identifikation exonischer Varianten große Fortschritte in der Ursachenforschung besonders im Falle monogenetischer Erkrankungen verzeichnet werden (5,6). Letztlich bleiben die genetischen Ursachen einiger Erbkrankheiten jedoch ungeklärt, womit das nicht-kodierende Genom zu einem zentralen Forschungsgebiet wird (7). Gleichzeitig stehen aufgrund technologischer Fortschritte bei Sequenzierungsmethoden in steigender Anzahl Genotypdatensätze zur Durchführung von *genomweiten Assoziationsstudien* (GWAS) bereit. Aufgrund der zunehmenden Trennschärfe und Präzision dieser Fall-Kontroll-Studien können mittels statistischer Methoden immer mehr genetische Varianten im Genom identifiziert werden, die mit einem erhöhten Krankheitsrisiko assoziiert sind (8,9). Eine tiefere Betrachtung der Ergebnisse von GWAS zeigt, dass sich der Großteil der gefundenen krankheitsassoziierten Varianten nicht in Protein-kodierenden Bereichen des Genoms, sondern in intergenischen oder intronischen Genomabschnitten befindet (10). Auch Nachweise einer Vielzahl an noncoding RNAs (ncRNAs) verschiedener Länge, die durch umfangreiche Transkription des nicht-kodierenden Genoms produziert werden,

rücken die nicht-kodierenden Sequenzbereiche zunehmend in den Mittelpunkt der Forschung (11,12).

Diese Ausrichtung der Forschungsbemühungen hat zur Folge, dass im Laufe der Zeit im nicht-kodierenden Genom mehr und mehr strukturelle Elemente identifiziert wurden, welche die Genexpression regulieren (13). Zudem spielt eine präzise Regulation von Genexpression nicht nur bei Krankheitsätiologien, sondern beispielsweise auch in der Entwicklung des menschlichen Embryos eine wichtige Rolle (14). Somit scheinen besonders bessere Kenntnisse der Regulationsmechanismen von Genexpression der Schlüssel zu Fortschritten in vielen Forschungsbereichen zu sein.

1.2 Regulation von Genexpression und koregulierte Gene

Genregulation in Eukaryoten ist ein komplexer Mechanismus, der fein abgestimmte Prozesse auf mehreren Ebenen erfordert (15). Regulatorische Elemente wie Enhancer, Promoter und Isolationsfaktoren stellen sicher, dass die richtigen Gene zum richtigen Zeitpunkt in der richtigen Menge transkribiert werden. Oft bestehen regulatorische Beziehungen nicht exklusiv, sondern überschneiden sich zwischen den betroffenen Genen, was die Komplexität weiter erhöht. Beispielsweise zeigen Studien, dass Genpaarungen in divergenter Orientierung einen gemeinsamen bidirektionalen Promoter haben können, wodurch die Genexpression koreguliert wird (16,17). Weiter kann ein einzelner Enhancer mehrere Gene regulieren, während zugleich ein Gen von mehreren Enhancern reguliert werden kann (18). Oftmals bleiben die zugrundeliegenden Mechanismen der Koregulation von Genen jedoch verborgen (19).

Bei Betrachtung auf höherer Ebene stellt man fest, dass regulatorische Elemente wie Enhancer, Silencer und Promoter genau wie aktiv transkribierte Gene vor allem in weniger dicht gepackten, euchromatischen Regionen lokalisiert sind, sodass sie für die Transkriptionsmaschinerie leicht zugänglich sind (20). Neben der Nukleosomenstruktur spielt auch die dreidimensionale Anordnung des Chromatins eine wichtige Rolle in der Regulation von Genexpression. So kann das Genom in *topologisch assoziierende Domänen* (TADs) gegliedert werden, wobei Sequenzabschnitte innerhalb derselben TAD vorrangig miteinander interagieren und nur begrenzt Kontakt zu benachbarten Bereichen zeigen. TADs erreichen

Größenordnungen von wenigen Kilobasen bis hin zu Megabasen und bilden übergeordnete regulatorische Einheiten im Genom (21). Regulatorische Elemente wie Enhancer und Promoter stehen in vielen Fällen über große Distanzen hinweg in regulatorischen Beziehungen zueinander. Solche Interaktionen, die eine gewisse örtliche Nähe der beteiligten Elemente voraussetzen, werden durch die Bildung von DNA-Schleifen innerhalb der TADs ermöglicht (**Abbildung 1**) (22,23).

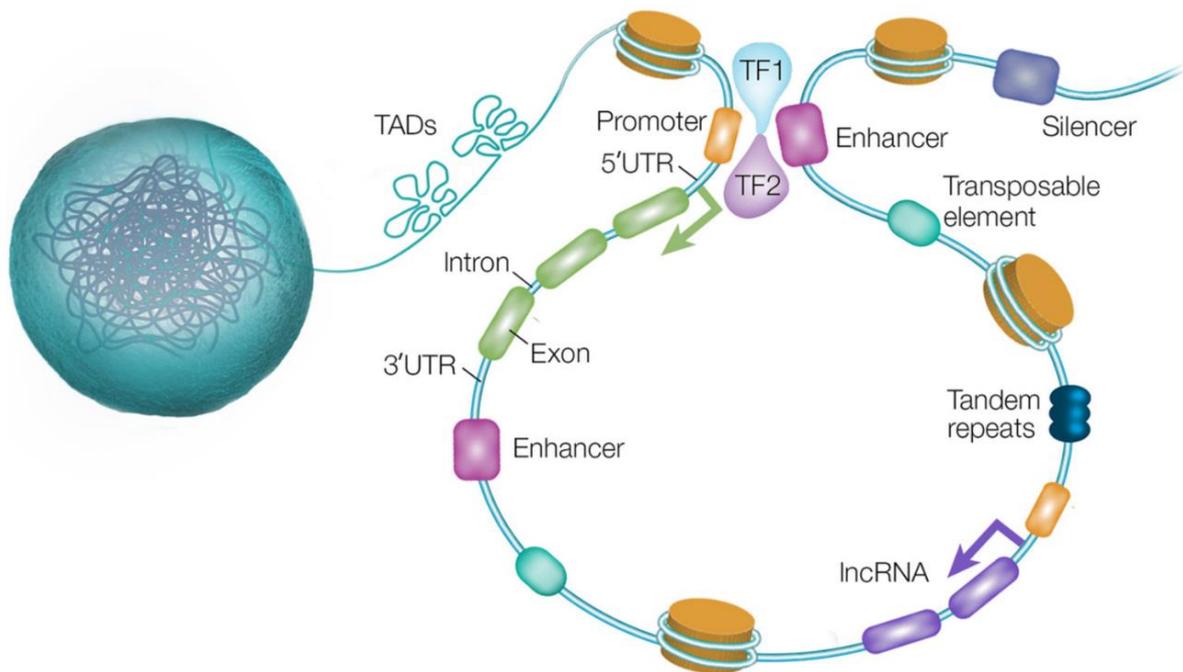


Abbildung 1: Funktionelle Anordnung regulatorischer Elemente des nicht-kodierenden Genoms im dreidimensionalen Raum

Die Abbildung zeigt schematisch, wie durch Ausbildung von DNA-Schleifen regulatorische Elemente wie Promoter und Enhancer innerhalb derselben TAD in räumlicher Nähe zueinander interagieren können, obwohl sie auf dem linearen Genom in großen Abständen zueinander lokalisiert sind. Nach French et al., 2020 (24).

Während die strukturellen Kontakte durch DNA-Schleifen innerhalb der TADs zwischen verschiedenen Zellen und Zelltypen variieren, sind deren Grenzen über mehrere Zellen und Zelltypen hinweg relativ konstant (21,25). Gene innerhalb derselben TAD tendieren dazu, zusammen exprimiert zu werden und sind mit einer höheren Wahrscheinlichkeit koreguliert als Gene, die in unterschiedlichen TADs lokalisiert sind (24,26). Dennoch gibt es Nachweise, dass manche Gene trotz Koregulation unterschiedlichen TADs zugeordnet sind (19) und dass auch TADs selbst in dynamischen „meta-TAD“-Anordnungen untereinander Interaktionen zeigen (27). Zusammengefasst lässt sich feststellen, dass das Wissen über die komplexen genregulatorischen Mechanismen im letzten Jahrzehnt einen enormen Zuwachs

erfahren hat und einen Eindruck davon vermittelt, wie vielschichtig und verzahnt regulatorische Prozesse sind. Trotz dieser Erkenntnisse bleibt das Gesamtbild der Regulation von Genexpression und besonders das Wissen über koregulierte Gene lückenhaft. Deshalb wurden verschiedene komplexe Methoden entwickelt, um solche regulatorischen Mechanismen und Interaktionen besser entschlüsseln zu können.

1.3 Ansätze zur Identifikation koregulatorischer Beziehungen

Mit der Identifikation von strukturellen Anordnungen mit unmittelbarer Assoziation zur Regulation der Genexpression wird immer deutlicher, dass die Organisationsstruktur des Chromatins eine wichtige Rolle in der Transkription spielt. Zur Untersuchung der dreidimensionalen Anordnung des Chromatins nehmen Methoden basierend auf *Chromosome Conformation Capture* (3C) eine zentrale Rolle ein (28). Durch verbesserte Ansätze wie 4C, 5C und Hi-C können in der Analyse höhere Auflösungen erzielt und in zunehmender Zahl Kombinationen interagierender Chromatinabschnitte gleichzeitig getestet werden (29,30). Mittels des aktuellen Goldstandards Hi-C kann genomweit die Häufigkeit physischer Chromatininteraktionen einer Vielzahl verschiedener Loci gemessen werden, wodurch die Hierarchie struktureller Ebenen des Chromosoms offengelegt wird (30,31). Hierfür wird nach der Fixierung physisch interagierender Loci durch Formaldehyd das Chromatin durch Restriktionsenzyme gespalten. Dann werden die Chromatinabschnitte in starke Verdünnung gebracht, sodass sich nur die Enden zuvor fixierter Stücke in räumlicher Nähe befinden und somit im folgenden Schritt bevorzugt die Enden interagierender Fragmente ligiert werden. Daraufhin wird die Interaktionsstelle mit Biotin markiert, bevor die Fragmente mittels Hochdurchsatz-Sequenzierung analysiert werden. Durch gepaarte Endsequenzierung können die interagierenden Fragmente den entsprechenden Sequenzen auf dem Genom zugeordnet und damit kartiert werden (30,32). Analysen zeigen, dass TADs in den Randbereichen eine erhöhte Dichte an Bindungsstellen für den Isolationsfaktor CTCF aufweisen (21,33). Diese scheinen zusammen mit Ringbildenden Cohesin-Proteinkomplexen zur Stabilisierung der dreidimensionalen Struktur eine wichtige Rolle in der Definition der Grenzen von TADS zu spielen (34). Die Ansätze zur Untersuchung der dreidimensionalen Chromatinstrukturen sind somit in der Lage, aufgrund struktureller Eigenschaften einzelne Loci mit potenziell regulatorischer Rolle zu identifizieren. Dadurch können beispielsweise Rückschlüsse auf entsprechende Enhancer-Promoter Beziehungen gezogen werden.

Bei einer anderen Herangehensweise steht besonders die präzise Identifikation koregulierter Gene im Fokus, ohne dabei den Anspruch zu verfolgen, die tatsächlich ursächlichen regulatorischen Elemente wie gemeinsame Enhancer oder bidirektionale Promoter aufzudecken. Hierbei basiert die Identifikation koregulierter Gene nicht auf messbaren physischen Interaktionen einzelner Genomabschnitte, sondern auf dem Vergleich von Mustern genetischer Varianten, die mit der Expression der betrachteten Gene assoziiert sind. Grundlage für diesen Ansatz sind *expression Quantitative Trait Locus* (eQTL) –Analysen, da sie Assoziationen zwischen genetischen Varianten und Veränderungen in der Genexpression aufzeigen (**Abbildung 2**). Dazu wird zuerst die Genexpression im Zielgewebe über mRNA-Sequenzierungen gemessen und anschließend werden genetische Varianten statistisch auf eine Assoziation mit interindividuellen Veränderungen innerhalb einer großen Stichprobe in der Genexpression getestet. Von lokalen oder cis-eQTL spricht man, wenn die Untersuchung auf Varianten in einer gewissen örtlichen Nähe zum Zielgen beschränkt wird. Durch diese Restriktion kann die Aussagekraft bei kleineren Stichprobengrößen gesteigert werden (35,36).

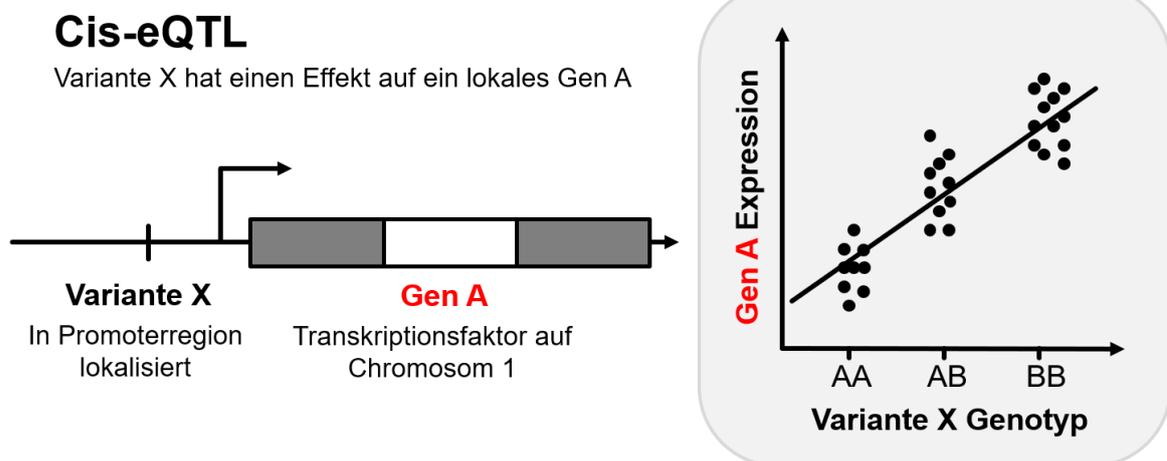


Abbildung 2: Cis-eQTL Schema

Diese Abbildung zeigt, wie durch Untersuchung von Varianten, die sich in einer gewissen Nähe zum Gen befinden, direkte und lokale Effekte identifiziert werden können. Dafür werden Datensätze von Genotypen der zu analysierenden Variante und den entsprechenden gemessenen Genexpressionen verschiedener Proben betrachtet. Durch Anwendung eines linearen Regressionsmodells werden Effektstärke und Effektrichtung der Assoziation einer Variante X mit der Expression eines Gens A ermittelt. Modifizierte Abbildung nach Westra et al., 2014 (37)

Da die Regulation der Genexpression zum Teil stark gewebeabhängig ist (38,39), werden in großer Anzahl Genotypdaten samt zugehöriger Expressionsdaten von verschiedenen Gewebeproben benötigt. Das *Genotype-Tissue Expression (GTEx)* Project hat sich die Durchführung von eQTL-Analysen in einer Vielzahl unterschiedlicher Gewebe zum Ziel gesetzt. Aus insgesamt 15.201 Proben von 49 verschiedenen Gewebetypen resultierte so ein Katalog mit über 4,2 Millionen genetischen Varianten (eVarianten), die in mindestens einem der betrachteten Gewebe einen signifikanten Effekt auf die Expression eines Gens (eGen) zeigen (40). Darunter existieren auch eVarianten, die mehr als ein eGen regulieren.

Einige Varianten treten besonders häufig zusammen auf, weil sie aufgrund ihrer örtlichen Nähe zueinander evolutionsgeschichtlich selten durch Crossing-Over Ereignisse getrennt und seither stets miteinander weitergegeben wurden. Aufgrund dieses Kopplungsungleichgewicht genannten Phänomens (Linkage Disequilibrium – LD) ist es nicht ausreichend, koregulierte Gene nur anhand einzelner signifikant assoziierter Varianten zu identifizieren (41,42). Vielmehr muss das Muster der Gesamtheit aller Varianten herangezogen werden, um koregulierte Gene unabhängig von komplexen LD-Strukturen der einzelnen Varianten unterscheiden zu können. Gleichzeitig bleibt aufgrund von LD-Strukturen die einzelne, tatsächlich kausale Variante für die gemessenen Effekte bei diesem Ansatz stets im Verborgenen (42,43). Die Gesamtheit aller auf Assoziation mit einem Gen betrachteter Varianten sowie deren relativen Beziehungen zueinander werden als genetisches Signal (**Abbildung 3**) bezeichnet und können mittels LocusZoom-Plots dargestellt werden (44).

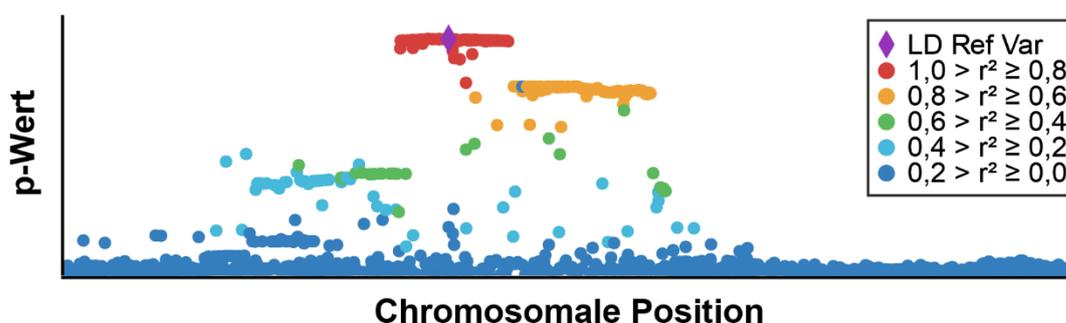


Abbildung 3: Beispiel eines genetischen Signals eines Gens im schematischen LocusZoom-Plot
Bei der Darstellung von genetischen Signalen in Locuszoom-Plots steht jeder Punkt für eine Variante, die bei der vorangestellten eQTL-Analyse auf Assoziation mit dem zu untersuchenden Gen überprüft wurde. Auf der y-Achse ist der p-Wert des linearen Regressionsmodells einer jeden Variante gegen die entsprechende chromosomale Position auf der x-Achse aufgetragen. Die Farbcodierung zeigt, welche Varianten in Kopplungsungleichgewicht zueinanderstehen.

Durch Kolokalisationsanalysen mittels statistischer Programme wie COLOC (45) oder eCAVIAR (42) werden die komplexen genetischen Signale der eQTL-Studien miteinander verglichen. Liegen kolokalisierende genetische Signale vor, können ohne exaktes Wissen über die ursächlichen regulatorischen Elemente koregulierte Gene innerhalb von TADs identifiziert werden. Größere regulatorische Verbünde an Genen können in regulatorische Cluster zusammengeführt werden. Somit stellt diese bioinformatische Methode – neben den auf physischer Interaktion basierenden Methoden wie Hi-C – einen alternativen Ansatz zur Identifikation koregulatorischer Beziehungen zwischen Genen dar und konnte in retinalem Gewebe bereits erfolgreich angewendet werden (46).

1.4 Modifikation von Genexpression in therapeutischen Ansätzen

Einerseits trägt die Identifikation koregulierter Gene im Rahmen der Grundlagenforschung zur weiteren Entschlüsselung des Genoms, besonders der nicht-kodierenden Bereiche und der dort begründeten regulatorischen Mechanismen, bei. Andererseits ist die Ursache der Entstehung vieler Krankheiten sehr wahrscheinlich in fehlerhafter Genregulation zu suchen (13,26,47). Somit ist das Wissen über koregulatorische Beziehungen von Genen auch für die medizinische Forschung unmittelbar von Relevanz. Erste gentherapeutische Methoden, wie beispielsweise die vektorvermittelte Integration entsprechender genetischer Information in das Genom zum Ersatz von defekten Genen, finden vereinzelt bereits in der Klinik Anwendung (48). Besonders großes Potenzial wird gentherapeutischen Ansätzen zugesprochen, die auf CRISPR-Systemen (Clustered Regulatory Interspaced Short Palindromic Repeats) basieren (49). Ein großer Vorteil der CRISPR-Systeme besteht darin, dass nur die etwa 20 Nukleotide lange Leitsequenz angepasst werden muss, um die universelle Nuklease Cas9 spezifisch an den Zielort im Genom zu lenken. Dort kann dann durch die Nukleaseaktivität ein exakt platzierter Doppelstrangbruch (DSB) induziert werden (**Abbildung 4**) (50,51).

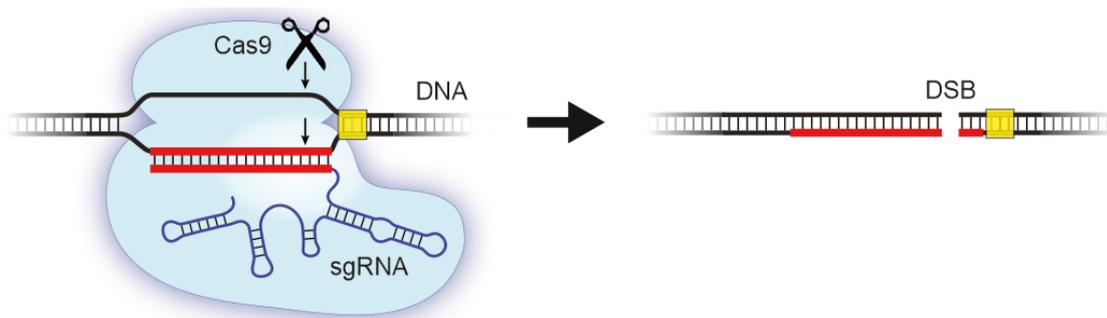


Abbildung 4: CRISPR/Cas9 Funktionsprinzip

Die Cas9 Nuklease bindet aufgrund der individuellen Leitsequenz (sgRNA) an eine spezifische Stelle im Genom, angrenzend an einen Locus mit PAM-Sequenz (protospacer adjacent motif, gelber Kasten). Durch die Nukleaseaktivität des universellen Cas9 Proteins wird ein exakt platzierter Doppelstrangbruch (DSB) der DNA induziert. Modifizierte Abbildung nach Sternberg et al., 2015 (50)

Liegt keine homologe Ersatzsequenz vor, werden bei der folgenden unzureichenden Reparatur durch die zelleigenen Reparaturmechanismen Insertionen oder Deletionen verursacht. Die Folge sind Frameshift-Mutationen oder Stopcodons, womit die Genexpression stark beeinträchtigt oder abgeschaltet wird. Wird ein Gen vollständig abgeschaltet, spricht man von einem Knockout. Alternativ können einzelne Genomabschnitte mit Längen von bis zu 10 kb gezielt entfernt und zielgerichtet durch Integration von beliebigen homologen DNA-Sequenzen in das Genom ersetzt werden (51,52). Auch CRISPR-Systeme mit inaktivierten Nukleasen können zur gezielten Repression oder Aktivierung von Genexpression eingesetzt werden, ohne jedoch bleibende Veränderungen im Genom zu verursachen (**Abbildung 5**) (53).

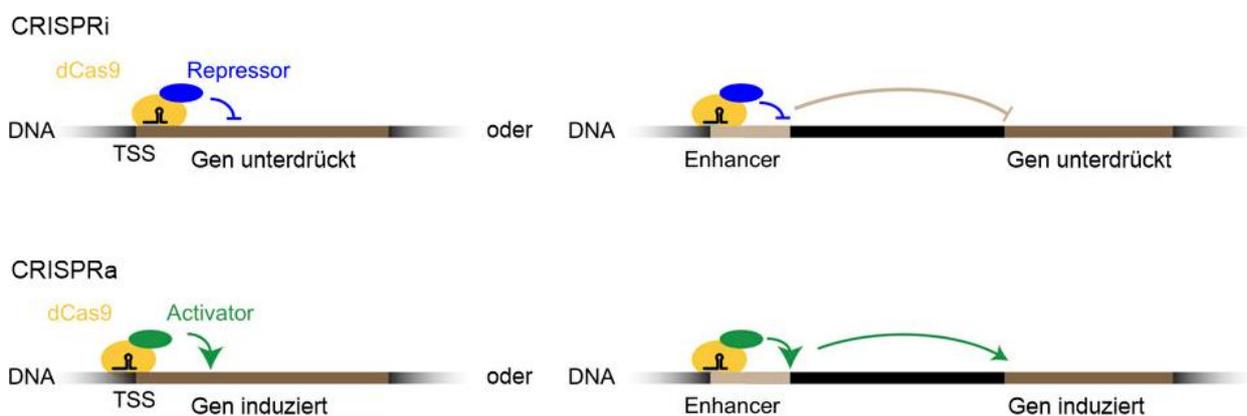


Abbildung 5: Regulation von Genexpression mittels CRISPR-Systemen mit deaktivierter Nuklease

Die Abbildung veranschaulicht, wie durch CRISPR-Systeme mit deaktivierter Nuklease regulatorische Elemente wie Repressoren oder Aktivatoren an spezifische genomische Loci von beispielsweise Promotern oder Enhancern dirigiert werden können, ohne dabei einen DSB zu induzieren. Repressoren oder Aktivatoren interagieren mit der DNA-Sequenz, sodass die Expression von Genen gezielt unterdrückt oder gefördert wird. Modifizierte Abbildung nach Kampmann, 2020 (54)

Bindet die inaktivierte Nuklease beispielsweise direkt am Promoter eines Gens, wird an dieser Stelle die Bindung von Transkriptionsfaktoren sterisch blockiert und die Transkription des Gens verhindert (53,55). Unter dem Begriff CRISPRa werden aktivierende Ansätze zusammengefasst, bei welchen synthetische Transkriptionsfaktoren fest mit der inaktivierten Nuklease verbunden sind (54,55). Diese Konstrukte werden mittels entsprechender Leitsequenz an bestimmte Loci dirigiert, wo die gebundenen Transkriptionsfaktoren beispielsweise mit Promotern und Enhancern interagieren und die Genexpression aktivieren (55). In einer Studie von Matharu et al. wird diese Methode exemplarisch im Mausmodell bereits erfolgreich eingesetzt, um mittels Überexpression der intakten Kopien haploinsuffizienter Gene den Wildtyp wiederherzustellen (56). Auch bei Betrachtung auf epigenetischer Ebene wurden in den letzten Jahren zahlreiche Erkrankungen beschrieben, die durch Mutationen in Chromatin-modifizierenden Enzymen ausgelöst wurden (13). Durch die resultierenden Abweichungen in den Methylierungs- oder Acetylierungsmustern der betroffenen Loci wird die Bindungsaffinität von regulatorischen Elementen verändert und folglich die Genexpression gestört. Genau nach dem gleichen Prinzip kann durch entsprechende CRISPR-basierte Methylierungs- und Acetylierungssysteme die Chromatinstruktur gezielt beeinflusst und die Transkription indirekt gesteuert werden (57).

Bei all diesen CRISPR-basierten Methoden müssen vor der Durchführung erster klinischer Studien vor allem unerwünschte Nebeneffekte, sogenannte Off-Target-Effekte, genau geprüft werden (55,58). In der Regel wird mittels bioinformatischer Methoden versucht, anhand der Leitsequenz weitere potenzielle Bindungsstellen zu erkennen und somit unerwünschte Effekte vorherzusagen (59). Daneben können jedoch auch koregulatorische Beziehungen Off-Target-Effekte verursachen. Solange die zugrundeliegenden Mechanismen nicht abschließend geklärt werden können, kann bei als koreguliert identifizierten Genen nicht ausgeschlossen werden, dass bei Modifikation der Genexpression des einen Gens auch koregulierte Gene mit beeinflusst werden. Deshalb können durch die Identifikation und Charakterisierung regulatorischer Cluster möglicherweise neue wertvolle Hinweise auf potenzielle Off-Target-Effekte gegeben werden.

1.5 Ziel der Arbeit

Eine Untersuchung von koregulierten Genen und regulatorischen Clustern wurde bisher nur für die Netzhaut des Menschen durchgeführt (46). Das Ziel dieser Arbeit besteht in der Identifizierung und Charakterisierung von regulatorischen Clustern in weiteren Geweben. Die Cluster sollen auf ihre Struktur über die verschiedenen Gewebe hinweg untersucht und aus gen-, cluster-, sowie gewebezentrierten Blickwinkeln analysiert werden. Dadurch soll eine umfassende Datenbank generiert werden, die neue Rückschlüsse auf genregulatorische Zusammenhänge ermöglicht. Mithilfe der Analyse in verschiedenen Geweben sollen mögliche gewebsübergreifende, oder aber -spezifische, regulatorische Effekte aufgedeckt werden. Das Wissen um gemeinsam agierende und regulierte Gene soll der Wissenschaftsgemeinschaft strukturiert und leicht zugänglich gemacht werden, denn Assoziationen dieser Art können wertvolle Hinweise auf neue Off-Target-Effekte bei der Evaluation von gentherapeutischen Ansätzen, wie beispielsweise CRISPR-basierter Methoden, liefern.

2 Material und Methoden

Die genregulatorische Clusteranalyse basiert auf Version 8 der Datenbank des GTEx Projekts (40). Die 838 Patientenproben wurden mittels Hauptkomponentenanalyse nach europäischer Abstammung selektiert, sodass letztlich Genotyp- und Genexpressionsdaten von 694 der 838 Patientenproben aus 49 Geweben zur Verwendung kamen. Dabei variierte die Stichprobengröße von 65 (Nierenrinde) bis 584 (Skelettmuskulatur). Filtern nach einer Mindestallelfrequenz des selteneren Allels von 1 % identifizierte 9.158.644 genetische Varianten. Die Hauptkomponentenanalyse sowie die Berechnung der eQTL wurden im Vorfeld von Dr. Tobias Strunz (Institut für Humangenetik, Universität Regensburg) durchgeführt und freundlicherweise zur Verfügung gestellt.

2.1 Berechnung der Expression Quantitative Trait Loci (eQTL)

Die Berechnung der cis-eQTL erfolgte sukzessive für jedes Gewebe entsprechend der GTEx v8 Analyse Pipeline (40) mit Hilfe des Programms *FastQTL* (60). Zur Kartierung der cis-eQTL wurden alle Varianten in einem Fenster von 1 Mb strangauf- und abwärts des Transkriptionsstartpunktes jedes Gens einbezogen. Als Kovariaten gingen im linearen Regressionsmodell das Alter und Geschlecht, die ersten fünf Faktoren der Genotyp-Hauptkomponentenanalyse sowie bis zu 60 PEER-Faktoren ein. Zudem wurden hier auch Parameter der Genotypisierung, wie z.B. die verwendete Plattform oder das Protokoll der Bibliothekenherstellung, berücksichtigt. Zuerst wurde *FastQTL* im Modus *permute 1000 10000* ausgeführt, um die Null-Verteilung der p-Werte der Varianten für jedes betrachtete Gen zu bestimmen. Mittels dieser Verteilung wurde für jedes Gen der nominelle p-Wert für die signifikanteste korrelierte Variante angepasst. Die daraus resultierenden empirischen p-Werte wurden anschließend mittels des Ansatzes der FDR (*false discovery rate*) nach Storey et al. (2003) (61) für multiples Testen adjustiert. Die Bestimmung der entsprechenden q-Werte erfolgte durch die Funktion *qvalue* aus dem R-Paket *qvalue* (61) mit dem Parameter $\lambda = 0,85$. Durch Anwendung einer FDR von $\leq 0,05$ wurden so alle eGene mit mindestens einer signifikanten Variante identifiziert.

Es folgten weitere Schritte, um für jedes zuvor identifizierte eGen alle signifikanten eVarianten zu bestimmen: Der empirische p-Wert, dessen adjustierter q-Wert der FDR von 0,05 am nächsten kam, wurde als genomweiter Signifikanz-Schwellenwert p_t

definiert. Mittels p_t sowie der genspezifischen Parameter der Null-Verteilung der p-Werte wurde individuell für jedes Gen ein nomineller p-Schwellenwert bestimmt. Letztlich wurden zu jedem eGen jeweils all diejenigen Varianten als signifikant korreliert, die einen nominellen p-Wert unterhalb des zuvor berechneten genspezifischen Schwellenwerts erreichten.

2.2 Identifikation von genregulatorischen Clustern

Regulatorische Cluster werden von eGenen gebildet, denen das gleiche genetische Signal zugrunde liegt. Zur Identifikation der genregulatorischen Cluster wurde ein zweistufiges Protokoll angewandt:

In einem ersten Schritt wurden die berechneten eQTL nach eVarianten gefiltert, die mit mehr als einem eGen signifikant korrelieren. Nur solche eGene besitzen möglicherweise gleiche genetische Signale und kamen somit als ein Element eines regulatorischen Clusters in Frage. Anschließend wurden potenziell regulatorische Regionen definiert, um zur Reduktion der benötigten Rechenkapazität und zur Gewährleistung der statistischen Power die gegeneinander zu testenden eGene vorab lokal einzugrenzen. In eine potenziell regulatorische Region wurden iterativ alle eVarianten eines Chromosoms inklusive ihrer eGene eingeschlossen, die einen Höchstabstand von 2 Mb zur vorhergehenden eVariante nicht überschritten. Da bei der Berechnung der eQTL ein Fenster von 1 Mb strangauf und -abwärts des Transkriptionsstartpunktes betrachtet wurde, konnte so sichergestellt werden, dass keine eGene mit gemeinsamen eVarianten als Grundlage für ein potenziell gemeinsames genetisches Signal in verschiedene Regionen eingeteilt wurden.

In einem zweiten Schritt wurden die eGene der zuvor definierten regulatorischen Regionen auf Kolo-kalisation hin getestet. Dies erfolgte mittels der Funktion *coloc.abf* aus dem R-Paket *coloc* (45,62), welche die genetischen Signale jeder möglichen Genpaarung innerhalb einer potenziell regulatorischen Region mittels bayesscher Statistik auf Kolo-kalisation hin testet. Die Funktion verwendete als Input pro Gen die nominellen p-Werte sämtlicher assoziierter Varianten, die Effektgrößen der Assoziationen einschließlich ihrer Varianz, sowie die Standardabweichung der gemessenen Genexpression. Aufgrund der bayesschen Methodik erfordert das *coloc*-Programm zudem die Festlegung dreier *a-priori*-Wahrscheinlichkeiten, nämlich die

Wahrscheinlichkeiten, dass eine der betrachteten Varianten zufällig nur mit Gen 1 (p_1), nur mit Gen 2 (p_2) oder mit beiden Genen assoziiert ist (p_{12}). Für p_1 und p_2 wurden die jeweiligen genspezifischen Schwellenwerte eingesetzt, während für p_{12} zunächst wie von C. Wallace beschrieben der Standardwert von 5×10^{-6} festgelegt wurde (62). Dieser wurde für die Fälle, in denen p_1 oder p_2 bereits kleiner als p_{12} war, auf die entsprechende kleinere Wahrscheinlichkeit korrigiert. Im Rahmen der Kollokalisationsanalyse wurde die *a-posteriori*-Wahrscheinlichkeit dafür berechnet, dass die betrachtete Genpaarung vom gleichen genetischen Signal reguliert wird (H_4). Für den Fall, dass die eGene aufgrund des zuvor festgelegten Fensters von 1 Mb keinerlei Varianten teilen, wurde der Wert von H_4 auf 0 gesetzt. Nur Genpaarungen mit einer *a-posteriori*-Wahrscheinlichkeit für ein gemeinsames genetisches Signal von $\geq 0,8$ wurden in der weiteren Analyse berücksichtigt. Zusätzlich wurde die Variante mit dem niedrigsten p-Wert, die beide Gene gemeinsam haben, zur Bestimmung der Effektrichtung der Regulation herangezogen. Aus der Menge aller koregulierten Genpaarungen wurden nun alle eGene, denen das gleiche genetische Signal zugrunde liegt, zu genregulatorischen Clustern zusammengeführt. Hierzu wurden die Daten unter Verwendung des R-Pakets *igraph* (63) in ungerichtete Graphen transformiert und mittels der Funktion *max_cliques* nach der maximalen Anzahl an größten vollständigen Subgraphen durchsucht, was exakt der Definition von genregulatorischen Clustern entspricht.

2.3 Integration von Datensätzen mit haploinsuffizienten Genen

Nach der Identifikation der regulatorischen Cluster wurde überprüft, inwiefern gewisse Bestandteile der Cluster potenziell Ziele von gentherapeutischen Ansätzen sein können. Hierfür wurden aus zwei verschiedenen Quellen Datensätze mit Genen verwendet, die bei Haploinsuffizienz einen krankheitsassoziierten Phänotyp hervorrufen. Der erste Datensatz mit 660 Einträgen wurde von Matharu et al. publiziert (56) und am 03.12.2020 heruntergeladen. Der zweite Datensatz wurde am 15.01.2021 aus dem Onlineportal von ClinGen bezogen, einer Ressource unter Federführung des National Institutes of Health (64). Es wurden nur Gene aus der höchsten Evidenz-Kategorie „Sufficient Evidence for Haploinsufficiency“ berücksichtigt. Diesen 312 Einträgen wurde anhand des Gensymbols die zugehörige ENSG-Nummer entsprechend der Ensembl Genannotationsdatenbank zugewiesen (65). Anschließend wurden beide Datensätze so kombiniert, dass jedes Gen genau einmal enthalten war,

was zu einer Liste von 713 Haploinsuffizienz-assoziierten Genen führte. Anhand der ENSG-Nummern des kombinierten Datensatzes folgte ein Suchlauf durch die aus **Kapitel 2.2** resultierende Tabelle mit genregulatorischen Clustern. Die entsprechenden Einträge wurden hervorgehoben und in zwei zusätzlichen Spalten dokumentiert, welches mit Haploinsuffizienz assoziierte Gen im jeweiligen Cluster vorgefunden wurde und in welchem Datensatz es ursprünglich enthalten war.

2.4 Programme und Tools

Die Berechnung der Daten sowie die statistische Auswertung erfolgte vollumfänglich mittels der Scriptsprache R in Version 3.3.1 und 4.0.3 (66). Plots der genetischen Signale einschließlich der Struktur des Kopplungsungleichgewichts der betrachteten Varianten wurden mittels der Online-Plattform von LocusZoom.js (44) erstellt. Zur Berechnung von linearen Regressionsmodellen wurde die R-Funktion *lm* verwendet. Diagramme wurden mittels des R-Pakets ggplot2 (67) erzeugt und zum Teil in Corel Draw nachbearbeitet, während schematische Abbildungen vollständig in Corel Draw erstellt wurden.

3 Ergebnisse

3.1 Identifikation von regulatorischen Clustern

Das Ziel der Arbeit bestand in der Identifikation und Charakterisierung regulatorischer Cluster. Hierfür wurde ein mehrstufiger Arbeitsablauf entwickelt (**Abbildung 6**).

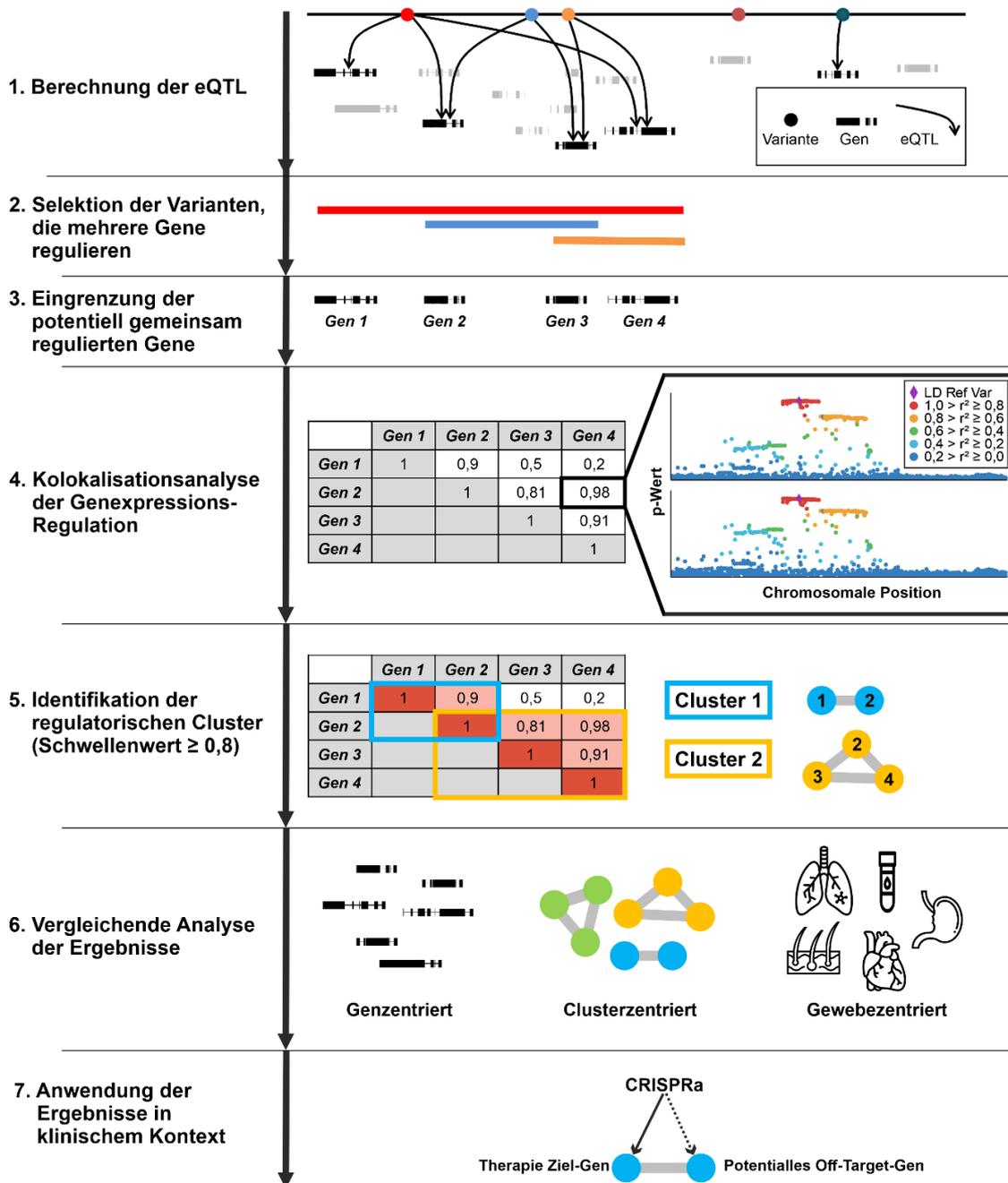


Abbildung 6: Arbeitsablauf zur Identifikation von regulatorischen Clustern anhand von Genexpressionsdaten verschiedener Gewebe.

Nach Berechnung der eQTL wurde lokal nach Varianten gefiltert, die mehr als ein Gen regulieren. Auf diese Weise sollten potenziell koregulierte Gene identifiziert werden (Schritte 1-3). Durch die Analyse der genetischen Signale wurden aus der Menge der Genpaarungen mit einer Kolokalisationswahrscheinlichkeit $\geq 0,8$ regulatorische Gencluster definiert (Schritt 4 und 5). Die Resultate werden aus Gen-, Cluster- und Gewebeperspektive analysiert (Schritt 6) und geben im klinischen Kontext Hinweise auf mögliche Off-Target-Gene bei gentherapeutischen Therapieansätzen (Schritt 7).

In einem ersten Schritt erfolgte die Berechnung der eQTL in 49 Geweben mit Stichprobengrößen von 65 (Nierenrinde) bis 584 (Skelettmuskel). Als Grundlage dafür dienten Genexpressions- und Genotypdaten des GTEx Projekts von 694 Patienten europäischen Ursprungs. Im nächsten Schritt wurden alle eVarianten nach denjenigen eVarianten gefiltert, die mit mehr als einem eGen signifikant korreliert waren. Die von diesen eVarianten regulierten eGene wurden sukzessive zu potenziell regulatorischen Regionen zusammengeführt, um die Anzahl der Tests in der folgenden Analyse möglichst einzugrenzen. Im vierten Schritt wurde eine Kolokalisationsanalyse durchgeführt, bei welcher jede mögliche Genpaarung innerhalb der potenziell regulatorischen Regionen auf ein gemeinsames genetisches Signal hin getestet wurde. Ab einer Kolokalisationswahrscheinlichkeit von $\geq 0,8$ kann davon ausgegangen werden, dass die betrachteten Gene koreguliert werden. Im fünften Schritt wurden aus der Menge aller gemeinsam regulierten Genpaarungen in jedem Gewebe größtmögliche regulatorische Cluster gebildet, die sich dadurch auszeichnen, dass alle beteiligten Gene vom gleichen genetischen Signal reguliert werden. Auf dieser Basis wurden die regulatorischen Cluster im sechsten Schritt der Studie aus verschiedenen Blickwinkeln analysiert, um die Ergebnisse zu charakterisieren und für künftige Experimente möglichst einfach zugänglich zu machen. Abschließend wurde durch Integration zweier Datensätze mit haploinsuffizienten Genen aufgezeigt, welche der zuvor identifizierten regulatorischen Cluster bei der Evaluation von gentherapeutischen Ansätzen von besonderer Bedeutung sein könnten. Die Ergebnisse der Analysen der letzten beiden Arbeitsschritte werden in den folgenden Kapiteln präsentiert.

3.2 Genzentrierte Analyse

3.2.1 Zugrundeliegende Genexpressionsdaten

Aus genzentrierter Perspektive wurde zunächst untersucht, welche Anzahl an Genen des Ausgangsdatsatzes in den einzelnen Geweben exprimiert ist, wie viele dieser Gene genetisch reguliert werden und wie viele eGene sich mit anderen eGenen koreguliert zeigen. Die Resultate wurden in **Abbildung 7** visualisiert, wobei die Gewebe auf der y-Achse nach aufsteigender Stichprobengröße von oben (KDNCTX, $n = 65$) nach unten (MSCLSK, $n = 584$) sortiert sind. Die mittlere Stichprobengröße der 49 Gewebe beträgt 259,4 (SD = 138,9) Proben. Eine vollständige Legende mit

ausgeschriebenen Gewebebezeichnungen zu den in dieser Arbeit verwendeten Gewebeabkürzungen ist **Anhangstabelle A1** zu entnehmen.

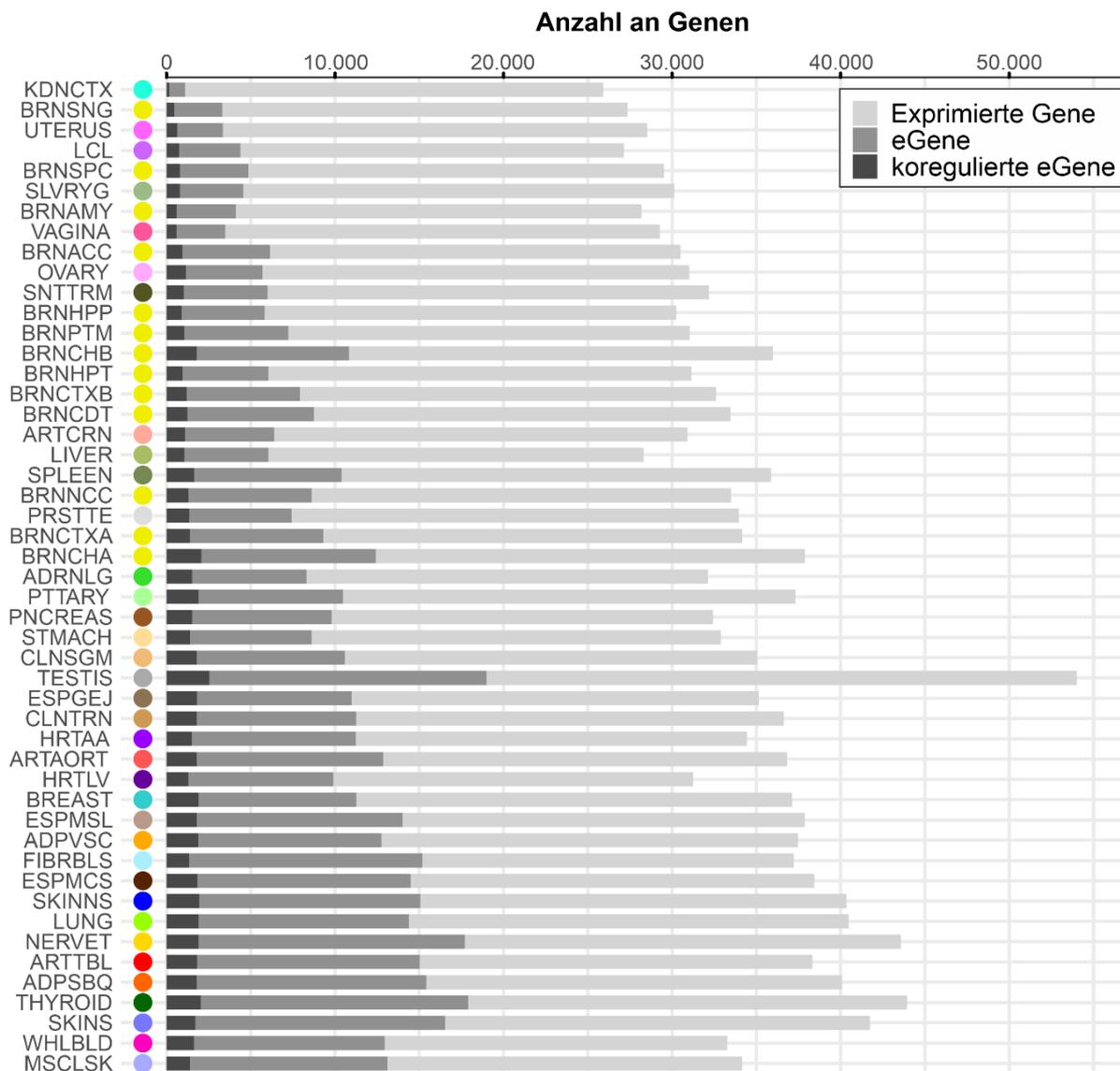


Abbildung 7: Zugrundeliegende Genexpressionsdaten

Dargestellt ist die Anzahl der exprimierten Gene, der eGene sowie der mit mindestens einem weiteren eGen koregulierten eGene pro Gewebe, sortiert nach aufsteigender Stichprobengröße; von $n(\text{KDNCTX}) = 65$ bis $n(\text{MSCLSK}) = 584$. Ausgeschriebene Bezeichnungen zu den Gewebekürzeln sind **Anhangstabelle A1** zu entnehmen.

Insgesamt lagen zu 39.832 Genen in mindestens einem der 49 Gewebe Genexpressionsdaten vor. Ein Anteil von 33.488 Genen (84,1 %) dieser 39.832 exprimierten Gene wurde in mindestens einem Gewebe genetisch reguliert, wovon wiederum 14.636 dieser eGene (43,7 %) mit mindestens einem weiteren eGen koreguliert waren. Betrachtet man die Mittelwerte der genannten Kennzahlen über alle

49 Gewebe hinweg, so waren im Durchschnitt 24.652 Gene (SD = 2.052) in den Geweben exprimiert, während zu durchschnittlich 8.447 (SD = 3.943) dieser Gene mindestens eine signifikant korrelierte Variante identifiziert werden konnte. Im Mittel zeigten sich 1.410 eGene (SD = 500) mit mindestens einem anderen eGen koreguliert. Vergleichsweise große Messwerte wurden in den Gewebeproben des Hodens festgestellt, wobei hier von einer Gesamtanzahl von 35.007 exprimierten Genen 16.443 (47,0 %) eGene waren, von welchen wiederum 2.556 (15,5 %) in regulatorischen Clustern wiedergefunden wurden.

3.2.2 Einfluss der Stichprobengröße auf die Anzahl an eGenen sowie auf die Anzahl koregulierter eGene

In der Gesamtschau des Histogramms in **Abbildung 7** ließ sich der Trend erkennen, dass die Anzahl an eGenen mit steigender Stichprobengröße zunimmt. Aus diesem Grund erfolgte eine Analyse des Einflusses der Stichprobengröße auf die Anzahl identifizierter eGene. Im Vergleich dazu wurde der Einfluss der Stichprobengröße auf die Anzahl an eGenen, die sich mit mindestens einem weiteren eGen koreguliert zeigten, untersucht. Aus **Abbildung 8A** ist ersichtlich, dass die Anzahl an eGenen mit steigender Stichprobengröße linear ansteigt. Ein wichtiger Parameter linearer Regressionsmodelle ist das Bestimmtheitsmaß r^2 . Es kann Werte zwischen einschließlich 0 und 1 annehmen, wobei gilt: Je näher r^2 an 1 liegt, desto stärker ist der lineare Zusammenhang; ist $r^2 = 0$, liegt kein linearer Zusammenhang vor. In diesem Fall zeigte das lineare Regressionsmodell eine starke Korrelation beider Variablen mit einem Bestimmtheitsmaß r^2 von 0,744 (p-Wert = $1,6 \times 10^{-15}$) über den ganzen Wertebereich der Stichprobengröße hinweg. Im Gegensatz dazu stieg die Anzahl an koregulierten eGenen nur bis zu einer Stichprobengröße von etwa 200 linear an (**Abbildung 8B**). Für die lineare Regression dieses Abschnittes betrug das Bestimmtheitsmaß $r^2 = 0,736$ bei einem p-Wert von $4,2 \times 10^{-8}$. In den 24 Geweben mit einer Stichprobengröße über 200 war kein signifikanter Einfluss der zugrundeliegenden Probenanzahl auf die Anzahl an koregulierten eGenen nachweisbar ($r^2 = 0,010$; p-Wert = 0,64).

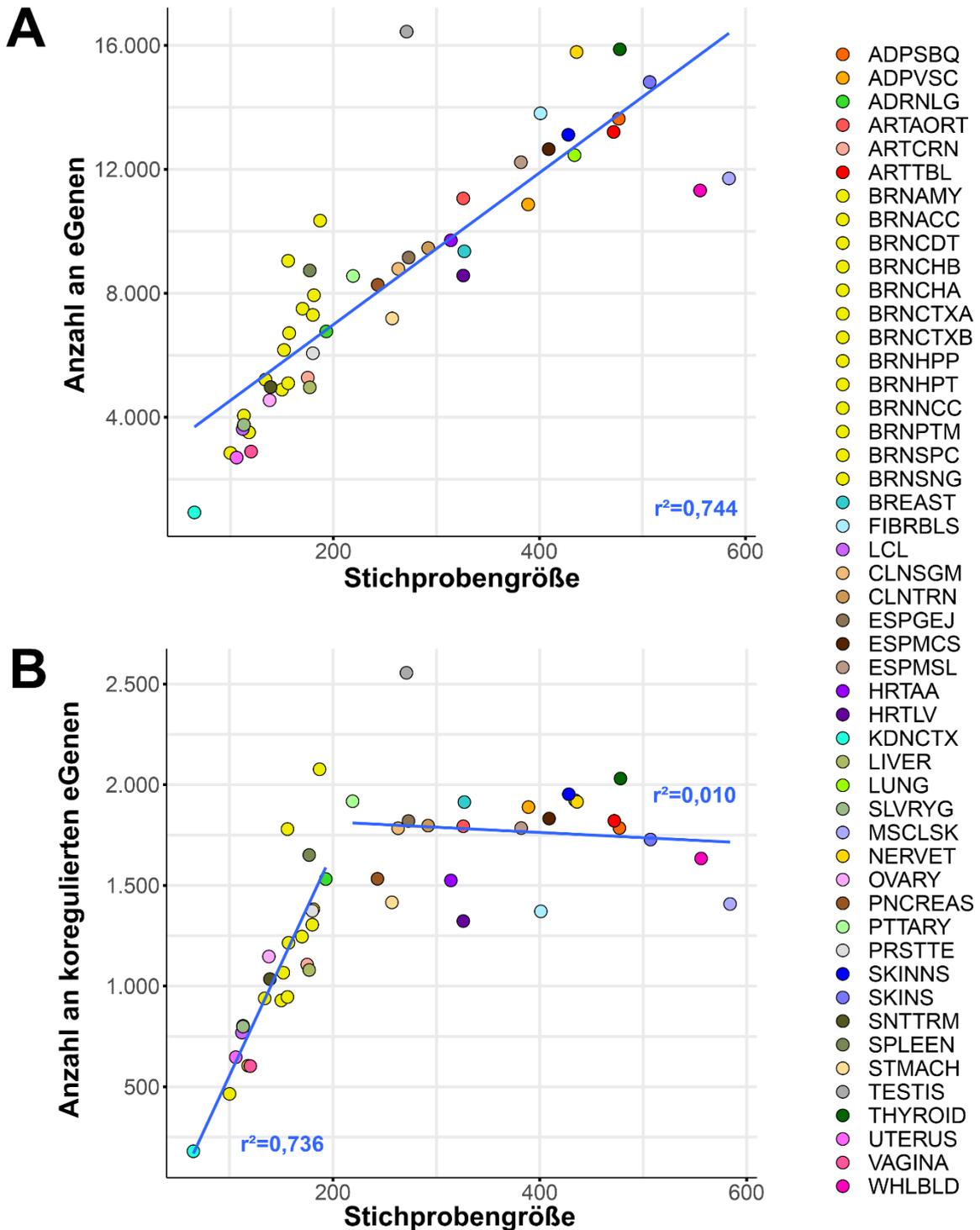


Abbildung 8: Einfluss der Stichprobengrößen der Gewebe auf die Anzahl der eGene sowie auf die Anzahl der eGene in Clustern

(A) Das Streudiagramm zeigt die Zahl an eGenen (y-Achse) in Abhängigkeit der Stichprobengröße der jeweiligen Gewebe (x-Achse). In blau ist die Regressionsgerade zum linearen Modell mit einem Bestimmtheitsmaß von $r^2 = 0,744$ und einem p-Wert = $1,6 \times 10^{-15}$ eingezeichnet. (B) In diesem Streudiagramm ist die Anzahl der eGene, die Teil eines Clusters sind, gegen die Stichprobengröße aufgetragen. In blau sind die Regressionsgeraden nach Aufteilung in zwei Abschnitte eingezeichnet. Unter Einbezug der Gewebe mit Stichprobengröße < 200 betragen die Parameter des Regressionsmodells $r^2 = 0,736$ und p-Wert = $4,2 \times 10^{-9}$, während unter Einbezug der Gewebe mit Stichprobengröße > 200 das Bestimmtheitsmaß $r^2 = 0,010$ und der p-Wert = 0,64 betragen.

Zusammengefasst lässt sich feststellen, dass bis zu einer Stichprobengröße von 200 die Anzahl an koregulierten eGenen ansteigt und deshalb in diesen Geweben aufgrund der kleinen Stichprobengröße ein Teil der regulatorischen Beziehungen wahrscheinlich nicht identifiziert werden kann. Aufgrund dieser Beobachtung wurden im Folgenden immer dann, wenn es Effekte zwischen den verschiedenen Geweben zu vergleichen galt, nur Gewebe mit einer Mindeststichprobengröße von 200 berücksichtigt, um so eine gewisse Vergleichbarkeit zu gewährleisten.

3.2.3 Kolokalisationsanalyse genetischer Signale

Im Rahmen der Kolokalisationsanalysen wurden die genetischen Signale der einzelnen Gene paarweise betrachtet, um koregulierte Gene zu identifizieren. Unter einem genetischen Signal versteht man die Gesamtheit aller Varianten eines Genortes, einschließlich der Varianten mit nicht signifikanten p-Werten, sowie deren exakte Beziehung zueinander. Beim Vergleich von genetischen Signalen zweier Gene konnten alle überlappenden Varianten berücksichtigt werden, die zuvor für beide betrachtete Gene analysiert wurden. Anschließend wurden die Wahrscheinlichkeiten dafür berechnet, dass die genetischen Signale der jeweils betrachteten Genpaarungen kolokalisieren. Bei Werten von $\geq 0,8$ ist davon auszugehen, dass die beiden Gene koreguliert sind. Von den insgesamt 8.643.014 analysierten Genpaarungen zeigten 49.637 (0,57 %) eine Kolokalisationswahrscheinlichkeit von $\geq 0,8$, wobei 5.759.777 (66,6 %) der Tests aufgrund fehlender überlappender Varianten per Definitionem eine Wahrscheinlichkeit von 0 ergaben. Zur Identifikation aller 49.637 koregulierten Genpaarungen wurden im Mittel 6.167,3 überlappende Varianten (SD = 2.191,7) herangezogen. Aufgrund der Tatsache, dass manche Genpaarungen identisch in mehreren Geweben auftreten, wurde letztlich eine Zahl von 16.673 einzigartigen koregulierten Genpaarungen bestimmt. Bei Ausschluss von Geweben mit einer Stichprobengröße < 200 für vergleichende Analysen verblieben 12.470 einzigartige koregulierte Genpaarungen.

Am Beispiel einer potenziell regulatorischen Region auf Chromosom 1 mit 189 Genen zeigen nur 22 aller möglichen Genpaarungen eine Kolokalisationswahrscheinlichkeit $\geq 0,8$, darunter die Gene *FLG* und *HRNR* (**Abbildung 9**). Diesen beiden Genen lag zu 94 % das gleiche genetische Signal zugrunde, wobei die Kolokalisationsanalyse auf 4.629 überlappenden Varianten basierte. Während das Gen *TCHH* zwar genetisch

reguliert ist, unterscheidet sich das zugrundeliegende genetische Signal jedoch deutlich. Die Beziehung der koregulierten Gene *FLG* und *HRNR* wurde in *subkutanem Fettgewebe* sowie in identischer Form in 21 weiteren Geweben identifiziert.

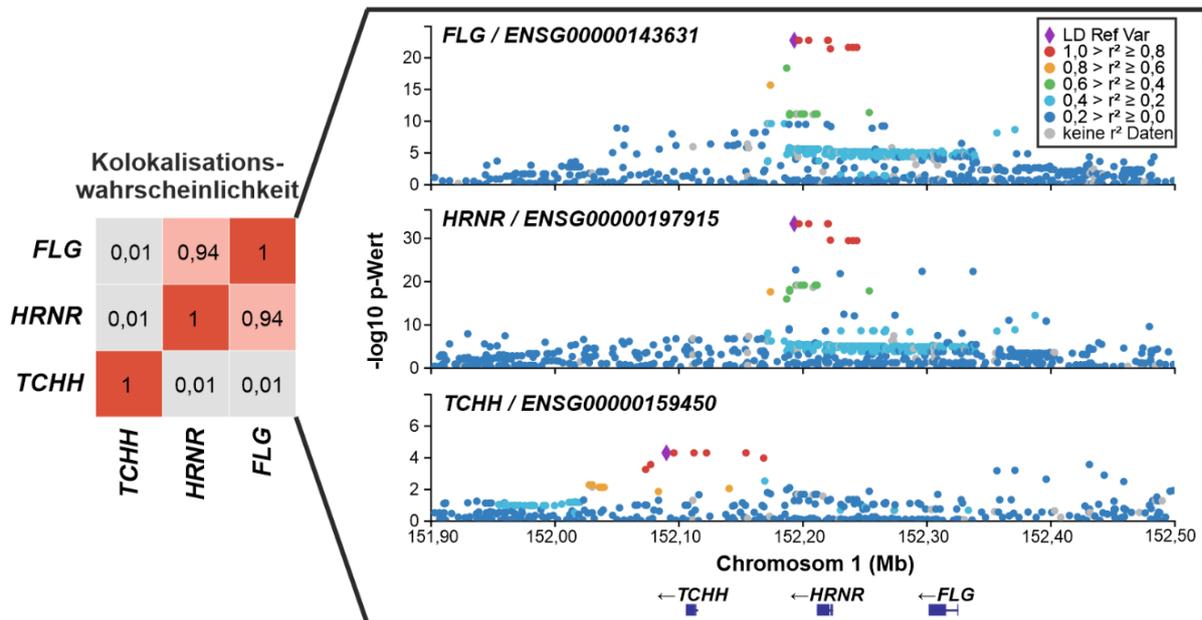


Abbildung 9: Kolokalisiertes genetisches Signal der Gene *FLG* und *HRNR* mit Negativkontrolle

Der linke Teil der Abbildung zeigt einen Ausschnitt einer Kolokalisationsmatrix mit den Genen *FLG*, *HRNR* und *TCHH*. Ab einer Kolokalisationswahrscheinlichkeit von 0,8 ist davon auszugehen, dass die genetischen Signale der betrachteten Genpaarungen übereinstimmen. Im rechten Teil der Abbildung sind die genetischen Signale in LocusZoom-Plots (44) dargestellt. Jeder Punkt steht für eine Variante, während die negativen \log_{10} Werte der p-Werte der eQTL-Berechnung auf der y-Achse gegen die genomische Position auf der x-Achse aufgetragen ist. Die Farbcodierung zeigt, welche Varianten in Kopplungsgleichgewicht zueinanderstehen.

3.2.4 Analyse koregulierter Genpaarungen

Im Anschluss an die Kolokalisationsanalyse wurde die Struktur der daraus resultierenden koregulierten Genpaarungen analysiert. Hierfür wurde zunächst untersucht, wie die gemeinsam regulierten Gene der 12.470 Genpaarungen aus den Geweben mit einer Stichprobengröße > 200 örtlich zueinander positioniert sind (**Abbildung 10**). Als Gengrenzen wurden die Start- und Endpositionen nach Ensembl (65) herangezogen und anschließend drei Fälle unterschieden: Von *vollständig überlappenden* Genen ist die Rede, wenn die Gengrenzen des einen Gens vollständig innerhalb der Gengrenzen des anderen Gens liegen. Diese Konstellation wurde in 1.329 Fällen identifiziert und lag somit in 10,7 % aller Genpaarungen vor. Als *partiell überlappend* werden Gene bezeichnet, deren Gengrenzen sich zumindest zum Teil überschneiden, was in 5,0 % ($n = 625$) der Genpaarungen zutraf. Im dritten Falle von *nicht überlappenden* Genen sind die Gene nebeneinander auf den jeweiligen

Chromosomen positioniert und teilen keinerlei Abschnitte miteinander. In dieser Anordnung war der Großteil der koregulierten Genpaarungen mit 84,3 % ($n = 10.516$) aller Fälle konfiguriert. Auffällig ist, dass der Abstand nicht überlappender Gene im Mittel 112.617,5 Basenpaare ($SD = 268.154,2$ bp) beträgt. Die größte gemessene Distanz zwischen den Gengrenzen beträgt 1.998.702 bp.

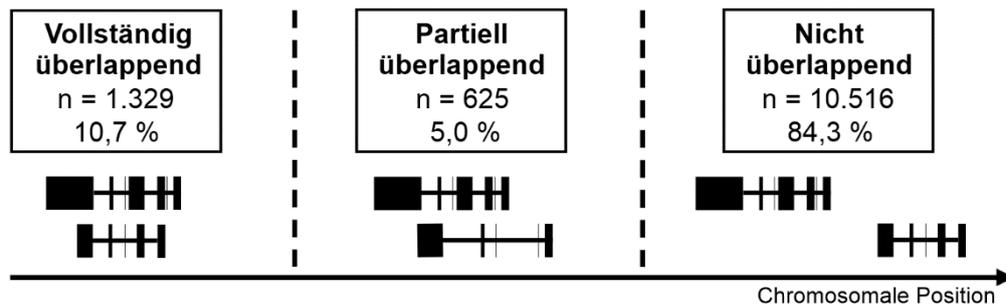


Abbildung 10: Relative Genpositionen in koregulierten Genpaarungen

Betrachtung aller kolokalisierender Genpaarungen ($n = 12.470$) in Geweben mit Stichprobengröße > 200 . Bei nicht überlappenden Genen beträgt die Distanz zwischen den Genen maximal 1.998.702 bp und im Mittel 112.617,5 bp ($SD = 268.154,2$ bp)

Neben den relativen Genpositionen wurden auch die Effektrichtungen der koregulierten Gene betrachtet. Es zeigte sich, dass in 9.153 (73,4 %) der 12.470 Genpaarungen beide Gene in allen Geweben, in welchen die betrachtete Genpaarung identisch erfasst wurde, jeweils in derselben Effektrichtung reguliert werden. Entsprechend ergab sich für 2962 (23,8 %) Genpaarungen die Situation, dass eine Regulation beider betrachteter Gene in allen Geweben konsistent in entgegengesetzter Effektrichtung vorliegt. 355 (2,8 %) der Genpaarungen weisen in mindestens einem Gewebe dieselbe Effektrichtung und zugleich in mindestens einem anderen Gewebe verschiedene Effektrichtungen auf.

3.3 Clusterzentrierte Analyse

Nach Analyse der Struktur von koregulierten Genpaarungen – den Grundbausteinen von Clustern – sollte nun die Gesamtheit der regulatorischen Cluster betrachtet werden. Regulatorische Cluster zeichnen sich dadurch aus, dass jede mögliche Genpaarung innerhalb eines Clusters kolokalisierende genetische Signale aufweist. So wurden aus der Menge der 16.637 koregulierten Genpaarungen insgesamt 14.727 möglichst große einzigartige Cluster identifiziert. Bei Berücksichtigung mehrfach identisch vorkommender Cluster wurden über alle 49 Gewebe hinweg insgesamt

33.261 erfasste Cluster gezählt. Für eine bessere Vergleichbarkeit aus clusterzentrierter Perspektive wurden im Folgenden nur die 10.482 einzigartigen Cluster der Gewebe mit einer Stichprobengröße > 200 herangezogen, wobei pro Gewebe im Mittel 848 (SD = 127) verschiedene Cluster erfasst wurden.

Zuerst wurde untersucht, wie viele genregulatorische Cluster aller vorkommenden Clustergrößen in den Geweben mit einer Stichprobengröße > 200 existieren (**Abbildung 11**). Von insgesamt 10.482 einzigartigen Clustern waren 8.386 (80,0 %) Cluster der Größe 2. Die durchschnittliche Clustergröße betrug 2,3 (SD = 0,6). Im Gewebe Colon Transversum wurden auf Chromosom 9 die zwei größten Cluster mit neun koregulierten Genen identifiziert. Beide Cluster enthielten die Gene *TMEM141*, *CCDC183*, *CCDC183-AS1*, *MAMDC4*, *LCN12*, *ABCA2*, *NPDC1*, *ENTPD2* und unterschieden sich nur im neunten Gen: Das eine Cluster enthielt zusätzlich das Gen *PHPT1*, während das andere stattdessen das Gen *RABL6* beinhaltete.

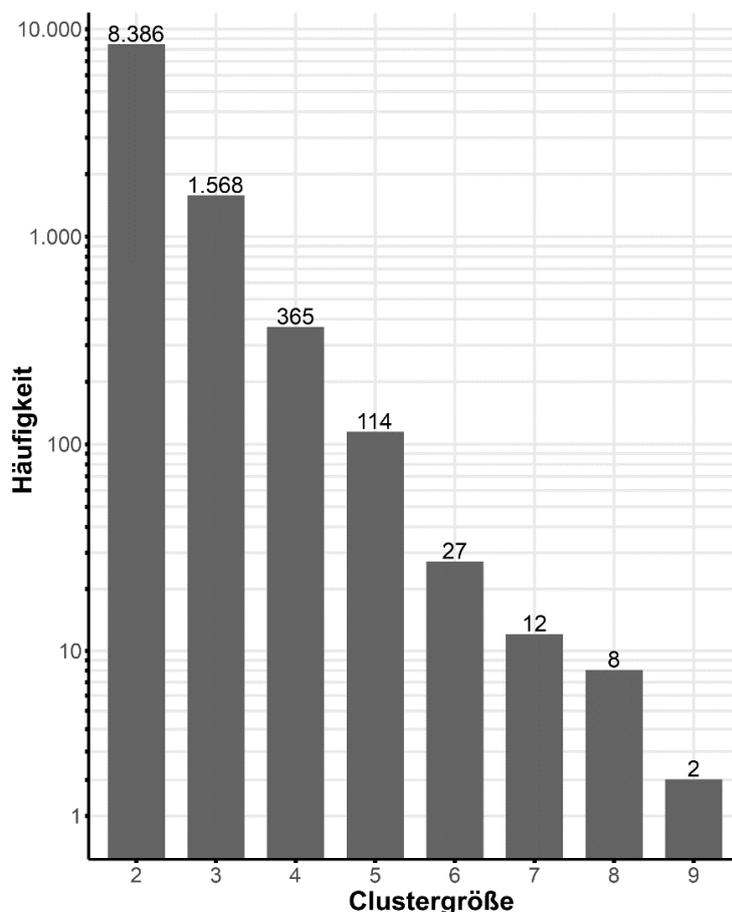


Abbildung 11: Verteilung der Clustergrößen

Das Säulendiagramm zeigt auf der y-Achse auf logarithmischer Skala die absolute Häufigkeit, mit welcher die auf der x-Achse aufgetragenen Clustergrößen in allen Geweben mit einer Stichprobengröße > 200 vorkommen.

Weiter wurden die Verteilungen der Clustergrößen getrennt für jedes Gewebe mit einer Stichprobengröße > 200 betrachtet und untereinander verglichen (**Abbildung 12**). Hierbei zeigten sich die Anteile der Clustergrößen gleichmäßig verteilt und unabhängig von der Stichprobengröße. Das Gewebe *Herz – linker Ventrikel* (HRTLTV) besaß von insgesamt 648 Clustern mit einem Anteil von 88,6 % (574 Cluster) den höchsten Anteil an Zweier-Clustern sowie mit 0,3 % (2 Cluster) den geringsten Anteil an Clustern der Größen 5+. Im Gegensatz dazu wurde in *Blut* (WHLBLD) von insgesamt 765 Clustern mit 2,1 % (16 Cluster) der größte Anteil an großen Clustern identifiziert, während der Anteil der Zweier-Cluster mit 81,2 % (751/925 Cluster) in Gewebeproben der *Hypophyse* (PTTARY) am geringsten ausfiel.

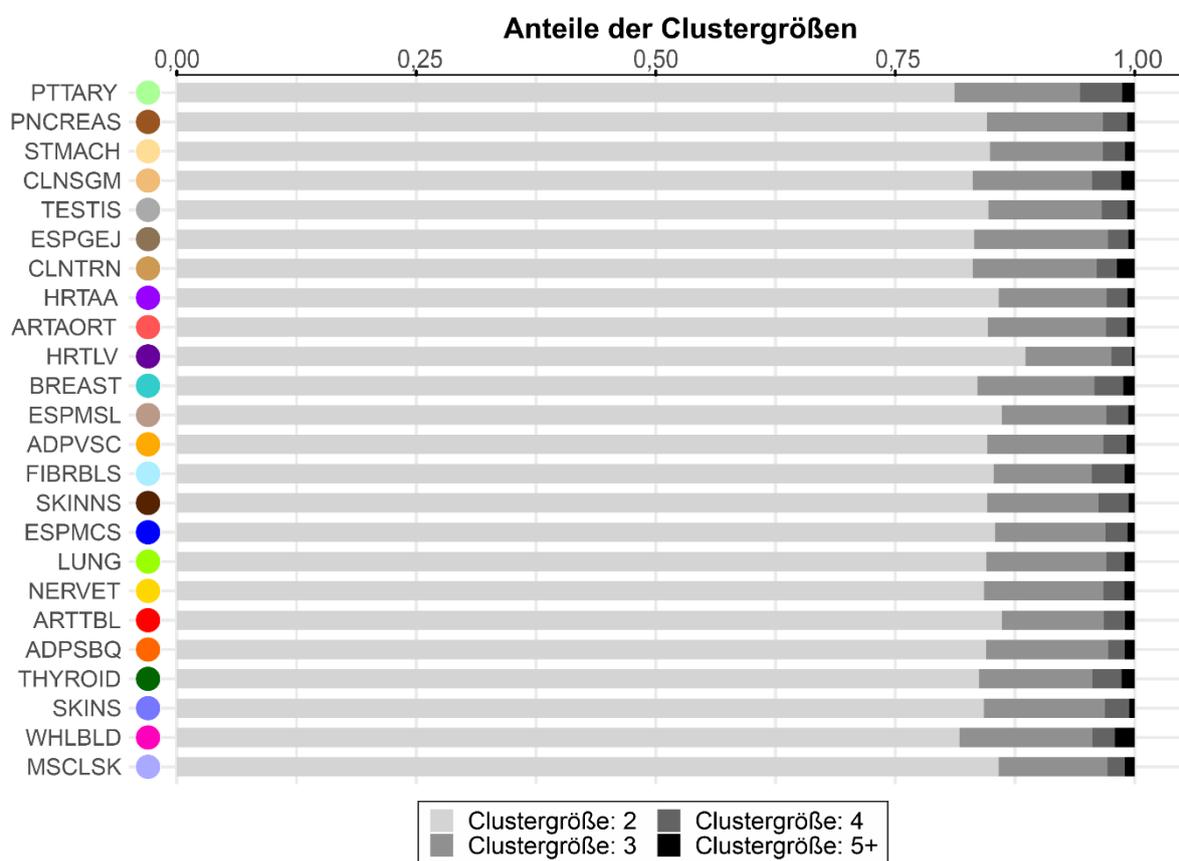


Abbildung 12: Verteilung der Clustergrößen pro Gewebe
 Dargestellt sind die relativen Häufigkeiten der Clustergrößen für jedes Gewebe ab einer Stichprobengröße von 200, aufsteigend nach Stichprobengröße sortiert.

3.4 Gewebezentrierte Analyse

Nachdem zuvor unterschiedliche Gesichtspunkte zu Struktur und Größe der einzelnen Cluster untersucht wurden, folgte nun eine Analyse aus gewebezentrierter Perspektive. Sie sollte aufzeigen, welche Rückschlüsse zu Beziehungen zwischen den einzelnen Geweben auf Grundlage von regulatorischen Clustern gezogen werden können. Zur Beantwortung dieser Fragestellung musste erfasst werden, wie viele Gemeinsamkeiten jede Gewebepaarung hinsichtlich regulatorischer Cluster aufwies. Hierfür wurde zunächst für jedes mögliche Gewebepaar ermittelt, welche Cluster identisch in beiden Geweben identifiziert wurden. Zu diesen Clustern wurde zusätzlich überprüft, ob sie in einem der beiden Gewebe Teil eines größeren Clusters sind, denn häufig waren Cluster über die Gewebe hinweg zwar nicht identisch konfiguriert, teilten jedoch einen gemeinsamen Kern. Die Summe aus identischen Clustern und Clustern mit gemeinsamem Kern wurde als Kennzahl zur Ermittlung von Gemeinsamkeiten zwischen Geweben bestimmt und in einer Heatmap visualisiert (**Abbildung 13**). Es wurden alle Gewebe mit einer Stichprobengröße > 200 eingeschlossen.

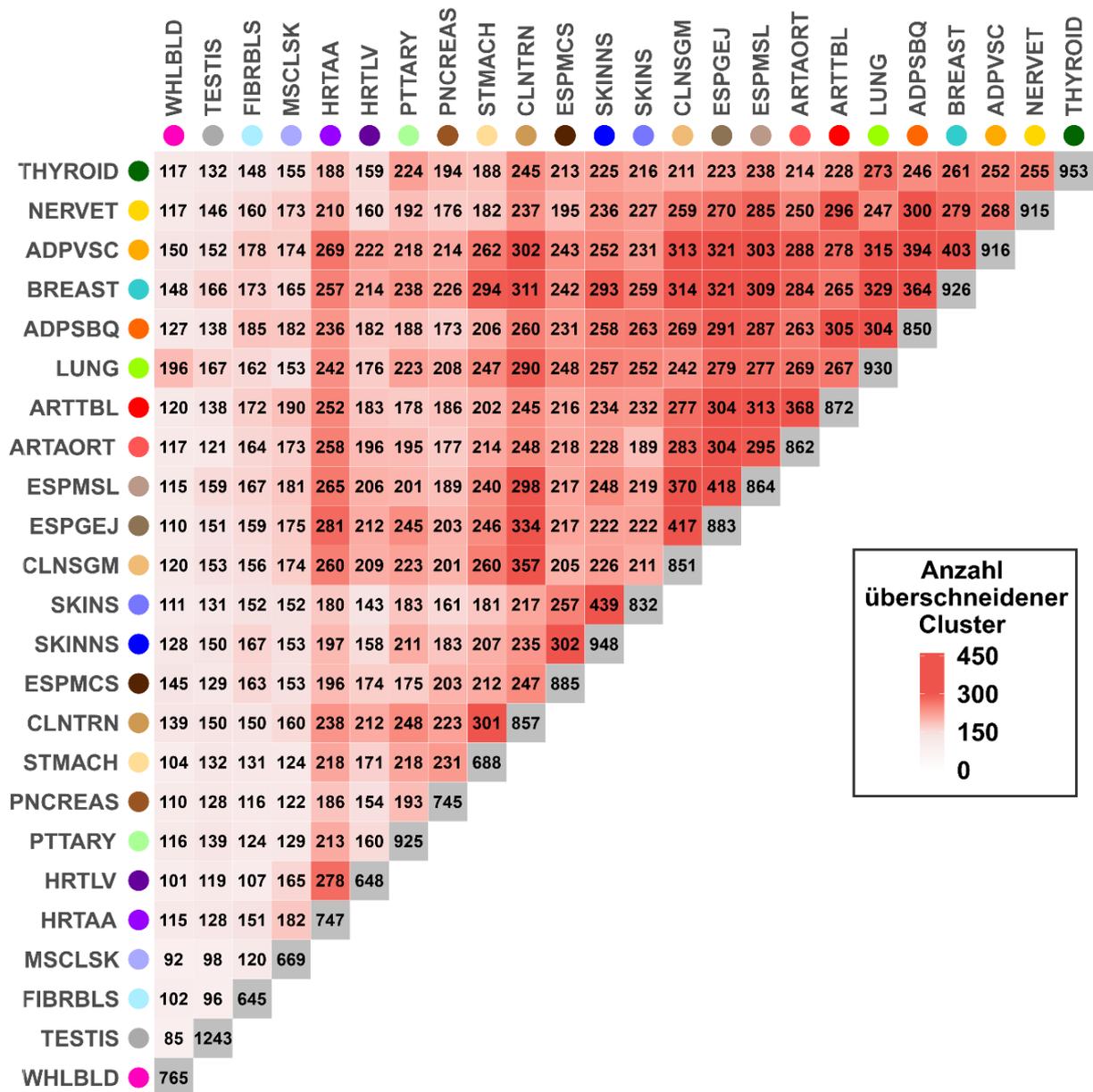


Abbildung 13: Heatmap zu Überschneidungen von Clustern für jede Gewebepaarung

In der Heatmap ist dargestellt, wie viele Cluster des einen Gewebes identisch oder als Teil eines größeren Clusters des anderen Gewebes vorkommen. Berücksichtigt sind alle Gewebe mit einer Stichprobengröße > 200. Die grau hinterlegte Diagonale zeigt die Anzahl der Cluster des jeweiligen Gewebes. Die Anordnung der Gewebe entlang den Achsen erfolgte anhand eines Algorithmus, der möglichst viele Gewebe mit hohen Kennzahlen für überschneidende Cluster möglichst nebeneinander platziert.

Die Gesamtanzahl an Clustern pro Gewebe nahm Werte von 645 in Proben von *Fibroblasten* (FIBRBLS) bis zu 1.243 in Gewebeproben von *Hoden* (TESTIS) an. Im Mittel zwischen allen Geweben ergab sich eine Zahl von 212,4 (SD = 66,2) überschneidender Cluster. Beispielsweise hatten die beiden Gewebeproben aus *nicht-Sonne-ausgesetzter, suprapubischer Haut* (SKINNS) und aus *Sonne-ausgesetzter Haut des unteren Beins* (SKINS) mit 439 Clustern in etwa die Hälfte aller ihrer

regulatorischen Cluster gemeinsam und erreichten damit den höchsten absoluten Wert aller betrachteten Gewebepaarungen. Zudem zeigte sich, dass sich die vier Gewebe *Blut* (WHLBLD), *Hoden* (TESTIS), *Fibroblasten* (FIBRBLS) und *Skelettmuskel* (MSCLSK) mit keinem anderen Gewebe in mindestens 200 Clustern überschritten und damit scheinbar die größten Besonderheiten zu allen anderen Geweben aufwiesen. Insgesamt hatte jedes der betrachteten Gewebe eine Mindestanzahl von 85 regulatorischen Clustern mit einem anderen Gewebe gemeinsam.

Weiter wurde untersucht, inwiefern die in dieser Studie ermittelten genregulatorischen Effekte spezifisch für ein einzelnes Gewebe waren (**Abbildung 14**).

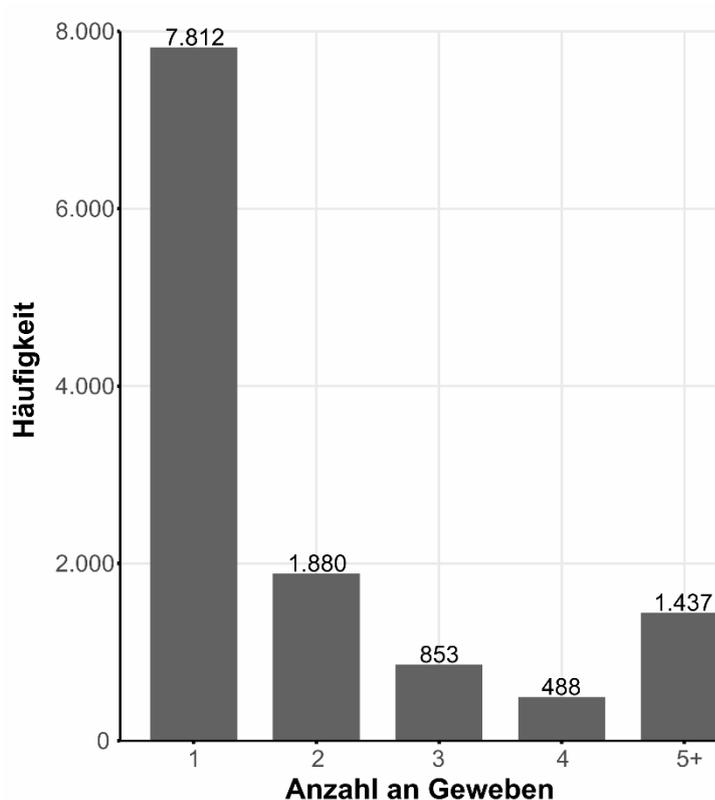


Abbildung 14: Gewebespezifität von koregulierten Genpaarungen

Dargestellt ist auf der y-Achse die Anzahl an Genpaarungen aus Geweben mit Stichprobengröße > 200, eingeteilt in Kategorien auf der x-Achse, die angeben, in wie vielen Geweben eine Genpaarung identisch vorliegt. Die Kategorie 5+ zeigt an, wie viele Genpaarungen in 5 bis 24 Geweben identisch vorkommen.

Hierfür wurden erneut die den Clustern zugrundeliegenden koregulierten Genpaarungen herangezogen, um unabhängig von den tatsächlichen Strukturen der Cluster alle gefundenen Effekte zu berücksichtigen. Bei Betrachtung der 24 Gewebe mit einer Stichprobengröße > 200 erfolgte die Analyse entsprechend auf Basis von 12.470 einmaligen, koregulierten Genpaarungen. Dann wurde bestimmt, in welcher

Anzahl an verschiedenen Geweben identische koregulierte Genpaarungen identifiziert wurden. Genpaarungen, die in 5 oder mehr Geweben identisch vorkamen, wurden in der willkürlich festgelegten Gruppe 5+ zusammengefasst. Es zeigte sich, dass 7.812 (62,7 %) der 12.470 Genpaarungen in nur einem Gewebe identifiziert wurden und damit als gewebespezifisch betrachtet werden können. 4.658 (37,3 %) aller Genpaarungen wurden hingegen in mehr als einem Gewebe gefunden, darunter 1.437 (11,5 %) die der Gruppe 5+ zugeordnet wurden, womit hier von gewebeübergreifenden Effekten ausgegangen werden kann.

3.5 Klinischer Bezug von regulatorischen Clustern

Die Resultate der bisherigen Analysen zeigen, dass regulatorische Cluster ein sehr häufiges Phänomen sind. Daraus ergibt sich, dass die Ergebnisse auch in aktuellen klinischen Ansätzen berücksichtigt werden sollten. Weil die koregulierten Gene anhand gemeinsamer genetischer Signale bestimmt wurden, ist das Wissen dieser Studie über regulatorische Cluster potenziell für sämtliche Methoden und Mechanismen relevant, welche die Expression eines dieser koregulierten Gene auf Ebene der Transkription regulieren oder verändern. Ein Beispiel für einen gentherapeutischen Ansatz auf Transkriptionsebene sind CRISPRa-Systeme. Durch sie können Aktivatoren gezielt an spezifische Positionen auf dem Chromosom geführt werden, wo sie mit anderen regulatorischen Elementen interagieren und so letztlich die Expression eines Gens zielgerichtet modifizieren. Wie ein solches CRISPRa-System eingesetzt werden kann, um Ausfälle von haploinsuffizienten Genen zu kompensieren, ist in **Abbildung 15** schematisch dargestellt.

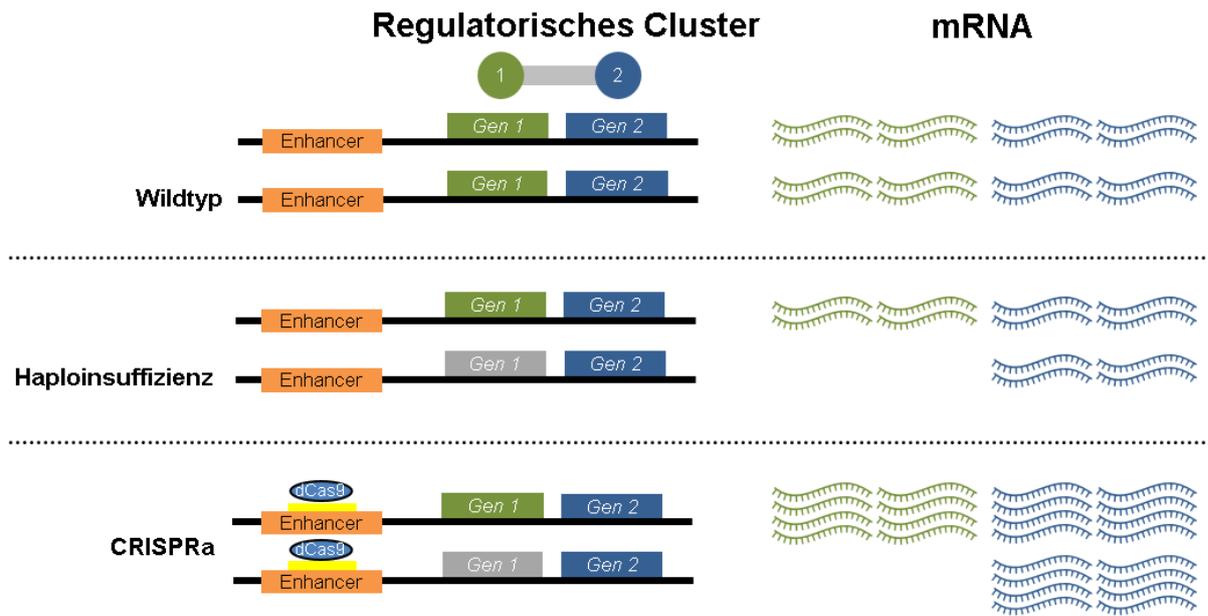


Abbildung 15: Regulatorische Cluster geben Hinweise auf potenzielle Off-Target-Effekte bei Behandlung von Haploinsuffizienzen mittels CRISPRa

Gen 1 und Gen 2 befinden sich in einem regulatorischen Cluster. Der obere Abschnitt zeigt, wie im Normalfall zwei gesunde Genkopien zur regulären Menge mRNA und damit zur Expression des Wildtyps führen. In der mittleren Zeile ist dargestellt, dass ein Ausfall des haploinsuffizienten Gens 1 weniger Genprodukt und somit eine Krankheit verursacht. Der untere Abschnitt zeigt, wie gezielte Überexpression der intakten Kopie von Gen 1 mittels CRISPRa-System zur wildtypischen Menge an mRNA von Gen 1 führt, jedoch gleichzeitig beispielsweise zu einer überschießenden Expression des koregulierten Gens 2 führen kann.

Wenn das Zielgen zudem Teil eines regulatorischen Clusters ist, kann auch die Expression weiterer Gene beeinflusst werden. Dieser Einfluss ist in der Regel ungewollt und die Folgen schwer abzuschätzen. Aus diesem Grund müssen die koregulierten Gene in den Clustern der Zielgene als potenzielle Off-Target-Gene betrachtet und bei der Evaluation von gentherapeutischen Ansätzen dieser Art berücksichtigt werden.

Dementsprechend wurde im Anschluss untersucht, wie viele regulatorische Cluster der hier präsentierten Arbeit ein haploinsuffizientes Gen beinhalten. Dafür wurden zwei verschiedene Datensätze von Matharu et al. (56) sowie des ClinicalGenome Projekts (64) herangezogen und zu einer Liste mit insgesamt 713 haploinsuffizienten Genen kombiniert. Anschließend wurde die Datenbank mit regulatorischen Clustern anhand des kombinierten Datensatzes durchsucht. Es wurden 429 Cluster gefunden, die mindestens ein potenziell haploinsuffizientes Gen enthielten. Diese Cluster beinhalteten neben den haploinsuffizienten Genen insgesamt 414 weitere koregulierte Gene, die somit als potenzielle Off-Target-Gene in Frage kommen. Die überwiegende

Anzahl (324 Cluster; 75,5 %) der Cluster bestand aus zwei Genen, wobei auch ein Cluster der Größe 7 ein haploinsuffizientes Gen enthielt. 295 (68,8 %) der 429 gefundenen Cluster waren spezifisch für das jeweilige Gewebe, wohingegen die übrigen 134 (32,2 %) Cluster in mehreren Geweben, in einem Fall sogar in 43 der 49 Gewebe, identisch erfasst wurden. Weiter wurde festgestellt, dass 10 der gefundenen Cluster zwei verschiedene haploinsuffiziente Gene beinhalten und da es sich bei allen dieser 10 Cluster um Zweier-Cluster handelt, bestehen diese damit vollständig aus haploinsuffizienten Genen. Von den 713 haploinsuffizienten Genen der Suchliste wurden 231 (32,4 %) in einem oder mehreren Clustern aufgefunden (**Abbildung 16**). Zusammengefasst können aufgrund der Ergebnisse dieser Arbeit allein im Kontext von Haploinsuffizienzen bereits für 231 mögliche Zielgene bestimmter genterapeutischer Ansätze neue Hinweise auf 414 potenzielle Off-Target-Effekte gegeben werden.

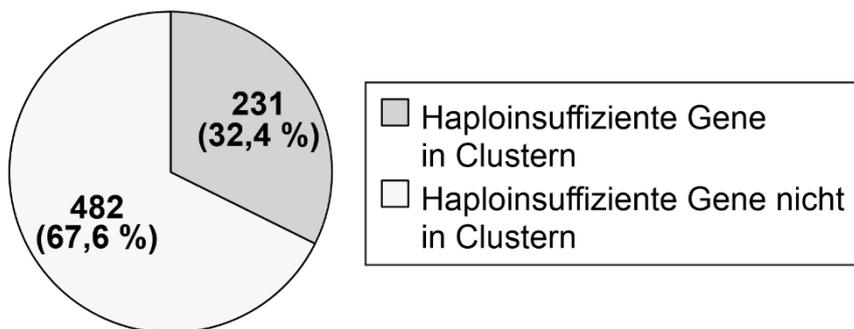


Abbildung 16: Anteil der in Clustern gefundenen haploinsuffizienten Gene
Das Tortendiagramm veranschaulicht, wie viele der 713 haploinsuffizienten Gene der Suchliste in mindestens einem regulatorischen Cluster gefunden wurden.

4 Diskussion

In dieser Studie erfolgte die Identifikation koregulierter Gene und regulatorischer Cluster anhand genetischer Signale basierend auf Genexpressionsdaten des GTEx Projekts. Von 39.832 exprimierten Genen wurden über alle 49 Gewebe hinweg 14.636 (36,7 %) Gene identifiziert, die sich mit mindestens einem anderen Gen koreguliert zeigen. Insgesamt wurden durch Kolokalisationsanalysen 49.637 koregulierte Genpaarungen identifiziert, welche anschließend zu 14.727 einzigartigen regulatorischen Clustern zusammengeführt wurden. Regulatorische Cluster erreichten Größen von bis zu neun verschiedenen Genen, die alle vom identischen genetischen Signal reguliert werden. Im Mittel betrug der Abstand zwischen koregulierten Genen 112.617,5 bp (SD = 268.154,2 bp). In der gewebezentrierten Analyse stellte sich heraus, dass die regulatorischen Effekte eines jeden Gewebes in mindestens 80 regulatorischen Clustern mit anderen Geweben übereinstimmen. Schließlich wurde am Beispiel haploinsuffizienter Gene gezeigt, warum bei der Modifikation von Genexpression im Rahmen von therapeutischen Ansätzen auch koregulierte Gene als potenzielle Off-Targets eine Rolle spielen können. Aufgrund der großen Anzahl betroffener regulatorischer Cluster ist es ratsam, koregulatorische Beziehungen bei der Entwicklung von gentherapeutischen Methoden, die die Genexpression auf der Ebene der Transkription modifizieren, zu berücksichtigen.

Seit das nicht-kodierende Genom zunehmend im Fokus der Forschung steht und die Bedeutsamkeit der Regulation von Genexpression berücksichtigt wird, wird an der Entschlüsselung des Gesamtbilds genregulatorischer Beziehungen gearbeitet. Bei der Identifikation von regulatorischen Elementen spielt die dreidimensionale Chromatinstruktur eine wichtige Rolle. Mittels Methoden wie 3C oder Hi-C werden TADs als übergeordnete regulatorische Einheiten beschrieben, die Genombereiche mit erhöhter Dichte an regulatorischen Beziehungen eingrenzen. TADs erreichen je nach Auflösung Größenordnungen von einigen Kilobasen bis wenigen Megabasen (21). Die Bildung von DNA-Schleifen innerhalb von TADs bietet eine Erklärung, wie regulatorische Elemente über große Distanzen hinweg miteinander interagieren können (22,23). Dieses Modell passt gut zur hier präsentierten Studie, welche zeigt, dass koregulierte Gene durchschnittlich 112.617,5 bp (SD = 268.154,2 bp) voneinander entfernt lokalisiert sind. Somit liegt nahe, dass mittels der

Kolokalisationsanalyse genetischer Signale tatsächlich regulatorische Beziehungen identifiziert werden können, die durch die Bildung dreidimensionaler Chromatinstrukturen bedingt sind. Inwieweit koregulierte Gene tatsächlich innerhalb zuvor identifizierter TAD-Grenzen liegen, bleibt jedoch zu validieren. Voraussetzung für einen aussagekräftigen Abgleich wären entsprechende Hi-C Experimente auf Basis vergleichbarer Gewebeproben, die aktuell nicht in diesem Maßstab vorliegen.

Trotzdem ist man mittels Hi-C in der Lage, Enhancer-Promoter Beziehungen zu erfassen und einzelne Loci mit potenziell regulatorischer Rolle aufgrund ihrer strukturellen Eigenschaften zu identifizieren. Ein Nachteil hierbei ist, dass die tatsächliche funktionelle regulatorische Aktivität nicht unmittelbar gemessen wird (68). Zudem ist die Auflösung zur Erfassung der räumlichen Interaktionen stark abhängig von der Sequenzierungstiefe der Hi-C Experimente (31,68). Die relativ geringen Sequenzierungstiefen aktuell verfügbarer Hi-C Datensätze, welche für vergleichende gewebebasierte Analysen nötig sind, erlauben zum Erlangen belastbarer Aussagen lediglich Auflösungen von etwa 40 kb (69). Deshalb kann in Regionen mit großer Gen-Dichte das tatsächlich betroffene Gen oftmals nicht eindeutig identifiziert werden. In diesen Punkten bieten eQTL-basierte Methoden eine Alternative, weil hier die koregulierten Gene anhand tatsächlich gemessener Genexpression festgemacht werden (35,36). Der größte Vorteil liegt darin, dass die koregulatorischen Beziehungen den Genen exakt zugeordnet werden können. Somit können anhand der in dieser Studie erstellten Datenbank koregulatorische Beziehungen einzelner Gene gezielt abgefragt und spezifische Fragestellungen differenziert nach Gewebetypen beantwortet werden. Trotzdem bleibt der auslösende Mechanismus für die Koregulation verborgen. Andererseits bestehen mittlerweile die technischen Voraussetzungen, um Hi-C Untersuchungen mit Auflösungen von wenigen Basen durchzuführen (70). Somit könnten die beteiligten Elemente auch in regulatorisch komplexen Regionen genau detektiert werden. Die Anwendung extrem hoher Auflösungen geht bei genomweiten Ansätzen wie Hi-C jedoch zu Lasten der erreichbaren Sequenzierungstiefe und macht die Studien anfällig für Störsignale (70). Zudem steigt bei Hi-C Experimenten aufgrund des genomweiten, massiv parallelen Untersuchungscharakters die zu überprüfende Anzahl möglicher Interaktionen mit zunehmender Genomgröße exponentiell an (22,70). In der Folge entstehen riesige und

zunehmend komplexe Datensätze, was die bioinformatische Prozessierung und Interpretation zur Herausforderung macht (70).

Aufgrund der Resultate dieser Studie kann jedoch der Fokus weiterer Experimente auf bestimmte regulatorische Cluster gelegt werden. Somit werden die zu untersuchenden Regionen vorab eingegrenzt, welche dann in weiterführenden Experimenten gezielt mittels Hi-C unter hochauflösenden Bedingungen charakterisiert werden könnten. Zur Identifikation von koregulierten Genen reicht die alleinige Betrachtung der Ergebnisse von eQTL-Studien nicht aus. Aufgrund komplexer LD-Strukturen kann die für die beobachteten Effekte kausale Variante nicht sicher bestimmt werden (41,42). So können beispielsweise auch irrelevante Varianten aufgrund von LD-Strukturen hochsignifikante p-Werte erreichen und gleichzeitig lässt die Höhe der p-Werte keinerlei Rückschlüsse auf die ursächliche Variante zu. Dieses Problem kann durch Betrachtung aller untersuchter Varianten und deren relativen Beziehungen zueinander adressiert werden. Beim Vergleich von Mustern genetischer Signale im Rahmen von Kolokalisationsanalysen wird die Gesamtheit der LD-Struktur miterfasst und ermöglicht so die Bestimmung koordiniert regulierter Gene, ohne eine einzelne ursächliche Variante festlegen zu müssen und damit zu stark zu gewichten. Somit bleibt die tatsächlich kausale Variante jedoch in jedem Fall verborgen. Weiter ist bekannt, dass in eQTL-Studien die Anzahl an signifikant assoziierten Varianten stark abhängig von der zugrundeliegenden Stichprobengröße ist (40). Diese Beobachtung konnte im Rahmen dieser Studie einmal mehr bestätigt werden.

Zudem zeigte sich, dass auch die Anzahl an koregulierten Genen bis zu einem gewissen Wert zunächst eine starke lineare Korrelation mit der Stichprobengröße zeigt, bis schließlich kein Zusammenhang mehr festgestellt werden kann. Eine mögliche Erklärung hierfür ist, dass mit steigender Stichprobengröße und damit differenzierteren eQTL-Ergebnissen zunehmend genetische Subsignale erfasst werden. In dieser Arbeit erfolgt die Identifikation koregulierter Gene jedoch lediglich anhand der Primärsignale, die bei ausreichender Stichprobengröße vermutlich entweder bereits vollständig abgebildet oder durch Subsignale sogar zunehmend gestört werden. Kolokalisationsanalysen, die Subsignale berücksichtigen, sind technisch möglich (62), wurden aber aufgrund der enormen benötigten Rechenkapazitäten für die Vielzahl an zu testenden Kombinationen von Subsignalen

pro Genpaarung im genomweiten Ansatz dieser Studie nicht durchgeführt. Eine zwischenzeitlich publizierte Arbeit integriert in die verwendete Software COLOC zur Kollokalisationsanalyse alternative Ansätze zur Regressionsberechnung, basierend auf der Summe von Einzeleffekten, mithilfe welcher auch unter Annahme multipler ursächlicher Varianten mehrere genetische Signale simultan evaluiert werden können und deshalb für mögliche weiterführende Studien zu beachten ist (71). Letztlich lässt sich festhalten, dass besonders in Geweben mit kleiner Stichprobengröße wahrscheinlich ein Teil der vorliegenden koregulierten Gene nicht erfasst werden konnte. Im Umkehrschluss sind dafür vermutlich alle in dieser Studie identifizierten regulatorischen Cluster mit einer hohen Wahrscheinlichkeit valide und bilden eine solide Grundlage für Hypothesen weiterführender Experimente.

Wie bei vielen Methoden wird auch bei dem im Rahmen der hier präsentierten Studie verfolgten Ansatz die Aussagekraft der Resultate stark dadurch bestimmt, welche Daten als Grundlage der Analyse dienen. Es wurde gezeigt, dass auch Faktoren wie das Alter und äußere Einflüsse wie beispielsweise Stressoren die Regulation von Genen sowie die Konfiguration von TADs verändern und somit die Genexpression beeinflussen können (26,72,73). Entsprechend muss bei der Interpretation dieser Studie berücksichtigt werden, dass die Proben des GTEx Projekts den Spendern verschiedenen Alters posthum aus gesundem Gewebe entnommen wurden und somit genau den Gewebezustand unter den zuvor beschriebenen Bedingungen abbilden (40). Zudem wurden im Rahmen anderer Studien bei der Untersuchung von regulatorischen Prozessen neben gewebespezifischen Effekten (38,39) auch zellspezifische Effekte (21,25,43) nachgewiesen. Während erstere aufgrund der gewebebasierten Anlage der hier präsentierten Studie hervorragend erfasst werden dürften, können letztere im Rahmen dieser Arbeit nicht explizit adressiert werden, weil die eQTL-Analyse nicht für einzelne Zelltypen, sondern basierend auf Gewebeproben durchgeführt wurde. Die Erstellung eines zellbasierten eQTL Atlas, organisiert im Rahmen eines Konsortiums, ist in Arbeit (74) und könnte als Grundlage einer erneuten Analyse regulatorischer Gencluster dienen. Selbst bereits vorliegende zellspezifische Studien, beispielsweise in T-Gedächtniszellreihen, zeigen dynamische eQTL-Ergebnisse in Abhängigkeit des vorliegenden Zell-Status und damit die Relevanz äußerer Einflussfaktoren (75). Andererseits wurde im Rahmen der gewebezentrierten Analyse im Rahmen der hier vorgestellten Untersuchung gezeigt, dass zwischen allen

untersuchten Geweben Schnittmengen bei mindestens 80 regulatorischen Clustern bestehen. Diese gewebeübergreifenden Effekte können als vielversprechender Ansatz weiterführender Untersuchungen dienen.

Neben methodischen Überlegungen stellt sich die Frage, inwieweit die Koregulation von Genen auf bereits validierte regulatorische Mechanismen zurückzuführen ist und ob die generierte Datenbank Rückschlüsse auf ursächliche regulatorische Elemente zulässt. So zeigen sich bei der Analyse der Resultate der hier dargelegten Studie Überschneidungen mit bereits überprüften und anerkannten Effekten publizierter Arbeiten. Die Ergebnisse der hier präsentierten Arbeit zeigen, dass die koregulierten Gene von 9.153 (73,4 %) der 12.470 Genpaarungen in derselben Effektrichtung reguliert werden. Diese Beobachtung passt beispielsweise zu Forschungsergebnissen, die zeigen, dass die Transkription zweier divergent angeordneter Gene von einem zentralen bidirektionalen Promoter aus strangauf- oder strangabwärts initiiert werden kann, wobei bidirektionale Promoter in der Regel eine Präferenz für eine Ableserichtung aufweisen (76,77). Trotz der relativ geringen Anzahl genau charakterisierter bidirektionaler Promoter geht man davon aus, dass die Transkription von etwa 10 % aller Gene von bidirektionalen Promotern bestimmt wird (78,79). Zum Beispiel sind pathologische Mutationen im Gen *FANCA* mit der seltenen Erbkrankheit Fanconi-Anämie (80) sowie mit Eierstock- und Brustkrebs assoziiert (81). Weitergehende Untersuchungen zeigten, dass *FANCA* einen bidirektionalen Promoter besitzt, welcher neben *FANCA* auch das Gen *SPIRE2* steuert. Daraus ergeben sich weitere Ansätze zur Erforschung der Beziehung von Tumorsuppressorgenen (82). Entsprechend wurde im Rahmen dieser Studie ein regulatorisches Cluster erfasst, welches die beiden Protein-kodierenden Gene *FANCA* und *SPIRE2* enthält. Es wurde in 23 der 49 untersuchten Gewebe identifiziert und beide Gene weisen gleichgesetzte Effektrichtungen auf. Divergente Genpaarungen werden meistens koexprimiert, weisen in manchen Fällen jedoch auch entgegengesetzte Effektrichtungen auf (76,77). An dieser Stelle sei beispielsweise das regulatorische Cluster mit den Genen *ATE1* und *ATE1-AS1* angeführt. Bei dieser Genpaarung wurden in diversen Geweben entgegengesetzte Effektrichtungen festgestellt, sodass eine Herabregulation der Expression des Antisense-Transkripts *ATE1-AS1* gleichzeitig mit einer erhöhten Transkription von *ATE1* assoziiert ist. Entsprechend dieser Beobachtung wird in der Literatur der bidirektionale Charakter des Promoters von *ATE1* beschrieben (79).

ATE1 kodiert für eine Arginyltransferase, die unter anderem als Hauptregulator viele krebsrelevante Stoffwechselwege beeinflusst, wobei verminderte *ATE1* Expression auch als Indikator für Metastasierung von Prostatakrebs diskutiert wird (83). Auch bei genauerer Analyse eines der im Rahmen dieser Studie größten identifizierten regulatorischen Clusters wurden bereits gut untersuchte regulatorische Komplexe von *Homöobox* (HOX) -Genen wiedergefunden. Die koordinierte Regulation einer Untergruppe der HOX-Gene, das HOXD-Cluster, spielt eine wichtige Rolle in der Entwicklung der proximalen und distalen Gliedmaßen (14,84). Entsprechend konnte auf Chromosom 2 in viszeralem Fettgewebe (Omentum) ein regulatorisches Cluster erfasst werden, welches mit den Genen *HOXD10*, *HOXD-AS2*, *HOXD9*, *HOXD8*, *HOXD3*, *HOXD4*, *RP11-387A1.5* und *HAGLR* im Prinzip ausschließlich Gene aus der HOXD-Familie enthält. *HAGLR* und das nicht weiter charakterisierte Gen *RP11-387A1.5* kodieren für Transkripte des Antisense-Strangs des HOXD-Clusters (65). Unklar ist, warum das regulatorische Cluster in dieser Ausprägung gerade in viszeralem Fettgewebe (Omentum) erfasst wurde. Eine mögliche These ist, dass bereits die geringe Expression eines der Gene zur Aktivierung des ganzen Clusters ausreicht. Weiter zeigen andere Studien, dass die Transkription der HOXD-Gene in zwei unabhängigen Phasen erfolgt. Diese werden von Enhancer-Systemen reguliert, welche jeweils an den Rändern des Genclusters lokalisiert sind (85,86). Interessanterweise befindet sich das HOXD-Cluster zudem am Rande zweier TADs. Deshalb wird diskutiert, dass das HOXD-Cluster als dynamische TAD-Grenze fungiert und die genaue Grenzposition der TAD je nach Status der Transkription und dem entsprechenden Entwicklungsfortschritt variiert (87). Im Kontext der hier präsentierten Studie bestätigen diese Beispiele einerseits die Validität der regulatorischen Cluster und zeigen andererseits eindrucksvoll, wie das Wissen über koregulatorische Beziehungen den Blick auf bestimmte Forschungsbereiche erweitern kann. Zudem wird klar, dass die Durchführung weiterführender Studien notwendig ist, um resultierende Thesen zu prüfen und entsprechende Fragestellungen zu klären.

Weiterhin besteht auch besonderes Interesse daran, die Ergebnisse dieser Studie in einen klinischen Kontext zu transferieren. So bietet das Wissen um regulatorische Cluster neue Möglichkeiten zur Evaluation sämtlicher gentechnischer Methoden, welche Genexpression auf Ebene der Transkription modifizieren. Stand 2020 befinden sich gemäß einer Übersichtsarbeit von Uddin et al. 21 CRISPR-basierte Gentherapien

mit 12 verschiedenen Zielgenen in diversen klinischen Studienphasen (59). Eines dieser Gene, *PDCD1*, ist auch in regulatorischen Clustern der hier präsentierten Studie zu finden. Das Genprodukt PD-1 des *PDCD1* Gens fungiert als Immuncheckpoint-Inhibitor von cytotoxischen T-Zellen (CTCs). Als inhibitorischer Transmembran-Immunrezeptor ist PD-1 für eine negative Regulation der CTC-Aktivierung zuständig (88,89). Maligne Zellen können durch Expression von Liganden zum PD-1 Rezeptor die Immunantwort unterdrücken und somit einem Angriff der CTCs entgehen (90). Deshalb ist eine Strategie der gentherapeutischen Ansätze, das Gen *PDCD1* zur Therapie verschiedenster Tumorentitäten zu deaktivieren und die Expression von PD-1 in CTCs zu verhindern. Die daraus folgende fehlende negative Regulierung der CTCs soll so zu einem erhöhten Aktivierungsniveau und folglich verstärkten Angriff auf maligne Zellen führen (59,88,89). Neben den nicht weiter beschriebenen Transkripten *AC114730.2* und *AC114730.3* sowie der lncRNA *LINC01237* zeigt sich das Zielgen *PDCD1* in unterschiedlich konfigurierten regulatorischen Clustern zudem mit den Protein-kodierenden Genen *NEU4*, *RTP5* und *GAL3ST2* koreguliert. *NEU4* kodiert für eine von vier Neuraminidase-Formen und verminderte *NEU4* mRNA Spiegel werden im Zusammenhang mit Dickdarmkrebs diskutiert (91), genau wie eine Herunterregulierung des Gens *GAL3ST2*, welches für eine Galactose-3-O-Sulfotransferase kodiert (92). *RTP5* kodiert für ein nicht genauer untersuchtes Rezeptor-Transporter-Protein (20). Die Literatur gibt jedoch keine zwingenden Hinweise auf Zusammenhänge zwischen *PDCD1* und eines der koregulierten Protein-kodierenden Gene. Da in der Regel eine direkte Veränderung der kodierenden DNA-Sequenz zum Knockout von *PDCD1* angestrebt wird (59,89) und kein übergeordnetes regulatorisches Element Ziel des therapeutischen Ansatzes ist, ist die Wahrscheinlichkeit vermutlich eher gering, dass auch die koregulierten Gene beeinträchtigt werden. Trotzdem ist nicht sicher auszuschließen, dass beispielsweise PD-1 selbst oder die zu modifizierende kodierende Zielsequenz zusätzlich regulatorische Eigenschaften besitzt und damit den Einfluss auf koregulierte Gene bedingen. Solange die regulatorischen Verhältnisse dieser Gene nicht abschließend geklärt sind, sollten sie als potenzielle Off-Target-Effekte nicht außer Acht gelassen und sorgfältig überprüft werden.

Besonders wenn übergeordnete regulatorische Elemente modifiziert werden sollen, spielt das Wissen über koregulierte Gene eine wichtige Rolle, was im Rahmen dieser

Studie am Beispiel haploinsuffizienter Gene als mögliche Zielgruppe solcher gentherapeutischen Ansätze gezeigt wurde. Beispielsweise wurden Funktionsverlust-Mutationen im Filaggrin Gen (*FLG*) als Auslöser von Ichthyosis vulgaris, der häufigsten erblichen Verhornungsstörung, identifiziert. Zudem stellen Ausfälle von *FLG* einen großen genetischen Risikofaktor für atopische Ekzeme dar und zeigen starke Assoziationen zu weiteren Phänotypen aus dem atopischen Formenkreis (93). Als Krankheit mit unvollständiger Penetranz und variabler Expressivität schwankt der Schweregrad der Erkrankung innerhalb der von Ichthyosis vulgaris betroffenen Familien zum Teil stark (94), wobei aufgrund des autosomal semidominanten Charakters Individuen mit zwei *FLG*-Null-Mutationen schwerer betroffen sind als Personen mit nur einer *FLG*-Null-Mutation (93). Als mögliche Therapieansätze kommen folglich unter anderem gentherapeutische Methoden in Frage, die die Expression der intakten Kopie von *FLG* hochregulieren. Beispielsweise könnten mittels eingangs erwähnter CRISPRa-Systeme Transkriptionsfaktoren gezielt zur Interaktion mit einem Enhancer des Gens *FLG* gebracht werden. Durch die resultierende Überexpression des haploinsuffizienten Gens könnte die Schwere der Krankheit gemildert werden. Indessen resultiert aus der genregulatorischen Analyse der hier präsentierten Arbeit, dass *FLG* mit dem Gen Hornerin (*HRNR*) koreguliert ist. Dementsprechend könnte Manipulation der Expression von *FLG* gleichzeitig eine veränderte Expression von *HRNR* verursachen. *HRNR* seinerseits ist ebenfalls mit atopischer Dermatitis assoziiert (95), wird bei erhöhter Expression jedoch auch mit Tumorprogression in Verbindung gebracht (96,97). Bei den identifizierten koregulierten Genen und regulatorischen Clustern resultierend aus dieser Arbeit handelt es sich wie auch in diesem Fall zwar bisher lediglich um Hypothesen, jedoch können dadurch Hinweise auf regulatorische Beziehungen gegeben werden, die bei der Evaluation von gentherapeutischen Ansätzen dringend beachtet werden sollten. Deshalb wäre dringend zu empfehlen das regulatorische Cluster mit den Genen *FLG* und *HRNR* und deren koregulatorische Beziehung gezielt zu untersuchen, bevor die Expression eines dieser Gene mittels gentherapeutischer Ansätze modifiziert wird. In bestimmten Fällen mag es sogar von Vorteil sein, die Expression mehrerer koregulierter Gene in ein und demselben Ansatz kombiniert zu modifizieren, um in therapeutischen Ansätzen den Ausfall mehrerer Gene ganzer regulatorischer Cluster einheitlich zu kompensieren. Auch in solchen Situationen kann die Datenbank dieser

Studie mit der Kartierung genregulatorischer Beziehungen als gute Ausgangslage zur Durchführung von Machbarkeitsanalysen und Folgeexperimenten dienen.

Generell kommen als Therapieziel genregulatorischer Therapieansätze beispielsweise haploinsuffiziente Gene wie *FLG* in Frage. Deshalb wurde untersucht, wie viele regulatorische Cluster in der hier präsentierten Arbeit mindestens ein haploinsuffizientes Gen enthalten. Es zeigte sich, dass von 713 haploinsuffizienten Genen einer Suchliste eine Zahl von 231 (32,4 %) Genen in einem oder mehreren von insgesamt 429 betroffenen regulatorischen Clustern aufgefunden wurden. Dies zeigt, dass aufgrund der Ergebnisse dieser Arbeit allein im Kontext von Haploinsuffizienz bereits eine Vielzahl neuer Hypothesen mit Hinweisen auf potenzielle Off-Target-Effekte generiert wurde. Daraus abgeleitet dürfte auch für sämtliche anderen Gengruppen als Therapieziele von gentherapeutischen Ansätzen eine gewisse Relevanz gegeben sein.

Allgemein betrachtet besitzen die relativ kostengünstigen und robusten CRISPR-basierten Systeme mit vielfältigen Ansätzen zur Modifikation von Genexpressionsregulation großes Potenzial (53,59). Der Einsatz all dieser Methoden steht und fällt jedoch mit der Präzision der CRISPR-Systeme. Bisher sind relativ häufige Off-Target-Effekte der Grund für die größten Bedenken für den klinischen Einsatz (58). Deshalb müssen entsprechende gentherapeutische Ansätze vorab genau evaluiert werden. In der Regel wird mittels bioinformatischer Methoden versucht, anhand der Leitsequenz weitere potenzielle Bindungsstellen zu erkennen und somit Off-Target-Effekte vorherzusagen (59). Gerade wenn unter Einsatz von CRISPR-basierten Methoden die Regulation von Genen modifiziert werden soll, ist das Wissen über koregulierte Gene von besonderem Interesse. Denn auch wenn das CRISPR-System präzise an der geplanten Stelle bindet, könnten zusätzlich zum eigentlichen Zielgen die entsprechend koregulierten Gene mit beeinflusst werden. So bietet die Datenbank der hier präsentierten Studie neue Hinweise auf potenzielle Off-Target-Effekte, die mittels herkömmlicher Methoden nicht beachtet worden wären.

Zusammengefasst lässt sich festhalten, dass im Rahmen dieser Arbeit, welche zwischenzeitlich in Teilen publiziert wurde (98), eine umfangreiche Datenbank mit Informationen über koregulierte Gene und regulatorische Cluster in 49 Geweben

erarbeitet wurde. Die Datenbank ist als Supplementary Material der Publikation abrufbar unter <https://www.mdpi.com/article/10.3390/cells10092395/s1>. Durch Analyse unter verschiedenen Perspektiven wurden die Resultate zur Beantwortung spezifischer Fragestellungen aufbereitet. Zudem ermöglicht die Datenbank die Generierung einer Fülle vielseitiger Hypothesen. Diese können als Grundlage zur Durchführung weiterführender Experimente dienen, besonders zur weiteren Entschlüsselung koregulatorischer Mechanismen sowie im klinischen Kontext zur Evaluation moderner gentherapeutischer Ansätze hinsichtlich wertvoller Hinweise auf potenzielle unerwünschte Nebeneffekte.

Abkürzungsverzeichnis

Abkürzung	Erklärung
3C	Chromosome Conformation Capture
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
CTC	Cytotoxische T-Zellen
eQTL	Expression Quantitative Trait Locus
FLG	Filaggrin
GTEx	Genotype Tissue-Expression
GWAS	Genomweite Assoziationsstudie
HOX	Homöobox
HRNR	Hornerin
LD	Kopplungsungleichgewicht / Linkage Disequilibrium
ncRNAs	Noncoding RNAs
PAM	Protospacer adjacent motif
TAD	Topologisch assoziierender Domäne
WES	Whole exome sequencing

Abbildungsverzeichnis

Abbildung 1: Funktionelle Anordnung regulatorischer Elemente des nicht-kodierenden Genoms im dreidimensionalen Raum	8
Abbildung 2: Cis-eQTL Schema	10
Abbildung 3: Beispiel eines genetischen Signals eines Gens im schematischen LocusZoom-Plot	11
Abbildung 4: CRISPR/Cas9 Funktionsprinzip	13
Abbildung 5: Regulation von Genexpression mittels CRISPR-Systemen mit deaktivierter Nuklease	13
Abbildung 6: Arbeitsablauf zur Identifikation von regulatorischen Clustern anhand von Genexpressionsdaten verschiedener Gewebe	20
Abbildung 7: Zugrundeliegende Genexpressionsdaten	22
Abbildung 8: Einfluss der Stichprobengrößen der Gewebe auf die Anzahl der eGene sowie auf die Anzahl der eGene in Clustern	24
Abbildung 9: Kolokalisiertes genetisches Signal der Gene FLG und HRNR mit Negativkontrolle	26
Abbildung 10: Relative Genpositionen in koregulierten Genpaarungen	27
Abbildung 11: Verteilung der Clustergrößen	28
Abbildung 12: Verteilung der Clustergrößen pro Gewebe	29
Abbildung 13: Heatmap zu Überschneidungen von Clustern für jede Gewebepaarung	31
Abbildung 14: Gewebespezifität von koregulierten Genpaarungen	32
Abbildung 15: Regulatorische Cluster geben Hinweise auf potenzielle Off-Target-Effekte bei Behandlung von Haploinsuffizienzen mittels CRISPRa	34
Abbildung 16: Anteil der in Clustern gefundenen haploinsuffizienten Gene	35

Anhang

Anhangstabelle A1

Abkürzung	Gewebebezeichnung	Stichproben- größe	Anzahl an eQTLs	Anzahl an eGenen	Anzahl an regulatorischen Clustern	Anzahl an eGenen in Clustern	Anzahl an Zweier- Clustern	Anzahl an Dreier- Clustern	Anzahl an Vierer- Clustern	Anzahl an Fünfer-Clustern oder größer	Größe des größten Clusters
ADPSBQ	Subkutanes Fettgewebe	477	2628341	13631	850	1867	718	108	15	9	7
ADPVSC	Viszerales Fettgewebe (Omentum)	389	1810577	10867	916	2022	775	111	22	8	8
ADRNLG	Nebenniere	193	812494	6770	724	1617	595	102	21	6	8
ARTAORT	Aorta	326	1845556	11063	862	1895	730	106	19	7	7
ARTCRN	Koronararterie	175	610660	5279	523	1185	421	78	16	8	7
ARTTBL	Tibialarterie	472	2648963	13206	872	1907	751	93	19	9	8
BREAST	Brustgewebe	327	1438060	9355	926	2057	774	113	28	11	7
BRNACC	Gehirn - Anteriorer cingulärer Cortex	134	497491	5213	457	1025	365	75	16	1	6
BRNAMY	Gehirn - Amygdala	118	324724	3514	301	670	247	42	10	2	5
BRNCDT	Gehirn - Caudatum	170	824725	7503	612	1354	505	88	16	3	6
BRNCHA	Gehirn - Kleinhirn	187	1405447	10346	992	2222	812	139	30	11	8
BRNCHB	Gehirn - Kleinhirnhemisphäre	156	1088502	9050	839	1896	674	125	30	10	8
BRNCTXA	Gehirn - Kortex	181	968888	7939	691	1516	586	83	17	5	6
BRNCTXB	Gehirn - Frontallappen BA9	157	720559	6715	598	1304	513	67	14	4	6
BRNHPP	Gehirn - Hippocampus	150	498323	4892	468	1051	377	70	19	2	6
BRNHPT	Gehirn - Hypothalamus	156	532236	5100	454	1003	378	64	8	4	7
BRNNCC	Gehirn - Nucleus accumbens	180	833789	7302	649	1423	548	80	18	3	5
BRNPTM	Gehirn - Putamen	152	638141	6170	516	1139	433	68	9	6	8
BRNSNG	Gehirn - Substantia nigra	100	244389	2846	222	504	181	30	3	8	5
BRNSPC	Gehirn - Rückenmark C1	113	382467	4055	383	860	306	65	8	4	6
CLNSGM	Colon sigmoideum	263	1267696	8789	851	1903	707	106	26	12	7
CLNTRN	Colon transversum	292	1394820	9459	857	1926	712	111	18	16	9
ESPG EJ	Ösophagus - Gastroösophagealer Übergang	273	1388401	9154	883	1949	735	123	19	6	7

Abkürzung	Gewebebezeichnung	Stichproben- größe	Anzahl an eQTLs	Anzahl an eGenen	Anzahl an regulatorischen Clustern	Anzahl an eGenen in Clustern	Anzahl an Zweier- Clustern	Anzahl an Dreier- Clustern	Anzahl an Vierer- Clustern	Anzahl an Fünfer-Clustern oder größer	Größe des größten Clusters
ESPMCS	Ösophagus - Mucosa	409	2302281	12651	885	1937	756	102	20	7	7
ESPMSL	Ösophagus - Muscularis	382	2223614	12229	864	1884	744	94	20	6	7
FIBRBLs	Fibroblasten	401	2621062	13807	645	1425	550	66	22	7	7
HRTAA	Herzohr	314	1469599	9710	747	1634	641	84	16	6	7
HRTLv	Herz - linker Ventrikel	326	1275371	8573	648	1392	574	58	14	2	7
KDNCTX	Nierenrinde EBV-transformierte	65	91381	929	79	198	57	13	5	4	8
LCL	Lymphozyten	112	360006	3622	356	819	279	56	15	6	7
LIVER	Leber	177	558529	4966	518	1145	440	59	13	6	9
LUNG	Lunge	434	2180823	12457	930	2046	786	116	18	10	6
MSCLSK	Skelettmuskel	584	2345057	11706	669	1461	574	76	12	7	7
NERVET	Nervus tibialis	436	3164521	15786	915	2016	771	114	20	10	7
OVARY	Ovar	138	476496	4549	553	1265	436	85	24	8	7
PNCREAS	Pankreas	243	1158245	8275	745	1645	630	90	19	6	8
PRSTTE	Prostata	180	707135	6067	657	1499	520	107	19	11	7
PTTARY	Hypophyse	219	1181213	8559	925	2093	751	121	41	12	7
SKINNS	Nicht-Sonne-ausgesetzte, suprapubische Haut	428	2408343	13115	948	2089	802	110	30	6	7
SKINS	Sonne-ausgesetzte Haut des Unterschenkels	507	2949425	14817	832	1830	701	105	21	5	7
SLVRYG	Speicheldrüse	113	342555	3757	379	851	314	44	17	4	7
SNTRM	Dünndarm - Terminales Ileum	139	511754	4975	494	1129	394	74	17	9	8
SPLEEN	Milz	177	1038239	8736	793	1761	657	109	16	11	6
STMACH	Magen	257	982360	7190	688	1516	584	81	16	7	8
TESTIS	Hoden	271	2574843	16443	1243	2733	1053	147	33	10	7
THYROID	Schilddrüse	478	3356546	15873	953	2124	798	113	29	13	7
UTERUS	Gebärmutter	106	269189	2700	303	708	233	52	12	6	7
VAGINA	Vagina	120	289918	2893	281	656	218	43	14	6	7
WHLBLD	Blut	556	2234431	11319	765	1726	625	106	18	16	7

Literaturverzeichnis

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860–921.
2. Little PFR. Structure and function of the human genome. Vol. 15, *Genome Research*. Cold Spring Harbor Laboratory Press; 2005. p. 1759–65.
3. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012 Sep 6;489(7414):57–74.
4. Craig Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* (80-). 2001 Feb 16;291(5507):1304–51.
5. Petersen BS, Fredrich B, Hoepfner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. Vol. 18, *BMC Genetics*. BioMed Central Ltd.; 2017.
6. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? Vol. 48, *Journal of Medical Genetics*. BMJ Publishing Group Ltd; 2011. p. 580–9.
7. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Vol. 97, *American Journal of Human Genetics*. Cell Press; 2015. p. 199–215.
8. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. *PLoS Comput Biol*. 2012 Dec;8(12):e1002822.
9. Höglund J, Rafati N, Rask-Andersen M, Enroth S, Karlsson T, Ek WE, et al. Improved power and precision with whole genome sequencing data in genome-wide association studies of inflammatory biomarkers. *Sci Rep*. 2019 Dec 1;9(1):1–14.
10. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009 Jun 9;106(23):9362–7.
11. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al.

- Molecular biology: The transcriptional landscape of the mammalian genome. *Science* (80-). 2005 Sep 2;309(5740):1559–63.
12. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature*. 2012 Sep 6;489(7414):101–8.
 13. Smith M, Flodman PL. Expanded Insights Into Mechanisms of Gene Expression and Disease Related Disruptions. *Front Mol Biosci*. 2018 Nov 27;5(NOV):101.
 14. Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, et al. A regulatory archipelago controls hox genes transcription in digits. *Cell*. 2011 Nov 23;147(5):1132–45.
 15. Lanctôt C, Cheutin T, Cremer M, Cavalli G, Cremer T. Dynamic genome architecture in the nuclear space: Regulation of gene expression in three dimensions. Vol. 8, *Nature Reviews Genetics*. *Nat Rev Genet*; 2007. p. 104–15.
 16. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, et al. Bidirectional promoters generate pervasive transcription in yeast. *Nature*. 2009 Feb 19;457(7232):1033–7.
 17. Zhang LF, Ding JH, Yang BZ, He GC, Roe C. Characterization of the bidirectional promoter region between the human genes encoding VLCAD and PSD-95. *Genomics*. 2003;82(6):660–8.
 18. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Vol. 167, *Cell*. *Cell Press*; 2016. p. 1170–87.
 19. Soler-Oliva ME, Guerrero-Martínez JA, Bachetti V, Reyes JC. Analysis of the relationship between coexpression domains and chromatin 3D organization. *PLoS Comput Biol*. 2017 Sep 1;13(9):e1005708.
 20. Sheffield NC, Furey TS. Identifying and characterizing regulatory sequences in the human genome with chromatin accessibility assays. Vol. 3, *Genes*. *Multidisciplinary Digital Publishing Institute (MDPI)*; 2012. p. 651–70.
 21. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012 May 17;485(7398):376–80.
 22. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. Vol. 14, *Nature Reviews Genetics*. *Nat Rev Genet*; 2013. p. 390–403.

23. Spitz F. Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles. Vol. 57, *Seminars in Cell and Developmental Biology*. Academic Press; 2016. p. 57–67.
24. French JD, Edwards SL. The Role of Noncoding Variants in Heritable Disease. Vol. 36, *Trends in Genetics*. Elsevier Ltd; 2020. p. 880–91.
25. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, et al. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. 2013 Sep 25;502(7469):59–64.
26. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. Vol. 32, *Trends in Genetics*. Elsevier Ltd; 2016. p. 225–37.
27. Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol*. 2015 Dec;11(12):852.
28. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science (80-)*. 2002 Feb 15;295(5558):1306–11.
29. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Vol. 164, *Cell*. Cell Press; 2016. p. 1110–21.
30. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-)*. 2009 Oct 9;326(5950):289–93.
31. Gibcus JH, Dekker J. The Hierarchy of the 3D Genome. Vol. 49, *Molecular Cell*. Mol Cell; 2013. p. 773–82.
32. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, et al. Hi-C: A method to study the three-dimensional architecture of genomes. *J Vis Exp*. 2010;(39):1869.
33. Bell AC, West AG, Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*. 1999 Aug 6;98(3):387–96.
34. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*. 2008 Feb 14;451(7180):796–801.
35. Jansen RC, Nap JP. Genetical genomics: The added value from segregation. Vol. 17, *Trends in Genetics*. Elsevier Ltd; 2001. p. 388–91.

36. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Vol. 10, *Nature Reviews Genetics*. Nature Publishing Group; 2009. p. 184–94.
37. Westra HJ, Franke L. From genome to function by studying eQTLs. Vol. 1842, *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier B.V.; 2014. p. 1896–902.
38. Cotney J, Leng J, Oh S, DeMare LE, Reilly SK, Gerstein MB, et al. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res*. 2012 Jun 1;22(6):1069–80.
39. Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*. 2013 Jun 6;153(6):1281–95.
40. Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* (80-). 2020 Sep 1;369(6509):1318–30.
41. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: Models and data. Vol. 69, *American Journal of Human Genetics*. University of Chicago Press; 2001. p. 1–14.
42. Hormozdiari F, van de Bunt M, Segrè A V., Li X, Joo JWJ, Bilow M, et al. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am J Hum Genet*. 2016 Dec 1;99(6):1245–60.
43. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet*. 2013 Feb;45(2):124–30.
44. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: Regional visualization of genome-wide association scan results. In: *Bioinformatics*. Oxford University Press; 2011. p. 2336–7.
45. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet*. 2014;10(5).
46. Strunz T, Kiel C, Grassmann F, Ratnapriya R, Kwicklis M, Karlstetter M, et al. A mega-analysis of expression quantitative trait loci in retinal tissue. Liu Y, editor. *PLOS Genet*. 2020 Sep 1;16(9):e1008934.
47. Spielmann M, Mundlos S. Looking beyond the genes: The role of non-coding

- variants in human disease. Vol. 25, Human Molecular Genetics. Oxford University Press; 2016. p. R157–65.
48. High KA, Roncarolo MG. Gene Therapy. *N Engl J Med.* 2019 Aug 1;381(5):455–64.
 49. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. 2012 Aug 17;337(6096):816–21.
 50. Sternberg SH, Doudna JA. Expanding the Biologist’s Toolkit with CRISPR-Cas9. Vol. 58, *Molecular Cell*. Cell Press; 2015. p. 568–74.
 51. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-)*. 2014 Nov 28;346(6213):1258096.
 52. Zheng Q, Cai X, Tan MH, Schaffert S, Arnold CP, Gong X, et al. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *Biotechniques*. 2014 Sep 1;57(3):115–24.
 53. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013 Feb 28;152(5):1173–83.
 54. Kampmann M. CRISPR-based functional genomics for neurological disease. Vol. 16, *Nature Reviews Neurology*. Nature Research; 2020. p. 465–80.
 55. Hirakawa MP, Krishnakumar R, Timlin JA, Carney JP, Butler KS. Gene editing and CRISPR in the clinic: Current and future perspectives. Vol. 40, *Bioscience Reports*. Portland Press Ltd; 2020.
 56. Matharu N, Rattanasopha S, Tamura S, Maliskova L, Wang Y, Bernard A, et al. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science (80-)*. 2019 Jan 18;363(6424).
 57. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Vol. 13, *Nature Methods*. Nature Publishing Group; 2016. p. 127–37.
 58. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target effects in CRISPR/Cas9-mediated genome engineering. Vol. 4, *Molecular Therapy - Nucleic Acids*. Nature Publishing Group; 2015. p. e264.
 59. Uddin F, Rudin CM, Sen T. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. Vol. 10, *Frontiers in Oncology*. Frontiers Media S.A.; 2020.

60. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016 May 15;32(10):1479–85.
61. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*. 2003 Aug 5;100(16):9440–5.
62. Wallace C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. Epstein MP, editor. *PLOS Genet*. 2020 Apr 20;16(4):e1008720.
63. Gábor C, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Syst*. 2006;1695.
64. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen — The Clinical Genome Resource. *N Engl J Med*. 2015 Jun 4;372(23):2235–42.
65. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. *Nucleic Acids Res*. 2020 Jan 1;48(D1):D682–8.
66. R Core Team (2020). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet]. Available from: www.R-project.org/
67. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York; 2009.
68. Perenthaler E, Yousefi S, Niggli E, Barakat TS. Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. Vol. 13, *Frontiers in Cellular Neuroscience*. Frontiers Media S.A.; 2019. p. 352.
69. Yu J, Hu M, Li C. Joint analyses of multi-tissue Hi-C and eQTL data demonstrate close spatial proximity between eQTLs and their target genes. *BMC Genet*. 2019 Apr 30;20(1):43.
70. Pal K, Forcato M, Ferrari F. Hi-C analysis: from data generation to integration. Vol. 11, *Biophysical Reviews*. Springer Verlag; 2019. p. 67–78.
71. Wallace C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet*. 2021 Sep 29;17(9).
72. Madlung A, Comai L. The effect of stress on genome regulation and structure. Vol. 94, *Annals of Botany*. Oxford University Press; 2004. p. 481–95.
73. De Nadal E, Ammerer G, Posas F. Controlling gene expression in response to

- stress. Vol. 12, *Nature Reviews Genetics*. Nature Publishing Group; 2011. p. 833–45.
74. van der Wijst MGP, de Vries DH, Groot HE, Trynka G, Hon CC, Bonder MJ, et al. The single-cell eQTLGen consortium. *Elife*. 2020;9:1–21.
 75. Nathan A, Asgari S, Ishigaki K, Valencia C, Amariuta T, Luo Y, et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature*. 2022 Jun 2;606(7912):120–8.
 76. Beck CF, Warren RA. Divergent promoters, a common form of gene organization. Vol. 52, *Microbiological reviews*. American Society for Microbiology (ASM); 1988. p. 318–26.
 77. Wei W, Pelechano V, Järvelin AI, Steinmetz LM. Functional consequences of bidirectional promoters. Vol. 27, *Trends in Genetics*. NIH Public Access; 2011. p. 267–76.
 78. Trinklein ND, Force Aldred S, Hartman SJ, Schroeder DI, Otilar RP, Myers RM. An abundance of bidirectional promoters in the human genome. *Genome Res*. 2004 Jan;14(1):62–6.
 79. Hu RG, Brower CS, Wang H, Davydov I V., Sheng J, Zhou J, et al. Arginyltransferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. *J Biol Chem*. 2006 Oct 27;281(43):32559–73.
 80. de Winter JP, Joenje H. The genetic and molecular basis of Fanconi anemia. *Mutat Res - Fundam Mol Mech Mutagen*. 2009 Jul 31;668(1–2):11–9.
 81. Kobayashi H, Ohno S, Sasaki Y, Matsuura M. Hereditary breast and ovarian cancer susceptibility genes (Review). Vol. 30, *Oncology Reports*. *Oncol Rep*; 2013. p. 1019–29.
 82. Yang MQ, Koehly LM, Elnitski LL. Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes. *PLoS Comput Biol*. 2007 Apr;3(4):733–42.
 83. Birnbaum MD, Zhao N, Moorthy BT, Patel DM, Kryvenko ON, Heidman L, et al. Reduced Arginyltransferase 1 is a driver and a potential prognostic indicator of prostate cancer metastasis. *Oncogene*. 2019 Feb 7;38(6):838–51.
 84. Zakany J, Duboule D. The role of Hox genes during vertebrate limb development. Vol. 17, *Current Opinion in Genetics and Development*. Elsevier Current Trends; 2007. p. 359–66.
 85. Spitz F, Herkenne C, Morris MA, Duboule D. Inversion-induced disruption of the

- Hoxd cluster leads to the partition of regulatory landscapes. *Nat Genet.* 2005 Aug 3;37(8):889–93.
86. Nelson CE, Morgan BA, Burke AC, Laufer E, DiMambro E, Murtaugh LC, et al. Analysis of Hox gene expression in the chick limb bud. *Development.* 1996 May 1;122(5):1449 LP – 1466.
 87. Rodríguez-Carballo E, Lopez-Delisle L, Zhan Y, Fabre PJ, Beccari L, El-Idrissi I, et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* 2017 Nov 30;31(22):2264–81.
 88. Rupp LJ, Schumann K, Roybal KT, Gate RE, Ye CJ, Lim WA, et al. CRISPR/Cas9-mediated PD-1 disruption enhances anti-Tumor efficacy of human chimeric antigen receptor T cells. *Sci Rep.* 2017 Dec 1;7(1):1–10.
 89. Zhao Z, Shi L, Zhang W, Han J, Zhang S, Fu Z, et al. CRISPR knock out of programmed cell death protein 1 enhances anti-tumor activity of cytotoxic T lymphocytes. *Oncotarget.* 2018;9(4):5208–15.
 90. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. Vol. 26, *Annual Review of Immunology.* Annu Rev Immunol; 2008. p. 677–704.
 91. Yamanami H, Shiozaki K, Wada T, Yamaguchi K, Uemura T, Kakugawa Y, et al. Down-regulation of sialidase NEU4 may contribute to invasive properties of human colon cancers. *Cancer Sci.* 2007 Mar;98(3):299–307.
 92. Seko A, Nagata K, Yonezawa S, Yamashita K. Down-regulation of Gal 3-O-sulfotransferase-2 (Gal3ST-2) expression in human colonic non-mucinous adenocarcinoma. *Japanese J Cancer Res.* 2002;93(5):507–15.
 93. Brown SJ, McLean WHI. One remarkable molecule: Filaggrin. Vol. 132, *Journal of Investigative Dermatology.* Nature Publishing Group; 2012. p. 751–62.
 94. Smith FJD, Irvine AD, Terron-Kwiatkowski A, Sandilands A, Campbell LE, Zhao Y, et al. Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat Genet.* 2006 Mar 29;38(3):337–42.
 95. Esparza-Gordillo J, Weidinger S, Fölster-Holst R, Bauerfeind A, Ruschendorf F, Patone G, et al. A common variant on chromosome 11q13 is associated with atopic dermatitis. *Nat Genet.* 2009 May 6;41(5):596–601.
 96. Choi J, Kim D-I, Kim J, Kim B-H, Kim A. Hornerin Is Involved in Breast Cancer Progression. *J Breast Cancer.* 2016 Jun 1;19(2):142.

97. Fu SJ, Shen SL, Li SQ, Hua YP, Hu WJ, Guo BC, et al. Hornerin promotes tumor progression and is associated with poor prognosis in hepatocellular carcinoma. *BMC Cancer*. 2018 Aug 13;18(1).
98. Strunz T, Kellner M, Kiel C, Weber BHF. Assigning co-regulated human genes and regulatory gene clusters. *Cells*. 2021 Sep 1;10(9).

Danksagung

Zunächst möchte ich mich ganz besonders bei meinem Doktorvater Prof. Dr. Bernhard Weber bedanken; für die herausragenden Bedingungen am Institut für Humangenetik, unter denen ich dieses spannende und anspruchsvolle Thema bearbeiten durfte, für das stimmige Maß an Forderung und Förderung und für diesen prägenden Abschnitt meiner klinischen Laufbahn.

Ausgesprochen großer Dank gilt auch meinen Betreuern Dr. Tobias Strunz und Dr. Christina Kiel, die unermüdlich meine Fragen beantwortet haben und mir jederzeit mit offenem Ohr und gutem Rat zur Seite standen. Danke, dass ich in vielerlei Hinsicht so viel von euch lernen durfte.

An dieser Stelle möchte ich auch die vielen weiteren Mitglieder des Instituts, einschließlich der medizinischen Doktorandinnen und Doktoranden dieses Abschnittes, nicht unerwähnt lassen. Vielen Dank für das, was den gemeinsamen Alltag Tag für Tag besonders gemacht hat.

Besonders herzlicher Dank gilt auch meiner langjährigen Partnerin Lena, die mir in dieser Zeit sicher nicht zum ersten Mal gezeigt hat, dass ich mich zu 100 % auf sie verlassen kann und die mir eine große Stütze war. Danke für deine motivierenden Worte, für deine Geduld und für den zugesprochenen Mut.

Herzlichen Dank auch an meine engste Familie; an meine Schwester Sonja und meinen Bruder Konrad und allen voran an meine Mutter Brigitte, die Unvorstellbares geleistet hat, um mir das und noch viel mehr überhaupt zu ermöglichen. In gleichem Atemzuge von Herzen Danke an meinen verstorbenen Vater Hans für das persönliche Fundament und für die unverändert währenden Einflüsse. Ihm widme ich diese Arbeit.

Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeit erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ort, Datum und Unterschrift