

**Einsatz eines modifizierten Cas9-Systems zur Promotorforschung
im endogenen genomischen Kontext**

**Dissertation zur Erlangung des akademischen Grades
doctor medicinae (Dr. med.)**

vorgelegt dem Rat der Medizinischen Fakultät
der Friedrich-Schiller-Universität Jena

von Steven Kirchner

geboren am 27.02.1990 in Neuhaus am Rennweg

Gutachter

- 1. Prof. Dr. med. Reinhard Bauer, Jena**
- 2. Prof. Dr. med. Holger Scholz, Berlin**
- 3. Prof. Dr. med. Ralf Mrowka, Jena**

Tag der öffentlichen Verteidigung: 06.09.2022

Inhaltsverzeichnis

1	Zusammenfassung	1
2	Einleitung.....	2
2.1	Regulatorische Elemente der eukaryotischen Transkription.....	2
2.2	Gene-Editing	4
2.2.1	Zink-Finger-Nukleasen und Transkriptionsfaktorartige-Effektor-Nukleasen	5
2.2.2	Das CRISPR/Cas9-System.....	7
2.2.3	Gene-Editing mit CRISPR.....	9
2.3	Der proximale Renin-Promotor als genomische Zielsequenz dieser Arbeit.	12
2.4	Reporterstudien in der Promotorforschung.....	14
2.5	Modellierung experimenteller Daten in der Transkriptionsforschung.....	15
2.6	Kombinatorischer Ansatz zur Erforschung <i>cis</i> -regulatorischer Interaktionen	16
3	Ziele der Arbeit	18
4	Publizierte Originalarbeit	19
5	Diskussion	20
5.1	Gen-Aktivierung durch einen modifizierten Cas9-Komplex	20
5.2	Erforschung <i>cis</i> -regulatorischer Interaktionen durch einen kombinierten experimentellen und modellierenden Ansatz	21
5.3	Promotorforschung im endogenen genomischen Kontext.....	23
5.4	Methodendiskussion.....	24
5.4.1	Einsatz der Cas9 zur Promotorforschung	24
5.4.2	Experimentelles Setup	27
6	Schlussfolgerungen	28
7	Literatur- und Quellenverzeichnis	29

8	Anhang.....	34
8.1	Danksagung	34
8.2	Abbildungsverzeichnis.....	35
8.3	Ehrenwörtliche Erklärung	36

Abkürzungsverzeichnis

bp	<i>Basenpaare</i>
cAMP	<i>cyclisches Adenosinmonophosphat</i>
Cas	<i>CRISPR associated</i>
CRISPR	<i>clustered regulatory interspaced short palindromic repeat</i>
GFP	<i>green fluorescent protein</i>
HEK	<i>human embryonic kidney</i>
mRNA	<i>messenger RNA</i>
NHEJ	<i>non-homologous end joining</i>
REN	<i>humanes Renin-Gen</i>
RVD	<i>repeat-variable diresidues</i>
SAM	<i>synergistic activation mediator</i>
sgRNA	<i>single guide RNA</i>
TALE	<i>transcription activator-like effector</i>
TALEN	<i>transcription activator-like effector nuclease</i>
TSS	<i>transcription start site</i>
ZF	<i>Zink-Finger-Domäne</i>
ZFN	<i>zinc-finger nuclease</i>

1 Zusammenfassung

Das Verständnis der komplexen Mechanismen der menschlichen Transkriptionsregulation bleibt trotz vollständiger Sequenzierung des humanen Genoms eine große Herausforderung. Zur Erforschung wichtiger nicht kodierender DNA-Elemente stehen heute mehrere molekulargenetische Werkzeuge zur Verfügung. Unter Verwendung eines modifizierten Cas9-Komplexes entwickelten wir einen kombinierten experimentellen und modellierenden Ansatz zur Untersuchung *cis*-regulatorischer Interaktionen im nativen genomischen Kontext. Es wurde zunächst untersucht, ob sich durch gezielte Einflussnahme auf einen Promotor eine Gen-Expression auslösen ließ. In der Folge wurde nach möglichen statistischen Interaktionen der untersuchten Promotorregion gesucht.

Mit Hilfe einer katalytisch aktiven Cas9 haben wir das native humane Renin-Gen mit einem Firefly-Luziferase-kodierenden Gen stabil markiert, sodass die potenzielle Induktion der Genexpression durch ein optisches Signal messbar wurde.

Für die anschließende experimentelle Arbeit fand eine modifizierte Cas9 Anwendung. Diese katalytisch inaktive Variante rekrutiert einen Komplex aus Transkriptionsfaktoren (SAM-Komplex), durch den die Expression eines gewünschten Zielgens induziert wird. Durch fünf designte Guide-RNAs konnten wir diesen SAM-Komplex zu spezifischen Promotorregionen des humanen Renins leiten. Die programmierten SAM-Komplexe induzierten in unseren Zellen die Aktivierung von Renin. Die resultierenden Genexpressionen wurden gemessen und in Computermodellen ausgewertet.

Durch die Anwendung unseres kombinierten experimentellen und modellierenden Ansatzes können wir zeigen, dass relevante Interaktionen innerhalb der ausgewählten Sequenzen der proximalen Renin-Promotorregion auftreten. Unsere Ergebnisse zeigen, dass mit Hilfe von Weiterentwicklungen moderner molekulargenetischer Verfahren wie des Cas9-Systems und einem kombinatorischen, experimentellen und modellierenden Setup, eine suffiziente Promotorforschung im endogenen Kontext möglich wird. Unser kombinierter Ansatz könnte auch für die Untersuchung anderer genomischer Regionen nützlich sein.

2 Einleitung

2.1 Regulatorische Elemente der eukaryotischen Transkription

Die Transkription ist der erste entscheidende Schritt der DNA-basierten Genexpression. Ein bestimmtes Segment kodierender DNA wird hierbei durch das Enzym RNA-Polymerase in RNA, im Falle der Proteinbiosynthese in kodierende mRNA kopiert. Die mRNA wiederum dient dann als Vorlage für die Synthese des entsprechenden Proteins durch Translation. Alternativ kann das transkribierte Gen auch für nicht-kodierende, regulatorische RNA-Moleküle kodieren. Die Transkription unterliegt zahlreichen regulatorischen Einflüssen. Eine zentrale Rolle nehmen dabei die sogenannten *cis*-regulatorischen Elemente und die Transkriptionsfaktoren ein.

Die transkriptionelle Regulation hat den größten Einfluss auf entweder Aktivierung oder Suppression von Genen und wird maßgeblich über *cis*-regulatorische Elemente kontrolliert (Hernandez-Garcia und Finer 2014).

Cis-regulatorische Elemente sind Bestandteil der nicht-kodierenden DNA und finden sich im Bereich von Promotoren. Sie dienen hier u.a. als Bindungsstelle für Transkriptionsfaktoren. Diese Elemente bilden komplexe Systeme. Durch die Regulierung dieser komplexen Systeme aus tausenden von Genen wird die morphologische Reifung von Zellen sowie deren differenzierte Funktion ermöglicht (Yuh et al. 1998). Das Verständnis der genomischen *cis*-regulatorischen Netzwerke ist daher von großer Bedeutung. Es gibt Beispiele für Mutationen von regulatorischen Promotorregionen, die zu schweren Erkrankungen führen (Soufi et al. 2018).

Oftmals befindet sich innerhalb der Promotorregion ein Kernpromotor im Bereich von ~40 Basenpaaren (bp) stromaufwärts der TSS (transcription start site) mit konservierter TATA-Box als Bindestelle für allgemeine Transkriptionsfaktoren (Molina und Grotewold 2005). Aber es finden sich auch *cis*-regulatorische Elemente weiter proximal der Kernpromotorregion. Je nach Wirkung in Bezug zur Genexpression werden diese Elemente als sogenannte Enhancer oder Silencer beschrieben. Auch distal und selbst innerhalb des zu transkribierenden Genes wurden regulatorische Regionen identifiziert (Voigtlander et al. 1999).

Neben den *cis*-regulatorischen Elementen sind die Transkriptionsfaktoren (TFs) essenziell, um durch Binden an die DNA die Aktivierung oder Unterdrückung der Gentranskription zu ermöglichen. Es gibt Unterschiede in der Art und Weise wie TFs bei der Regulierung der Genexpression wirken. In einer aktuellen Übersichtsarbeit wird zwischen ca. 1600 menschlichen TFs unterschieden, die ~8 % aller menschlichen Gene repräsentieren (Lambert et al. 2018). Es existieren mehrere Möglichkeiten TFs zu klassifizieren. Generell ist eine Einteilung in basale oder allgemeine TFs und spezifische TFs möglich. Basale TFs sind in allen Zellen ubiquitär vorhanden und für die Transkriptionsinitiation, z.B. über Bindung an TATA-Boxen, notwendig (Reese 2003, Shilatifard et al. 2003). Im Gegensatz dazu zeigen spezifische TFs nur in bestimmten Geweben und in bestimmten Entwicklungsstadien Aktivität. Sie können an spezifische *cis*-regulatorische Elemente, z.B. innerhalb von Promotoren, Enhancern oder Silencern binden und sind für die Regulation zentraler Mechanismen wie der Zellentwicklung oder der Reaktion auf Reize über Signalkaskaden notwendig (Lobe 1992, Pawson 1993).

2.2 Gene-Editing

Die gezielte Einflussnahme auf Gene ist zur Entschlüsselung deren Funktion und Rolle im Organismus essenziell. Über die letzten 2 Jahrzehnte haben sich in diesem Feld, dem sogenannten Gene-Editing, verschiedene molekulargenetische Techniken zur Untersuchung der Gene und ihrer regulatorischen Elemente etabliert. Zu den heute etablierten Methoden zählen die auch als Designer-Nukleasen bezeichneten Zink-Finger-Nukleasen (Zinc-Finger Nuclease, ZFN), die Transkriptionsfaktorartigen-Effektor-Nukleasen (Transcription Activator-like Effector Nuclease, TALEN) und das CRISPR/Cas9-System. Diese Methoden revolutionierten die molekulargenetische Grundlagenforschung und verdrängten bis dahin bestehende Verfahren wie die pronukleäre Mikroinjektion, die zielgerichtete Modifikation des Genoms in embryonalen Stammzellen oder die gezielte Inaktivierung von Genen durch RNA Interferenz (Bednarski und Cathomen 2015, Gaj et al. 2013).

Von den modernen Werkzeugen des Gene-Editing verspricht man sich zudem viel als Einsatz zur Therapie bestimmter Erkrankungen. TALENs und ZFNs konnten in kleinen Studien bereits erfolgreich zur Therapie von HIV, Leukämie, Sichelzellanämie und β -Thalassämie eingesetzt werden (Reardon 2015, Frangoul et al. 2021). Inzwischen wurde sogar ein Verfahren entwickelt, dass die Sicherheit des Einsatzes von TALENs und Cas9 im menschlichen Körper einzuschätzen versucht (Turchiano et al. 2021).

2.2.1 Zink-Finger-Nukleasen und Transkriptionsfaktorartige-Effektor-Nukleasen

ZFNs stellen die älteste Gruppe der modernen molekulargenetischen Verfahren des Gene-Editing dar. Die prinzipielle Wirkungsweise dieser synthetischen Enzyme beruht zum einen auf einer spezifischen Sequenzerkennung der DNA mittels einer DNA-bindenden Untereinheit. Über eine weitere katalytische Untereinheit wird dann die erwünschte Wirkung am Genom erzielt. Durch Fusion mit Endonukleasen können zielgerichtet DNA-Doppelstrangbrüche erzeugt werden, was zur Aktivierung zellulärer Reparaturmechanismen, wie der homologen Rekombination oder der nicht homologen Endverknüpfung (NHEJ) führt (Carroll 2011, Wu et al. 2007). Im Jahr 2003 gelang es zum ersten Mal, gezielt das Erbgut der Fruchtfliege mit Hilfe von ZFNs zu verändern (Bibikova et al. 2003). Die spezifische DNA-Bindung erfolgt hierbei über eine Zink-Finger-Domäne (ZF), einem ca. 30 Aminosäuren langen Peptid, welches sich um ein zentrales Zinkion organisiert. So kann eine spezifische Sequenz von 3 bp erkannt werden (Abbildung 1). ZFs enthaltende Proteine stellen dabei die größte Gruppe von bekannten Transkriptionsfaktoren des menschlichen Genoms dar (Carroll 2011, Gersbach und Perez-Pinera 2014). Für eine ZFN werden gewöhnlich 3 bis 4 ZFs für eine spezifische DNA-Bindung von 9 bis 12 bp fusioniert. Durch Koppeln an eine katalytische Domäne, wie der Endonuklease *FokI*, erhält man ein Gene-Editing-Tool zum gezielten Einführen eines DNA-Strangbruches. Abhängig von der katalytischen Domäne und dem Design der Bindedomäne können Alterationen der DNA im Sinne von Punktmutationen, Deletionen, Insertionen, Inversionen oder Translokationen vorgenommen werden. Auch gezieltes Aktivieren oder Inaktivieren von Genen durch Koppeln an entsprechende Domänen ist so möglich (Beerli et al. 2000, Urnov et al. 2010, Gaj et al. 2013). Konformationsänderungen durch ein freies Kombinieren neuer ZFs können jedoch schwer vorhersehbare Wechselwirkungen bedingen, wodurch eine zufriedenstellende Funktion der gesamten DNA-Bindedomäne ausbleiben kann und sich die de-novo Synthese einer ZFN oftmals als langwieriger und kostenaufwendiger Prozess darstellt (Wang et al. 2013).

Im Jahr 2011 gelang die Etablierung eines weiteren Werkzeuges des Gene-Editing, den TALENs (Bednarski und Cathomen 2015). Der prinzipielle Aufbau entspricht dabei dem der ZFNs. Auch hier wurde ursprünglich eine DNA-Bindedomäne an eine

katalytische *FokI*-Nukleasedomäne fusioniert. Die DNA-Bindedomäne basiert jedoch auf repetitiven Peptiden oder Modulen mit jeweils 33 – 35 Aminosäuren einer hoch konservierten Struktur, den TALEs (Transcription Activator-like Effectors). Diese wurden ursprünglich aus *Xanthomonas*, einem phytopathogenen gram-negativen Bakterium, isoliert (Greiner 2015). Die Spezifität der DNA-Bindung wird allein über 2 hypervariable Aminosäuren an Position 12 und 13, den sogenannten RVDs (repeat-variable diresidues) erreicht (Boch et al. 2009). Ein Modul eines TALE bindet dabei im Unterschied zu den ZFs spezifisch ein einzelnes Nukleotid (Abbildung 1). Dementsprechend können mit vier TALE-Modulen alle Nukleotide des humanen Genoms spezifisch gebunden werden. Üblicherweise besteht die DNA bindende Domäne aus 15 bis 18 solcher Module (Bednarski und Cathomen 2015, Joung und Sander 2013).

Der einfachere modulare Aufbau und weniger konformationsabhängige Wechselwirkungen im Vergleich zu ZFNs ließen die TALENs schnell zu einem effizienten und weit verbreiteten Werkzeug des Gene-Editing heranreifen. So gelang bereits kurz nach Vorstellung dieser Methode eine gezielte DNA-Insertion in das Zebrafisch Genom (Zu et al. 2013, Bedell et al. 2012). Durch die Möglichkeiten der beliebigen Fusion an katalytische Domänen wie Nukleasen, Transkriptionsaktivatoren oder -repressoren sind TALENs inzwischen als eine Standardmethode zur gezielten Genmodulation etabliert (Gaj et al. 2013).

2.2.2 Das CRISPR/Cas9-System

CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats) und die CRISPR assoziierten Proteine (Cas) bilden ein adaptives Immunsystem in Bakterien (Barrangou et al. 2007, Ishino et al. 1987). Mehrere Klassen von CRISPR-Systemen werden unterschieden, wobei die CRISPR assoziierte Cas9 aus dem Typ II-System des Bakteriums *Streptococcus pyogenes* heute die größte Rolle für CRISPR basiertes Gene-Editing spielt (Sander und Joung 2014).

Die Cas9 bindet eine frei programmierbare Guide-RNA (sgRNA), worüber eine spezifische DNA-Bindung ermöglicht wird. Über enzymeigene Endonukleasedomänen erfolgt die zielgerichtete Einführung eines DNA-Doppelstrangbruchs (Abbildung 1).

Die Art der DNA-Bindung stellt den wesentlichen Unterschied zu auf ZFNs oder TALENs basierendem Gene-Editing dar, bei welchen Proteinstrukturen für die spezifische Sequenzerkennung notwendig sind. Die sgRNA besteht aus einem Anteil, der für die Cas9-Bindung essenziell ist und einem ~ 20 bp langem Anteil (Spacer), der homolog zur genomischen Zielsequenz ist und eine Anlagerung der Cas9 an die genomische Zielsequenz über die Watson-Crick-Basenpaarung zwischen sgRNA und DNA ermöglicht (Perez-Pinera et al. 2013).

Eine wichtige Voraussetzung für das Design der sgRNA ist das Vorhandensein der PAM-Stelle (Protospacer Adjacent Motif) unmittelbar nach der Guide-Sequenz, durch welche viele Cas-Proteine erst die volle Funktionsfähigkeit entwickeln (Shah et al. 2013). Die PAM Sequenz besteht meist aus drei Nukleotiden und variiert in den verschiedenen CRISPR-Systemen. In *Streptococcus pyogenes* ist die Basenfolge -NGG-.

Für das Design potenzieller sgRNAs stehen offizielle bioinformatische Tools zur Verfügung, die entsprechende Voraussetzungen berücksichtigen (Zhang Lab 2020).

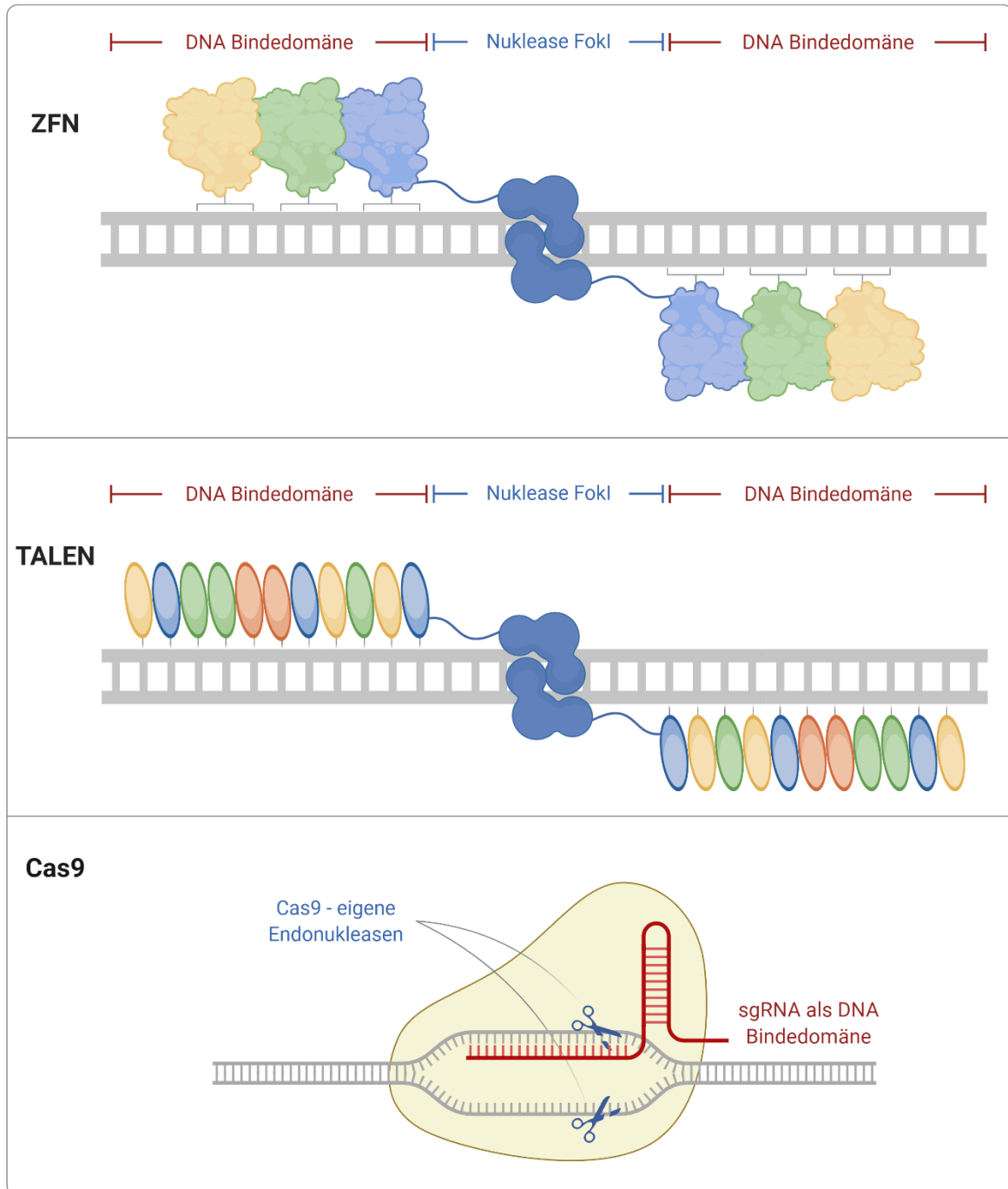


Abbildung 1: Designer-Nukleasen im Vergleich. ZFN: üblicherweise 3 bis 4 Zinkfingerdomänen (ZF) bilden eine DNA-Bindedomäne; eine ZF bindet jeweils ein Basentriplett; durch Dimerisierung der gekoppelten Nuklease *FokI* erfolgt die spezifische Einführung eines Doppelstrangbruches. TALEN: zumeist 15 bis 18 TALE-Module formen eine DNA-Bindedomäne; ein TALE-Modul bindet spezifisch ein Nukleotid; durch Dimerisierung der gekoppelten Nuklease *FokI* erfolgt die spezifische Einführung eines Doppelstrangbruches. Cas9: die DNA-Bindung erfolgt über eine frei programmierbare sgRNA; durch die Cas9-eigenen Nukleasen erfolgt die spezifische Einführung eines Doppelstrangbruches. Erstellt mit *Biorender.com*.

2.2.3 Gene-Editing mit CRISPR

2012 gelangen Jinek et al. mit Hilfe einer sgRNA gesteuerten Cas9 die ersten programmierten DNA-Doppelstrangbrüche (Jinek et al. 2012). Seitdem hat sich das System zu einem effizienten Werkzeug des programmierbaren Gene-Editings etabliert (Abbildung 2). Ein Gen-Knockout wird erzielt durch die oft fehlerbehaftete nicht homologe Endverknüpfung, die sich meist als ein möglicher DNA-Reparaturmechanismus an den erfolgten spezifischen DNA-Doppelstrangbruch anschließt. Die oftmals resultierenden Insertionen, Deletionen oder Frameshift-Mutationen können dann zur Stilllegung des entsprechenden Gens führen (Wyman und Kanaar 2006). Gezieltes Ausnutzen des zweiten möglichen DNA-Reparaturmechanismus, der homologen Rekombination, ermöglicht die gezielte Insertion einzelner Basen bis hin zu ganzen Genen. Durch Bereitstellen einer entsprechenden Gensequenz kann so die stabile Integration in das Genom erfolgen. So ist beispielsweise das Markieren eines Zielgenes mit einem Reporter-gen (Tagging) möglich (Yang et al. 2013).

Ein relevantes Problem beim gezielten Gene-Editing stellen off-target Effekte dar. Die designten sgRNAs können möglicherweise außerhalb der gewünschten genomischen Zielsequenz ausreichend gute Homologien aufweisen, die zum Aktivwerden der Cas9 an eben diesen Stellen führt. Dies resultiert in schwer vorhersehbaren Folgen für die Funktionsfähigkeit der Zelle bzw. des Organismus. Zahlreiche Weiterentwicklungen, besonders durch Mutationen der katalytischen Domänen, ermöglichten die Etablierung von Cas9 Varianten, die eine noch größere Spezifität bei weniger unerwünschten Nebeneffekten zeigen (Kleinstiver et al. 2016, Slaymaker et al. 2016).

Auch die Modifikation der Cas9 zu einer Nickase, welche gezielt DNA-Einzelstrangbrüche einführt, adressiert die Limitation potenzieller off-target Effekte. Hierdurch werden für einen gezielten DNA-Doppelstrangbruch zwei sgRNAs benötigt, welche die jeweils angelagerten Cas9-Nickasen zu eng benachbarten und gegenüberliegenden Sequenzen translozieren (Ran et al. 2013b, Richardson et al. 2016).

Durch gezielte Mutationen der katalytischen Domänen war es möglich, der Cas9 ihre ursprüngliche Funktion der Nuklease zu entziehen und sie somit zu einem effizienten

DNA-bindenden Protein (dCas9) weiterzuentwickeln (Jinek et al. 2012). Im Jahr 2015 stellten Konermann et al. eine dCas9 vor, die an einen Komplex aus verschiedenen Transkriptionsfaktoren gekoppelt wurde. So entstand ein hocheffizienter, sgRNA gesteuerter *trans*-Aktivierungskomplex, der sogenannte SAM-Komplex (Maeder et al. 2013, Konermann et al. 2013, Konermann et al. 2015, Perez-Pinera et al. 2013, Gilbert et al. 2013). Die Entwicklung dieses Gene-Editing-Tools stellt eine Erweiterung der Möglichkeiten der Erforschung von Genen und ihren regulatorischen Elementen im endogenen Kontext und in ihrer natürlichen Umgebung im Genom dar.

Auch das Abschalten von Genen ist über eine Bindung der dCas9 an inaktivierende Domänen möglich (Alerasool et al. 2020).

Durch Fusion eines Fluoreszenzmarkers wie GFP (green fluorescent protein) an eine katalytisch inaktive dCas9 erhält man einen programmierbaren DNA-Marker, mit dem gezielt genomische Loci in lebenden Zellen sichtbar gemacht werden können (Chen et al. 2013). Durch Fusion einer dCas9 an epigenetische Modifikatoren, sind beispielsweise Änderungen des Methylierungszustandes oder die Histonacetylierung programmierbar (Nuñez et al. 2021). Dies ermöglicht eine Einflussnahme auf die Genexpression ohne das Einführen von DNA-Doppelstrangbrüchen.

Inzwischen sind auch verschiedene Cas9-Systeme etabliert, die durch ein Einbringen unterschiedlicher sgRNAs innerhalb einer Zelle ein spezifisches und zeitgleiches Gene-Editing an mehreren genomischen Loci ermöglichen (Sakuma et al. 2014).

Darüber hinaus macht die einfache Generierung von sgRNAs sowie die Möglichkeit, nahezu jeden genomischen Locus anzusteuern, CRISPR zu einem sehr skalierbaren Gene-Editing-Tool, was sich in einem zunehmenden Einsatz für genomweite Screenings widerspiegelt (Doench et al. 2016, Sanson et al. 2018).

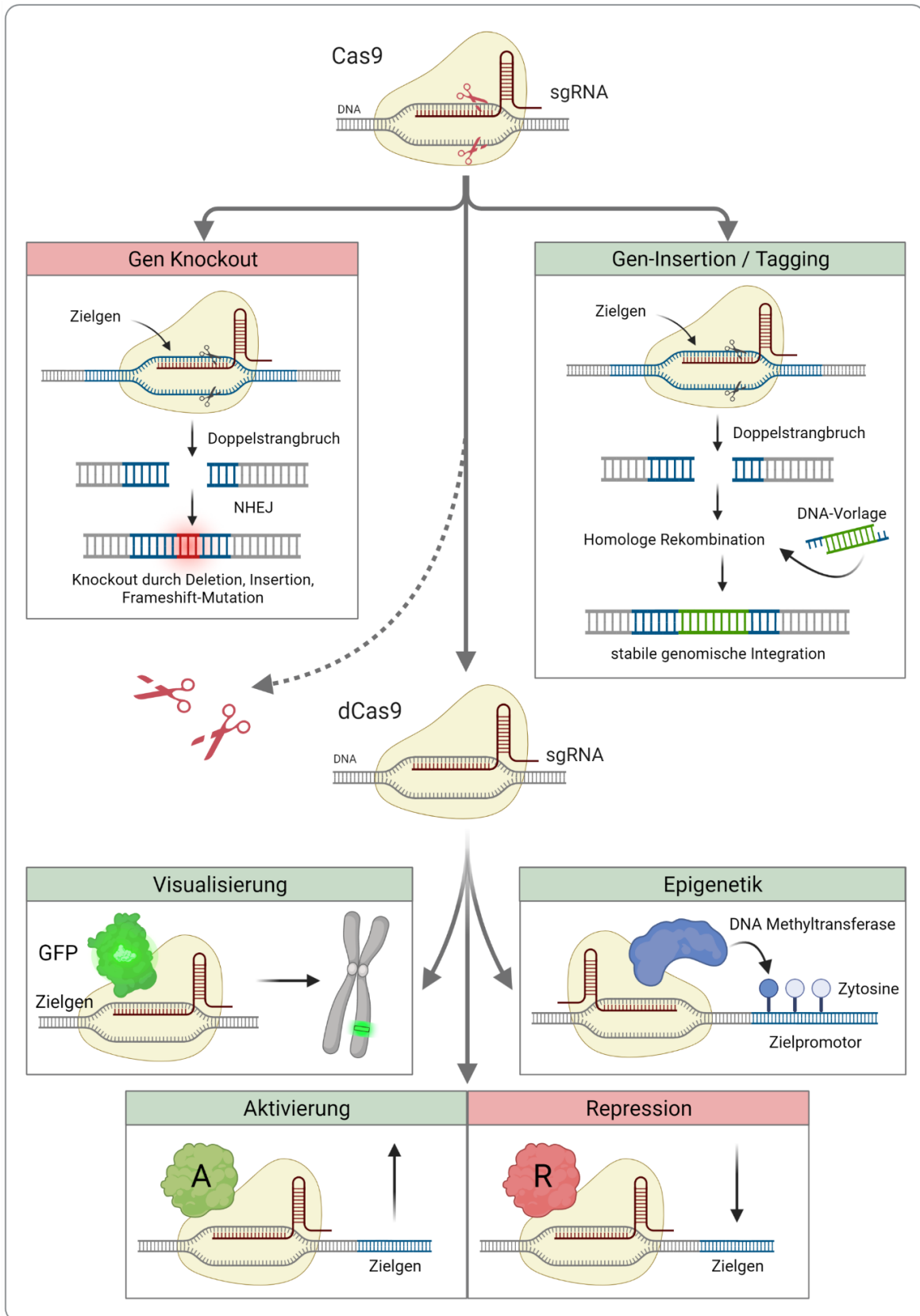


Abbildung 2: Gene-Editing mit CRISPR/Cas9. Dargestellt ist eine vereinfachte, schematische Übersicht der Möglichkeiten des Gene-Editing mit CRISPR/Cas9. Durch gezielte Mutationen und Deaktivierung der Endonuklease-Domänen wurde ein spezifisches DNA-Bindeprotein (dCas9) generiert, wodurch sich die Möglichkeiten des programmierbaren Gene-Editing erweitern. Erstellt mit *Biorender.com*.

2.3 Der proximale Renin-Promotor als genomische Zielsequenz dieser Arbeit

Die genomische Zielsequenz für die vorliegende Arbeit der Promotorforschung im endogenen Kontext bildet der proximale Renin-Promotor. Renin gilt als das Schlüsselenzym im Renin-Angiotensin-Aldosteron-System (RAAS). Durch die Regulation des Salz- und Wasserhaushaltes und des arteriellen Blutdrucks wird das RAAS zu einem lebenswichtigen System des menschlichen Körpers (Patel et al. 2017). Eine abnorme Aktivierung des RAAS kann u.a. zur Entwicklung von Bluthochdruck, kardialer Hypertrophie und Herzinsuffizienz beitragen (Sparks et al. 2014, Kurtz 2012). Nach Angaben der WHO leiden weltweit etwa 1,13 Milliarden Menschen an Bluthochdruck, einem alleinstehenden Hauptrisikofaktor vorzeitiger kardiovaskulärer Mortalität (World Health Organization 2019, september 13). Obwohl das Haupteffektormolekül des RAAS das Angiotensin II (ANG II) ist, wird die Regulation von ANG II und seinen Vorstufen hauptsächlich durch die Expression von Renin reguliert (Schweda et al. 2007).

Das humane Renin-Gen (REN) liegt auf Chromosom 1 (Chirgwin et al. 1984). Für REN konnten bereits wichtige *cis*-regulatorische Elemente identifiziert werden (Abbildung 3). So konnte ein "Renin-Enhancer" (ca. -12.000 bp zum Transkriptionsstart (TSS)) als Teil einer evolutionär konservierten Region (hREnc-Region) identifiziert werden, der als wichtig für die basale REN-Expression angesehen wird (Petrovic et al. 1996, Zhou et al. 2006, Mrowka et al. 2003). Weiter proximal wurde ein "Chorion-Enhancer" (ca. -5.500 bp zur TSS) gefunden. Seine Relevanz ist jedoch noch weitestgehend unklar (Fuchs et al. 2002).

In der Region von +200 bp bis -6 bp bezogen zum TSS befindet sich der proximale Renin-Promotor (Borensztein et al. 1994). Mit Hilfe von Experimenten an transgenen Mäusen wurde gezeigt, dass er eine bedeutende Rolle für die Gewebe- und Zellspezifität der REN-Expression spielt (Zhou et al. 2006, Glenn et al. 2008). Die Region zeigt bis hin zu einer vollständig konservierten TATA-Box ausgeprägte Homologien zwischen Maus und Mensch (Castrop et al. 2010). In murinen Reporterassays wurde dem proximalen Promotor nur ein geringes Aktivierungspotential zugesprochen (Petrovic et al. 1996, Pan et al. 2001).

Das Aktivierungspotential des murinen Enhancers wird jedoch als größer beschrieben als das des humanen Enhancers, was die Bedeutung des proximalen Promotors innerhalb des humanen REN aufwerten könnte (Itani et al. 2007).

Klassische Reporterstudien und Electrophoretic Mobility-Shift Assays (EMSA) haben bereits zahlreiche Bindestellen für Transkriptionsfaktoren der proximalen Promotorregion identifiziert. So gibt es beispielsweise Evidenz für Bindestellen, die für die Genregulation des REN über den sekundären Botenstoff cAMP wichtig sind (Germain et al. 1996, Tamura et al. 1994, Ying et al. 1997).

Es wurden jedoch auch regulatorische Elemente gefunden, deren Relevanz für die transkriptionelle Regulation von REN kontrovers diskutiert werden (Todorov et al. 2007, Todorov et al. 2008, Pan und Gross 2005).

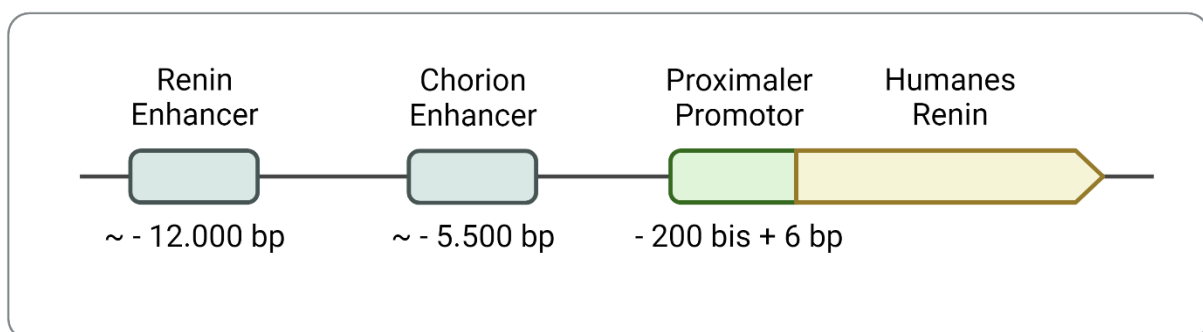


Abbildung 3: Humanes Renin-Gen. Das humane Renin-Gen verfügt über mehrere 5'-flankierende cis-regulatorisch wirksame Elemente. Dazu zählen der bei ca. -12 kbp gelegene Renin-Enhancer, der bei ca -5,5 kbp gelegene Chorion-Enhancer und die proximale Promotorregion. Erstellt mit *Biorender.com*.

2.4 Reporterstudien in der Promotorforschung

Eine molekulargenetische Standardmethode zur Erforschung *cis*-regulatorischer Elemente stellen Plasmid basierte Reporterassays dar. Hierzu werden zunächst Restriktionsstücke der zu untersuchenden DNA-Region in Vektoren kloniert. Den Sequenzen folgen nachgeschaltete Reportergene, die beispielsweise GFP oder Luziferase kodieren, deren Expression auf unterschiedliche Weise quantifiziert werden kann. Nach Transfektion der Vektoren in Zellen kann anschließend aus dem Expressionsniveau der Reportergene auf die Aktivität der untersuchten Promotorregion geschlossen werden. Auch die meisten Informationen über *cis*-regulatorische Elemente des REN konnten durch Zellkulturexperimente mit Hilfe solcher klassischen Reporterassays gewonnen werden (Castrop et al. 2010, Glenn et al. 2013). Allerdings werden die betreffenden DNA-Sequenzen bei dieser Methode aus ihrem endogenen Kontext herausgelöst (Borensztein et al. 1994, Petrovic et al. 1996, Shi et al. 1999, Voigtlander et al. 1999, Pan et al. 2001, Pan et al. 2004). Die Experimente erfolgen also außerhalb der natürlichen Umgebung des Promotors.

2.5 Modellierung experimenteller Daten in der Transkriptionsforschung

Die Möglichkeit mit Hilfe mathematischer Modellierung Rückschlüsse auf die Dynamik nicht kodierender DNA-Elemente ziehen zu können, wurde bereits 2003 in einer Studie am Bakterium *Escherichia coli* gezeigt (Setty et al. 2003). Darüber hinaus konnten in einer 1998 veröffentlichten Arbeit komplexe Prozesse einer wichtigen Promotorregion des Genoms von Seeigeln aufgedeckt werden, die für die Embryonalentwicklung von großer Bedeutung sind (Yuh et al. 1998). Bezogen auf REN konnten beispielsweise Prinzipien identifiziert werden, nach denen einzelne Abschnitte eines *cis*-regulatorischen Elements etwa 14.000 bp stromaufwärts von REN miteinander interagieren (Mrowka et al. 2007). So kann die mathematische Modellierung helfen, das Verständnis des komplexen transkriptionellen Netzwerks zu vertiefen.

Bei der statistischen Untersuchung von Faktoren, die eine Variable beeinflussen, ist es von allgemeinem wissenschaftlichem Interesse, herauszufinden, ob diese Faktoren unabhängig oder abhängig voneinander wirken. Diese Frage nach komplexen Interaktionen wurde unserer Meinung nach in der Forschung zu *cis*-regulatorischen Regionen, insbesondere zum proximalen Renin-Promotor, zu wenig beachtet. Im Jahr 2007 wurde nach unserem Kenntnisstand zum ersten Mal eine Studie veröffentlicht, die unter Verwendung eines modellierenden Ansatzes versuchte, Gesetzmäßigkeiten regulatorischer Elemente von REN aufzuklären (Mrowka et al. 2007). Allerdings waren auch hier die Experimente Plasmid basiert und erfolgten damit außerhalb des endogenen Kontextes. Außerdem stand eine regulatorische Region im Fokus, die etwa 14.000 Basenpaare stromaufwärts des TSS von REN liegt.

2.6 Kombinatorischer Ansatz zur Erforschung *cis*-regulatorischer Interaktionen

Um mögliche komplexe Interaktionen innerhalb der ausgewählten Promotorregion zu untersuchen, haben wir einen neuartigen kombinierten Ansatz angewendet. Zunächst besteht dieser Ansatz aus kombinatorischen Transfektionen von fünf ausgewählten sgRNAs, die den SAM-Komplex in eine jeweils spezifische Region des endogenen Zielpromotors translozieren. Die resultierenden Expressionsniveaus durch die verschiedenen Kombinationen der anvisierten Promotorregionen können durch eine Luziferaseaktivität quantifiziert werden (Abbildung 4). Nachfolgend wurden zwei verschiedene mathematische Modelle erstellt und unsere experimentellen Daten darin ausgewertet. Bei der Modellierung experimenteller Daten sollte das Prinzip der Simplizität beachtet werden. Die einfachste Annahme wäre, dass die untersuchten Regionen völlig unabhängig voneinander die Genexpression beeinflussen. Dies spiegelt sich in unserem ersten gewählten Modell (Summenmodell) wider. Das Summenmodell beschreibt eine unabhängige Beziehung der von uns untersuchten Promotorsequenzen zu ihrem Einfluss auf die Gen-Aktivierung. Dieser Ansatz ist vergleichbar mit Erkenntnissen, die klassische Promotorstudien liefern können, da diese wie bereits beschrieben, nur eine eingeschränkte Möglichkeit bieten, komplexe Interaktionen zu untersuchen. Eine wichtige wissenschaftliche Frage bei der Modellierung ist jedoch, ob eine statistische Interaktion existiert (Nieuwenhuis et al. 2011). Um diese Frage zu klären, wurde ein zweites mathematisches Modell erstellt (Interaktionsmodell). Dieses Modell stellt eine komplexere Annahme der Bedingungen in der Region des gewählten Zielpromotors dar. Das Interaktionsmodell erlaubt Abhängigkeiten zwischen den ausgewählten Promotorregionen, um die Gen-Aktivierung zu erklären. Hinsichtlich der Modellierung haben wir das multiple lineare Regressionsmodell verwendet, um die linearen Parameter anzupassen, was eine statistische Standardmethode ist (Fisher 1922). Um zu überprüfen, welches der generierten Modelle die gemessenen Daten am besten erklären kann, wurde zum einen der jeweilige absolute Vorhersagefehler berechnet. Dem Prinzip der maximalen Parsimonie bei der Modellbildung folgend, wurde weiterhin das jeweilige Akaike-Informationskriterium (AIC) zur Modellbeurteilung berechnet (Akaike 1981).

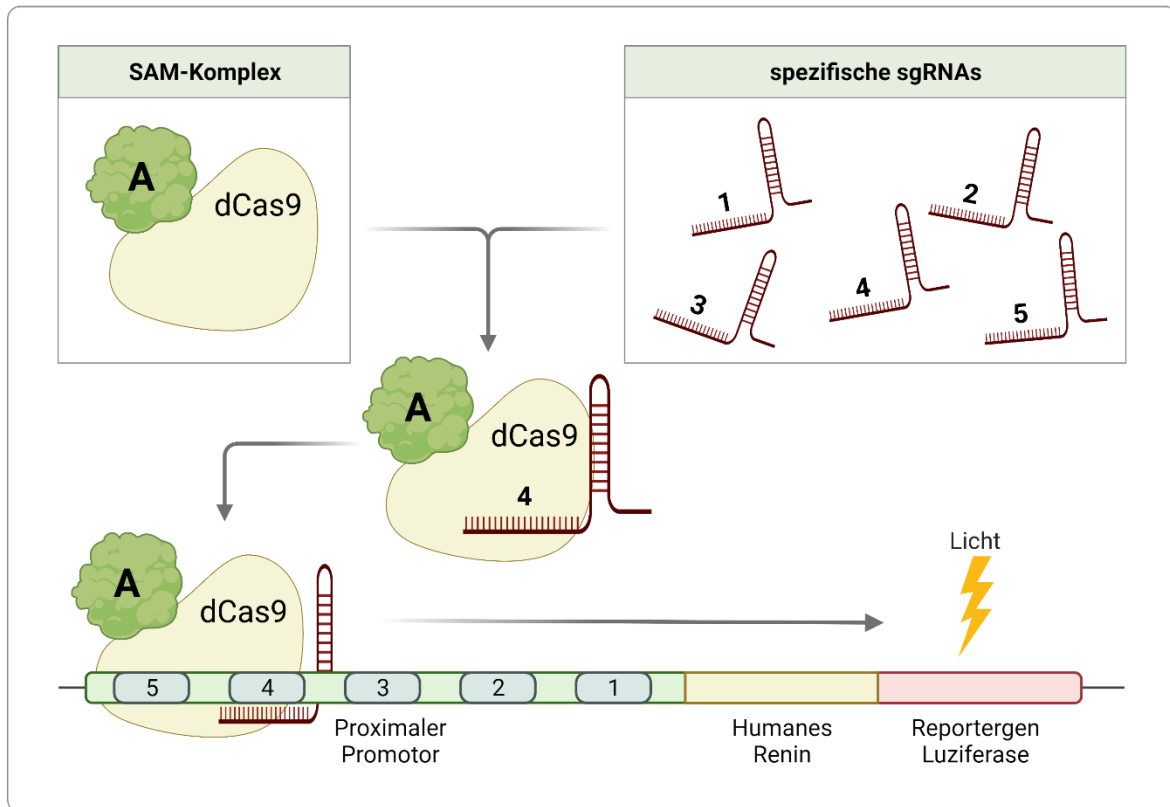


Abbildung 4: Programmierbare spezifische Aktivierung des endogenen REN. Nach der Assemblierung des stabil transfizierten SAM-Komplexes wird dieser über 5 spezifische sgRNAs an komplementäre Promotorregionen des endogenen REN transloziert. An der Promotorregion bewirkt der modifizierte Cas9-Komplex eine ortsspezifische Genaktivierung. Das resultierende Expressionslevel kann durch das stabil integrierte Reporter-gen Luziferase gemessen werden. Erstellt mit *Biorender.com*.

3 Ziele der Arbeit

Die Möglichkeiten des Gene-Editing sind durch stetige Weiterentwicklungen des Cas9-Systems in den letzten Jahren enorm gewachsen. Es existieren bereits kommerziell erhältliche Varianten zum gezielten Gen-Knockout, Aktivieren und Inhibieren von Zielgenen oder multiplexen Gene-Editing-Ansätzen. In dieser Arbeit wird ein modifizierter und aktivierender Cas9-Komplex verwendet, der auf spezifische Promotorregionen eines Zielgens programmiert wurde. Durch kombinatorische Transfektionen und computergestütztes Modellieren der gemessenen Daten soll mit einem neuartigen Ansatz nach möglichen statistischen Interaktionen der untersuchten regulatorischen DNA-Sequenzen gesucht werden.

Folgende Ziele verfolgt die vorliegende Arbeit:

Durch auf spezifische Sequenzen des proximalen Renin-Promotors programmierte aktivierende Cas9-Komplexe (SAM-Komplexe) soll eine Genexpression des endogenen REN bewirkt werden.

Darüber hinaus sollen durch kombinatorische Transfektionen und computergestützte Modellierung der gemessenen Daten mögliche statistische *cis*-regulatorische Interaktionen innerhalb der gewählten genomischen Zielsequenzen untersucht werden. Dieser neuartige kombinierte Ansatz soll darstellen, wie vielfältig und effizient Weiterentwicklungen des Cas9-Systems für molekulargenetische Fragestellungen Anwendung finden können.

Diese Arbeit soll damit hervorheben, dass sich die Einsatzmöglichkeiten der Cas9 durch unseren kombinierten Ansatz zur suffizienten Promotorforschung im endogenen genomischen Kontext erweitern lassen.

4 Publizierte Originalarbeit

Decipher the complexity of *cis*-regulatory regions by a modified Cas9

Steven Kirchner, Stefanie Reuter, Anika Westphal, Ralf Mrowka

PLoS ONE 02. Juli 2020, 15(7): e0235530

DOI: <https://doi.org/10.1371/journal.pone.0235530>

RESEARCH ARTICLE

Decipher the complexity of cis-regulatory regions by a modified Cas9

Steven Kirchner¹*, Stefanie Reuter, Anika Westphal, Ralf Mrowka

Experimental Nephrology Group, KIM III, Universitätsklinikum Jena, Jena, Germany

* steven.kirchner@uni-jena.de

Abstract

Background

Understanding complex mechanisms of human transcriptional regulation remains a major challenge. Classical reporter studies already enabled the discovery of *cis*-regulatory elements within the non-coding DNA; however, the influence of genomic context and potential interactions are still largely unknown. Using a modified Cas9 activation complex we explore the complexity of *renin* transcription in its native genomic context.

Methods

With the help of genomic editing, we stably tagged the native *renin* on chromosome 1 with the *firefly luciferase* and stably integrated a programmable modified Cas9 based *trans*-activation complex (SAM-complex) by lentiviral transduction into human cells. By delivering five specific guide-RNA homologous to specific promoter regions of *renin* we were able to guide this SAM-complex to these regions of interest. We measured gene expression and generated and compared computational models.

Results

SAM complexes induced activation of *renin* in our cells after *renin* specific guide-RNA had been provided. All possible combinations of the five guides were subjected to model analysis in linear models. Quantifying the prediction error and the calculation of an estimator of the relative quality of the statistical models for our given set of data revealed that a model incorporating interactions in the proximal promoter is the superior model for explanation of the data.

Conclusion

By applying our combined experimental and modelling approach we can show that interactions occur within the selected sequences of the proximal *renin* promoter region. This combined approach might potentially be useful to investigate other genomic regions. Our findings may help to better understand the transcriptional regulation of human *renin*.

 OPEN ACCESS

Citation: Kirchner S, Reuter S, Westphal A, Mrowka R (2020) Decipher the complexity of cis-regulatory regions by a modified Cas9. PLoS ONE 15(7): e0235530. <https://doi.org/10.1371/journal.pone.0235530>

Editor: Hans A Kestler, University of Ulm, GERMANY

Received: September 20, 2019

Accepted: June 18, 2020

Published: July 2, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0235530>

Copyright: © 2020 Kirchner et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Grant Support: Funding: Bundesministerium für Bildung und Forschung

(BMBF) grant: FKZ 400 01EK1612B to RM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Transcriptional regulation of genes is one of the key points for gene expression in general. *Cis*-regulatory elements are regions of non-coding DNA that regulate the transcription of neighbouring genes and, among others, serve as a binding site for *trans*-factors. These elements form complex systems. By regulating these complex systems of thousands of genes the morphological maturation of cells as well as their differentiated function are made possible [1]. The understanding of these genomic *cis*-regulatory networks is therefore of great significance. There are examples of mutations of *cis*-regulatory promoter regions leading to severe diseases [2]. An example of a gene and its regulation that has been focused on by numerous research initiatives for many years is the human *renin* (*REN*). *REN* is considered to be a key enzyme in the renin-angiotensin-aldosterone system (RAAS). RAAS is a vital system of the human body, as it maintains plasma sodium concentration, arterial blood pressure and extracellular volume [3]. Abnormal activation of the RAAS can contribute to the development of hypertension, cardiac hypertrophy, and heart failure [4, 5]. According to the WHO about 1.13 billion people worldwide have hypertension which is one of the major causes of premature death [6]. Hence, the understanding of transcriptional regulation is potentially important to understand basic principles of many cardio-vascular diseases. Although the main effector molecule of RAAS is angiotensin II (ANG II), the regulation of ANG II and its precursors is mainly regulated by the expression of *REN* [7].

Important *cis*-regulatory elements could already be identified for *REN*. Most of the information about those elements could be obtained by cell culture experiments using classical reporter assays or experiments with transgenic mice [8, 9]. To perform classical reporter assays, restriction pieces of the DNA region under investigation are cloned in vectors. The sequences are followed by downstream reporter genes such as *green fluorescent protein* (GFP) or *Luciferase*, whose expression levels can be quantified in different ways. After transfection of the vectors in cells, the activity of the promoter region can subsequently be deduced from the expression level of the reporter genes. However, the DNA sequences in question, e.g. the promoter regions are detached from their endogenous context [10–15]. Thus, the experiments must be performed outside the natural environment of the promoter. Furthermore, classical promoter studies assume an independent effect of *cis*-regulatory regions which is reflected in their experimental setup. In conclusion, it is not possible to study complex interactions of individual regulatory elements.

Focusing on human renin, a “*renin enhancer*” (about -12.000 base pairs (bp) to transcription start site (TSS)) as part of an evolutionary conserved region (hRENc region), which is considered to be important for basal *REN* expression [11, 16, 17]; and more proximally, a “*chorion enhancer*” (about -5,500 bp to TSS) was found. However, its relevance is still unclear [18].

The closest known regulatory region to transcriptional initiation is the proximal *renin* promoter, which has been shown to play a significant role in tissue and cell specificity of *REN* expression following experiments on transgenic mice [17, 19]. Research on the As4.1 cell line has identified a proximal promoter region of the murine *Ren-1C*, for which a position in the human *REN* is usually indicated at about -200 to +6 upstream of the transcription start site [8–10]. This region shows distinctive homologies between mice and humans up to a fully conserved TATA box [8]. It's considered to be essential for tissue-specific expression of *REN*, even though the proximal promoter region has reached only a slight enhancement of *REN* expression in murine reporter assays [11, 14]. However, the renal enhancer of *REN* has a lower *trans*-activating capacity compared to the murine renal enhancer [20], which could enhance the influence of the proximal promoter in the *REN*.

Reporter studies and Electrophoretic Mobility-Shift Assays (EMSA) have identified numerous transcription factor binding sites for the proximal promoter region. So, there is e.g. evidence for binding sites that are important for gene regulation of *REN* via the second messenger cAMP [21–23]. Thus, regulatory elements were also found, whose relevance to transcriptional regulation of *REN* seems to be questionable [24–26].

In addition to the *cis*-regulatory elements, the transcription factors (TFs) are essential to enable gene transcription. These proteins bind to DNA and can activate or repress the transcription of genes. There are differences in the way TFs act to regulate gene expression. Some TFs need to assemble with other proteins, others can directly recruit RNA polymerase which then leads to gene transcription [27]. In a current review, a distinction is made between approx. 1600 human TFs, which represent ~8% of all human genes [28]. There are several ways to classify TFs. In general, a division into basal or general TFs and specific TFs is possible. Basal TFs are ubiquitous in all cells and necessary for transcription to occur [29]. During assembly they are part of the preinitiation complex that enables the binding of the RNA polymerase and thus the initiation of the transcription via specific DNA binding sites such as TATA boxes [29, 30]. In contrast, specific TFs only show activity in specific tissues and/or at specific developmental stages. They may bind at specific DNA binding sites (*cis*-regulatory regions), e.g. promoters, enhancers or silencers and are necessary for the regulation of central mechanisms such as cell development or the response to stimuli via signal cascades [31, 32].

A novel experimental approach to explore and thus better understand the complex mechanisms of transcriptional regulation via *cis*-regulatory elements has become possible through further development of the CRISPR-Cas9 system.

This system has become a powerful gene editing engine. It facilitated and expanded the possibilities of loss-of-function and also gain-of-function studies and should even be highly valuable for studying circadian rhythms [33].

Apart from gene editing the guided binding of the Cas9-RNA-DNA complex can be used for other purposes. This includes gene activation, gene silencing and gene labelling, for example with fluorescent proteins. In this study we used a modified Cas9 system that is able to activate genes. Therefore, a non-cutting Cas9, that was fused to an activation transcription factor complex was chosen. We used this system to explore the *cis*-regulatory importance of genomic DNA, in this study for the proximal *renin* promoter.

In detail, we used the Synergistic activation mediator complex (SAM) [34] (Fig 1).

In 2015, Konermann et al. presented a modified Cas9 (dCas9) that was coupled to a complex of different transcription factors, allowing them to establish a highly efficient guide-RNA-directed *trans*-activating complex referred to as SAM complex [34–38]. The dCas9 lost its originally cutting function through mutations in its catalytic domain, making it an efficient DNA-binding protein [39]. Additionally, it is coupled to the transcription factor VP64 [40, 41]. Regions of the original tracr-RNA were also modified to bind a complex of trans-activating domains of p65 and HSF1 [34]. The dCas9 can be programmed by a single guide-RNA [39]. Regarding the design of the guide-RNA it is possible to target any sequence in the genome, given that a PAM site (NGG) follows the 20 nt sequence. In order to make the interactions specific, the sequence should be unique for the given species. This is incorporated by the bioinformatic tools for designing potential target sequences.

The development of this new gene editing tool enables the study of genes and their regulatory elements in their endogenous context and in their natural environment. This kind of research is not possible by classical reporter studies (as described above). In addition, this specific and constant *trans*-factor complex allows the isolated study of the *cis*-regulatory regions since the same trans-factors always bind to the regulatory elements.

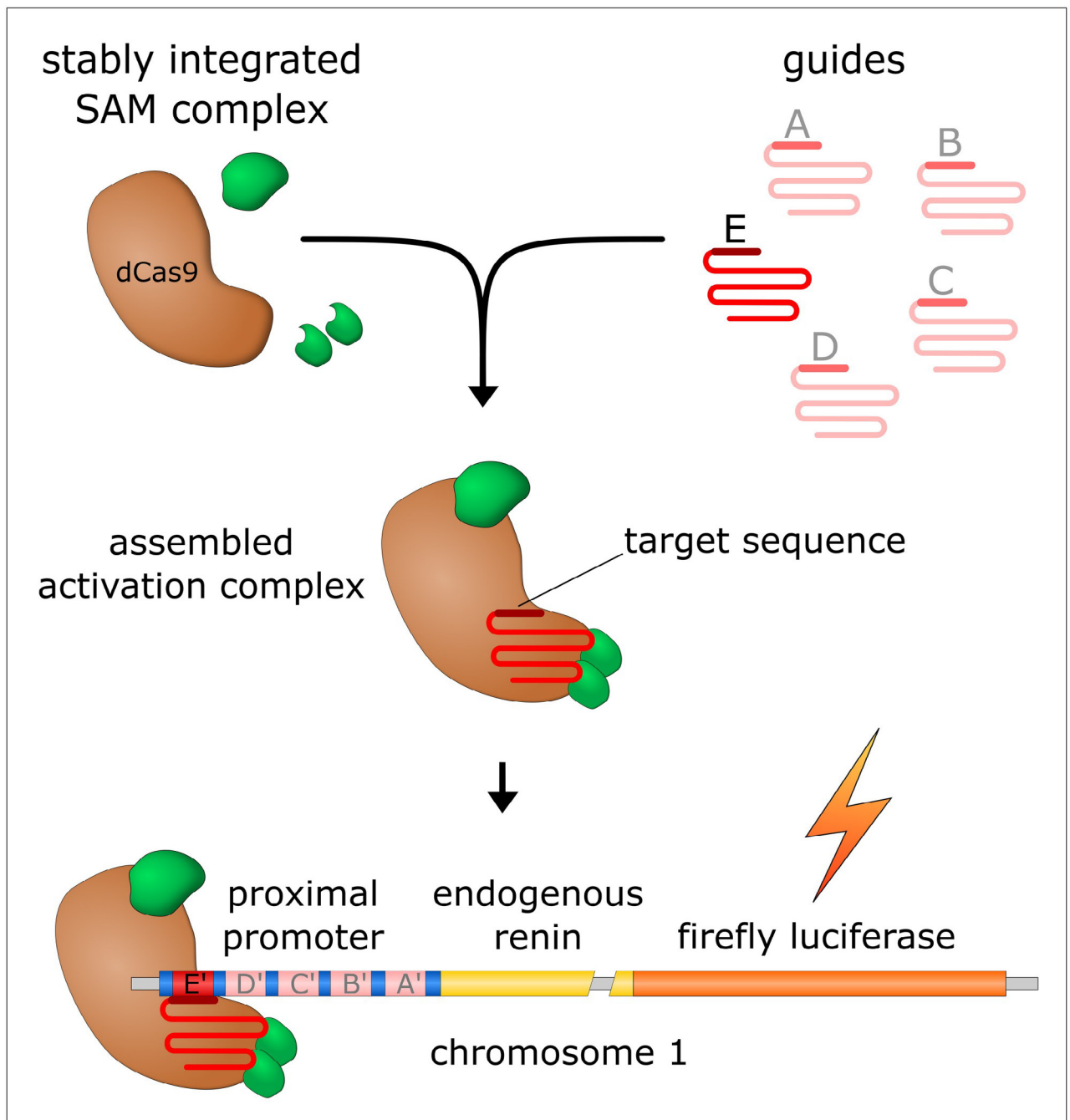


Fig 1. Programmable site-specific activation of REN with designed guides A-E. After assembling of the activation complex (SAM) the attaching guide (shown here by the example of guide E) leads the complex to its specific region of the proximal renin promoter (E'). Thus, the guides enable site specific programming of the SAM complex. The resulting gene activation can be measured by a tagged firefly luciferase.

<https://doi.org/10.1371/journal.pone.0235530.g001>

In order to obtain more understanding about the complex transcriptional network, mathematical modelling of experimental data was performed. For example, conforming principles have been identified, under which individual sections of a *cis*-regulatory element approximately 14 kb upstream of *REN* interact with each other [42]. The possibility of drawing

conclusions concerning the dynamics of *cis*-regulatory elements by using mathematical modelling was already shown in 2003 in a study on the *Escherichia coli* bacterium [43]. In addition, by mathematical modelling in 1998, complex processes at an important promoter region of *Endo1* had been revealed, which are of great importance for the embryonic development of sea urchins [1]. Thus, mathematical modelling can help to deepen the understanding of the complex transcriptional network.

When statistically examining factors that influence a variable, it is of general scientific interest to find out whether these factors act independently or interdependently. In our opinion, this question of complex interactions was given too little importance in the research on *cis*-regulatory regions of *REN*, especially on the proximal promoter [9, 26]. A study that attempted to address this problem with a modelling approach emerged in 2007 [42]. However, even here the experiments were plasmid-based and therefore outside of the endogenous context. In addition, a regulatory region of *REN* was in the focus, which is approximately 14.000 base pairs upstream of the start of transcription.

With this study on the one hand we want to show a novel approach to investigate interactions of *cis*-regulatory regions in an endogenous context, and on the other hand raise awareness of how important these interactions are for the understanding of the complex transcriptional network.

To study possible complex interactions within the proximal promoter region of *REN* we applied a novel combined approach. Firstly, this approach consists of combinatorial transfections from five selected guide-RNAs that translocate the SAM-complex to a specific region of the endogenous proximal promoter. The resulting expression levels of *REN* through the different combinations of the targeted promoter regions can be quantified by luciferase activity. Secondly, we generated and fitted two different mathematical models to our experimental data. When modelling experimental data, the principle of simplicity should always be considered. The simplest assumption would be that the regions examined influence the expression of *REN* completely independent of each other. This is reflected in our first model (sum model). The sum model describes an independent relationship of the promoter sequences we examined to their influence on the activation of *REN*. In this model each individual region is analysed in respect to its influence on gene expression. This approach is comparable to the knowledge that classical promoter studies can provide, since, as described above, these do not allow the possibility of studying complex interactions. However, an important scientific question in modelling is whether a statistical interaction exists [44]. In order to address this issue, a second mathematical model was generated (interaction model). This model represents a more complex assumption of the conditions in the region of the proximal promoter. The interaction model moreover allows interactions between the selected promoter regions to explain the *REN* activation. Regarding modelling we used the multiple linear regression model to fit the linear parameters, which is a standard statistical method [45]. In order to check which of the generated models can explain the measured data best, the respective absolute prediction error was calculated. Following the principle of maximum parsimony in modelling, the respective Akaike information criterion (AIC) was calculated for further model judging [46]. The objective of this study was to examine potential interactions of sequences within the proximal *renin* promoter. Through combinatorial transfections of specific guide-RNAs using the SAM complex and computational modelling of the measured data we want to show a novel combined approach that helps to enlighten the complexity of *cis*-regulatory regions in an endogenous context. We want to show that transcriptional regulation is even more complex than already known and that complex interactions should be considered when assessing the importance of specific *cis*-regulatory elements. The region of the proximal *renin* promoter is in the focus of this work. Thus, this study may help to better understand the transcriptional regulation of the

key enzyme of RAAS. Furthermore, in our opinion this approach is also suitable for evaluating interactions and dynamics of other *cis*-regulatory regions.

Material and methods

Cell line

Human embryonic kidney cells (HEK293) were cultured in T75 cell culture flasks in high-glucose DMEM (Thermo) supplemented with 10% fetal bovine serum (FBS) (Biochrom) and 1% penicillin/streptomycin (Biochrom) at 37°C and 5% CO₂ in a humidified incubator. The medium was changed every 3–4 days. At approximately 90% confluence, the cells were passaged at a 1:5 dilution and seeded in a new T75 flask.

Generating SAM-complex expressing cells by lentiviral transduction

24 hours prior to transfection, 1.4×10^6 cells were seeded into cell culture dishes (60 mm diameter) in 3 ml high-glucose DMEM supplemented with 10% fetal bovine serum. Transfection was performed at a confluency of 80–90%. Per dish 2 µg of either lenti dCAS-VP64_Blast (Addgene #61425) or lenti MS2-P65-HSF1_Hygro (Addgene #61426) [34], and the lentiviral packaging plasmids 2 µg pMDLg/PRRE (Addgene #12251) [47], 0.4 µg pRSVrev (Addgene #12253) [47] and pMD2.G (Addgene #12259) were diluted in BES-buffer (50 mM BES, 280 mM NaCl, 1.5 mM Na₂HPO₄ x 2 H₂O, pH 6.95). CaCl₂ was added to a final concentration of 0.125 M, the plasmids were incubated for 10 minutes at room temperature and added to the cells. Post transfection (12–16 hours later), the medium was changed to fresh DMEM with 10% FBS. Lentivirus-containing supernatant was harvested 48 hours after transfection. The supernatant was filtered sterile with a 0.45 µm cellulose acetate filter (Roth) and stored at -80°C.

Approximately 50,000 HEK293 cells were seeded per well of a 24-well plate in 500 µl high-glucose DMEM supplemented with 10% fetal bovine serum. The medium was changed to 500 µl of lentivirus supernatant after 12 and 18 hours. First, cells were generated expressing dCas9-VP64; afterwards, MS2-P65-HSF1 were selected with 10 µgml⁻¹ blasticidin (Invivogen) and 100 µgml⁻¹ hygromycin B (Invitrogen). Cells are now referred to as HEK dCas9-SAM.

Tagging of *REN*

REN was tagged in frame with the gene for firefly luciferase (Fig 1) and the G418 resistance using a Cas9 and a specially designed guide-RNA GGCTTCGCCTTGGCCCGCTG. The G418 resistance cassette was not used in this study but was cloned for potential further experiments with those cells or plasmids. The guide design was performed by the guide design tool “crispr.mit.edu” and cloned according to the manufacturer’s instructions into pGuide-it-tdTomato (Clontech). The stop codon was deleted. Luciferase and G418 resistance were linked to the *REN* via T2A and P2A sequences. The flanking homology arms were amplified from genomic DNA of HEK293 cells by PCR. The following primers were used for the reaction:

Renin-left-arm-forward-NotI GTACGCGGCCGCGCTCACCAGCGCGACTATGTAT,
Renin-left-arm-reverse-PaCI AGCTTTAATTAAGCGGGCCAAGGCGAAGCCAATGCG,
Renin-right-arm-forward-AarI acgtccacctgcgtgcttaaggccctctgccaccag
gcag,
Renin-right-arm-reverse-AscI AGCTGGCGCGCCGACCCAAGTCAGACGGGCTGGGTTC.

The homology arm PCR products were integrated into MV-PGK-Puro-TK vector (Transposagen), which was modified by integrating a cassette via NheI and HindIII digestion. The cassette contains a P2A-linker-firefly luciferase-T2A linker-G418 resistance-PGK promoter-

puromycin resistance-T2A linker-thymidine kinase (S1 Fig) and is flanked by restriction sites NotI and PacI for left homology arm integration and AarI and AscI for right homology arm integration.

For transfection of the plasmids 1.2×10^6 HEK dCas9-SAM were seeded into a well of a 6-well plate and transfected 12–16 hours later at a confluency of 80–90%. 1.5 μg of *REN*-tagging donor plasmid and 1.5 μg of pGuide-it-tdTomato vector with the integrated *REN*-guide were diluted in 100 μl Opti-MEM™ (Thermo) and mixed with 100 μl Opti-MEM™ including 12 μl Lipofectamine® 2000 (Thermo). After 5 minutes incubation at room temperature, the plasmid was added dropwise to the cells. Cells were selected with $2 \mu\text{gml}^{-1}$ puromycin. The cells are now referred to as HEK dCas9-SAM_Renin-luciferase.

***REN*-guides: Design and cloning**

The guides used for this work were designed using the online “SAM sgRNA design tool” [48] in the beginning of 2018. The guides are sorted in order of specificity (highest to lowest) based on a method described by Hsu et al in 2014 [49]. We have chosen the top 5 hits for human *REN* for our experimental approach. The respective 20-base-long promoter sections A’ to E’ are approximately evenly distributed from 60 to 159 base pairs upstream of the transcription start of *REN* (Table 1).

The DNA oligos were designed with overhangs for BbsI and cloned into backbone plasmid sgRNA (MS2) ordered from Addgene (#61424) [34] following the depositor’s advice. Thus, guides A-E were created.

Transfection of the guides

Approximately 50,000 HEK_dCas9-SAM_Renin-luciferase cells were seeded per well in 96-well plates in 140 μl each of high-glucose DMEM containing 10% FBS, 1% Pen/Strep, 1 mM L-Glutamin (Biochrom), 10 mM HEPES (Thermo) and 250 μM Luciferin D (Promega). The cells were incubated overnight in a cabinet humidified incubator at 37°C and 5% CO₂. After 16 hours, the transfection with the guides was carried out at a confluence of about 80%. The transfection was performed according to the protocol of Lipofectamine® 2000 reagent. 0.5 μl Lipofectamine® 2000 and a constant amount of guide-DNA per guide and per well were added in a total volume of 10 μl Opti-MEM™.

To study whether interactions between the investigated promoter sequences occur all possible combinations of the five guides A-E were transfected into the HEK_dCas9-SAM_Renin-luciferase cells. In each case 30 ng of guide-DNA per guide and per well were used, which means the total amount of DNA per well varied from 0 ng– 150 ng. Each possible combination was transfected with a sample size of $n = 6$. We have not performed a concentration dependent

Table 1. Chosen guides A to E for combinatorial transfections.

Guide	Sequence (5’-3’)	Distance to TSS
A	TCTGCCCTGATTTATTACCC	-60 bis -80
B	ACTGCCCTGCCATCTACCCC	-81 bis -101
C	TCTCACTGCCGGACAGAGCT	-111 bis -131
D	GTGAAGGGTACCCAGGTTTC	-137 bis -157
E	CAGCCCTCTGCTCCCCATC	-159 bis -179

Here the five chosen guides A to E with their respective distance to TSS are shown that are used for the experimental work. For guide-design we have used the online “SAM sgRNA design tool” in the beginning of 2018.

<https://doi.org/10.1371/journal.pone.0235530.t001>

Table 2. Pattern of the combinatorial transfections of the designed guides A-E.

	A	B	C	D	E
1					
2					X
3				X	
4				X	X
5			X		
6			X		X
7			X	X	
8			X	X	X
9		X			
10		X			X
11		X		X	
12		X		X	X
13		X	X		
14		X	X		X
15		X	X	X	
16		X	X	X	X
17	X				
18	X				X
19	X			X	
20	X			X	X
21	X		X		
22	X		X		X
23	X		X	X	
24	X		X	X	X
25	X	X			
26	X	X			X
27	X	X		X	
28	X	X		X	X
29	X	X	X		
30	X	X	X		X
31	X	X	X	X	
32	X	X	X	X	X

All performed combinatorial transfections of the designed guides A-E. To sort the combinations, a binary code was applied so that each combination could be uniquely assigned.

<https://doi.org/10.1371/journal.pone.0235530.t002>

analysis for each individual guide in this study. From other studies of other genes in our lab we have used similar concentrations for transfection of guide-RNAs that showed biologically relevant expression of RNA and protein. For this reason, we always used the concentration of 30 ng guide-RNA per well for transfection in the combinatorial setup.

In order to sort the combinations, a binary code was applied so that each combination could be uniquely assigned (Table 2).

Luciferase-reporter-assay

The luciferase assays were performed with HEK_dCas9-SAM_Renin-luciferase cells. The cells were transfected with all possible combinations of the five guides as described above. The sample size of each combination was $n = 6$. Immediately after transfection, the 96-well plates were

sealed and the luciferase activity was measured over time at 37°C in the TopCount[®] NXT (Perkin Elmer, Waltham, USA). Luciferase activity is expressed in counts per second (cps).

Modelling of received data and statistics

In order to find conforming patterns by which the promoter regions influence the expression of *REN*, models were generated with "R". The first model (sum model) describes an independent relationship of the five selected promoter sequences A'-E' with respect to their influence on the activation of *REN*. This model was chosen because it describes the simplest possible form of framework of the regulatory elements. In this model each individual region is analysed concerning its influence on gene expression. For the independent sum model, the measured activity y_i can be represented by formula 1:

$$y_i = \epsilon_i + \beta_0 + A_i\beta_1 + B_i\beta_2 + C_i\beta_3 + D_i\beta_4 + E_i\beta_5$$

where β_0 is the offset, $\beta_1 \dots \beta_5$ are the linear coefficients, $A_i \dots E_i$ are {0,1} depending on presence in experiment i and ϵ_i is the error.

Furthermore, a second model has been generated (interaction model) that additionally allows interactions between the promoter sequences in order to explain *REN* activation. For the interaction model the measured activity y_i for all the 32 possible combinations can be represented by formula 2:

$$\begin{aligned} y_i = & \epsilon_i + \beta_0 + \beta_1 A_i + \beta_2 B_i + \beta_3 C_i + \beta_4 D_i + \beta_5 E_i + \beta_6 A_i B_i + \beta_7 A_i C_i + \beta_8 B_i C_i + \beta_9 A_i D_i \\ & + \beta_{10} B_i D_i + \beta_{11} C_i D_i + \beta_{12} A_i E_i + \beta_{13} B_i E_i + \beta_{14} C_i E_i + \beta_{15} D_i E_i + \beta_{16} A_i B_i C_i + \beta_{17} A_i B_i D_i \\ & + \beta_{18} A_i C_i D_i + \beta_{19} B_i C_i D_i + \beta_{20} A_i B_i E_i + \beta_{21} A_i C_i E_i + \beta_{22} B_i C_i E_i + \beta_{23} A_i D_i E_i + \beta_{24} B_i D_i E_i \\ & + \beta_{25} C_i D_i E_i + \beta_{26} A_i B_i C_i D_i + \beta_{27} A_i B_i C_i E_i + \beta_{28} A_i B_i D_i E_i + \beta_{29} A_i C_i D_i E_i + \beta_{30} B_i C_i D_i E_i \\ & + \beta_{31} A_i B_i C_i D_i E_i \end{aligned}$$

where β_0 is the offset, $\beta_1 \dots \beta_{31}$ are the linear coefficients, $A_i \dots E_i$ are {0,1} depending on presence in experiment i and ϵ_i is the error.

These models have been fitted with the measured data. For fitting of the models represented in formula 1 and 2 the fitting functionality of the *lm* function of the built-in *stats* package of R version 1.1.423 was used [50]. The idea behind this is to minimize the error in prediction of y by optimizing the linear factors β for the experimental data. The *lm* function of the R package *stats* computes the linear factors β that fit the input variables A to E according to the proposed model including all the statistics of the fitted parameters. The complete data that was used including the R-script can be found in the supplementary material (S1 Table, S1 Script). Data was analysed at 60 hours after transfection.

For the evaluation of the models we used the obtained coefficients of the respective model and put the coefficients in formula 1 and 2, respectively. For statistical analysis of the coefficients itself the built-in p-value calculation of the multiple linear regression of the *lm* function of R was used. For comparison of the models we calculated the respective absolute prediction errors of the two models with the built-in *predict* function of the *stats* package of R [50]. After fitting of the linear factors β , the error ϵ_i could be calculated for the experiments i through conversion of the respective formula. For statistical analysis of the respective absolute prediction errors we used the independent 2-group Mann-Whitney U Test [50]. To compare the generated models following the principle of maximum parsimony in modelling, the AIC (Akaike Information Criterion) of each model were calculated for further model judging. The AIC was calculated in the R environment with the AIC function of the *stats* package [50].

Random controls were generated by the built-in *sample* function of R [50]. For this purpose, a dataset was created by randomizing the values of the respective variables A to E. This

randomized data was then fitted to model 1 and 2. No statistically significant p values for the coefficients were obtained in the random case (S1 Script).

The entire statistical evaluation including model fitting and model evaluation by 2-group Mann-Whitney U Test of the respective prediction errors and calculation of the AIC did not need any further parameters, apart from the experimental parameters such as incubation temperature or evaluation at about 60 hours after transfection.

Results

Our aim was to explore possible interactions within the proximal *renin* promoter. Therefore, HEK cells were transfected with all combinations of guides that represent the selected promoter sequences A' to E'. The luciferase signal of transfected HEK dCas9-SAM_Renin-luciferase cells, that indicates renin activation, was measured over time in the TopCount[®] NXT. With a modified Venn diagram the activation levels of *REN* via all of the 32 resulting possible combinations of the guides were visualised (Fig 2). The 32 individual areas result from overlaps of the five main areas A-E, which represent the guides. The size of the respective area is of no significance considering the results. The circle area surrounding the entire individual areas reflects the case in which none of the guides were used.

Each combination of guide RNAs has its own binary code assigned. Each transfection of HEK dCas9-SAM_Renin-luciferase was done according to the same protocol (see [Material and methods](#)).

After approximately 24 hours of transfection of the combinations of RNA guides to HEK dCas9-SAM_Renin-luciferase, expression of the *REN* could already be detected via the increase in the luciferase activity. Over time, the expression levels differed for the individual combinations of the promoter sequences (Fig 3). In Fig 3 expression levels of the 32 possible combinations are shown at three different time points after transfection using the modified Venn diagram as described above (Fig 2). Respective activation levels of *REN* were expressed in a colour ramp that rose in ascending order from white to yellow, further to red and up to blue. For analysis, the values of expression levels were scaled logarithmically. Otherwise, since

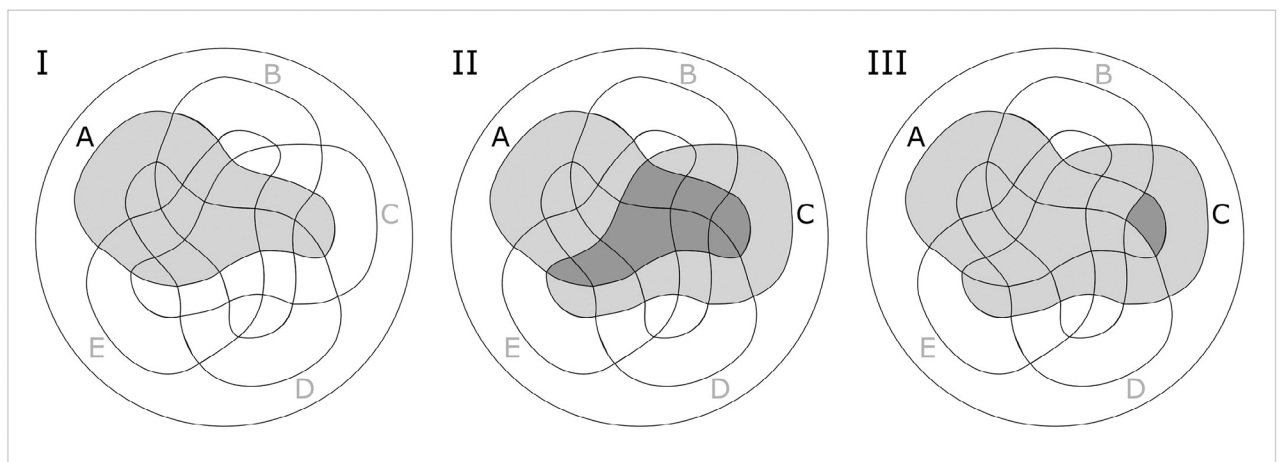


Fig 2. Visualisation of the 32 possible combinations of guides. This modified Venn diagram represents each of the 32 possible combinations of guides as a surface. The individual areas result from overlaps of the five main areas A-E, which represent the guides. The size of the respective area is of no significance considering the results. The circle area surrounding the entire individual areas reflects the case in which none of the guides were used. Three examples are shown here. (I) All combinations including A are filled light grey; (II) All combinations including A and C are filled in dark grey; (III) Combination consisting solely of A and C are marked dark grey.

<https://doi.org/10.1371/journal.pone.0235530.g002>

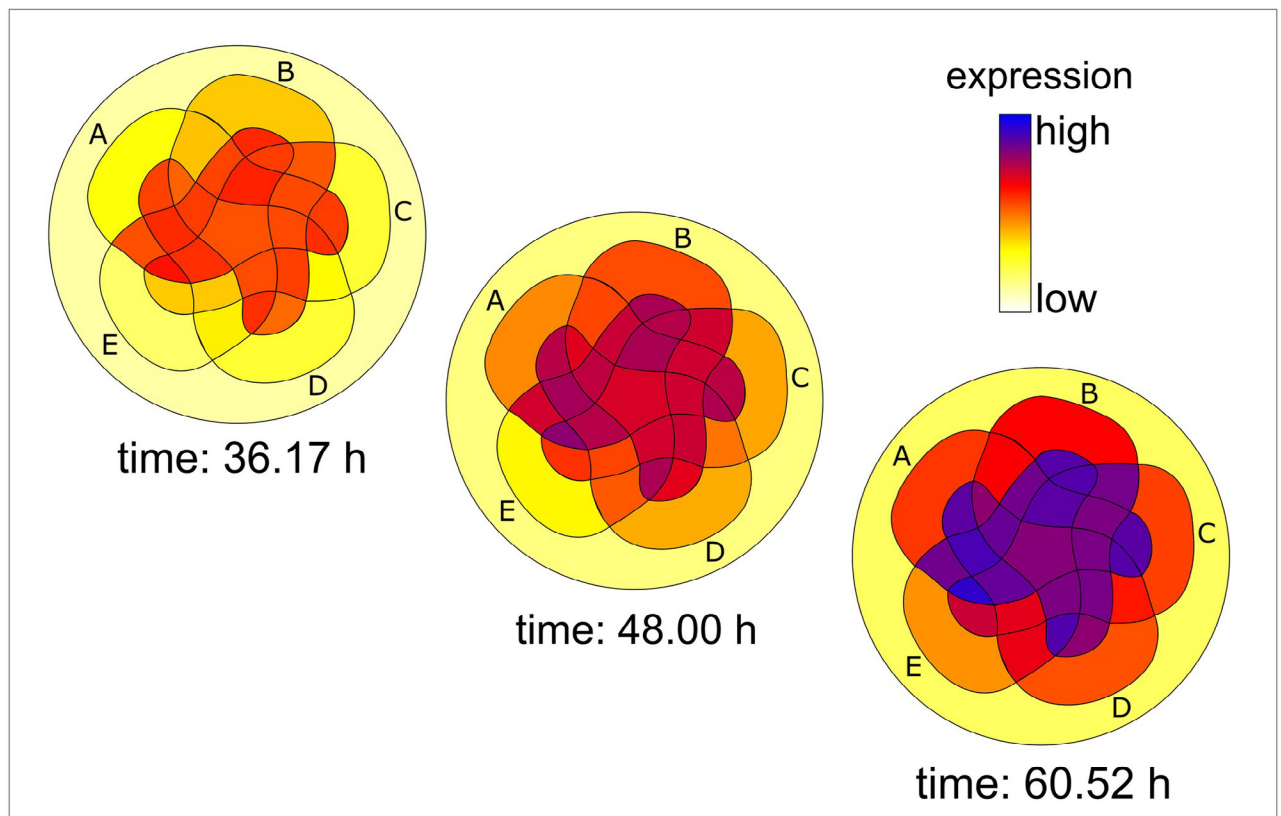


Fig 3. Visualisation of the luminescence shows the activation of renin via cis-regulatory regions for each individual combination of guides over time. The expression levels of the 32 possible combinations are shown here at three different time points after transfection of HEK dCas9-SAM_Renin-luciferase. By translating the luminescence values into a colour scale, the corresponding colour intensity resulted for each RNA guide combination. A time-dependent increase in the expression levels of the 32 guide combinations becomes obvious. Furthermore, individual guides and the 32 possible combinations caused different expression levels of REN. The axis "Expression" is scaled logarithmically (each area $n = 6$).

<https://doi.org/10.1371/journal.pone.0235530.g003>

these values were widely divergent, only a few surfaces would have been coloured in a linear scale, whereas most were white to slightly yellow. On the one hand the individual guides itself caused different levels of renin expression. On the other hand, the different combinations achieved different renin activation (Fig 3 and S1 Table). The combinations of the two guides C—B or A—E caused a renin activation that was higher than the simple summation of the respective individual activation levels. In contrast, combinations of the two guides A—B or also D—C have a less increasing effect on the renin activation.

To further investigate the potential interactions of the chosen proximal promoter sequences we generated two mathematical models in R and fitted them to the measured data described above.

The first model was the sum model. It describes an independent relationship of the five selected promoter sequences with respect to their influence on activation of *REN*. We chose this model because it assumes the simplest possible correlation between the promoter regions. Modelling of the data revealed that all sequences studied significantly affected the activation of *REN* (S2 Fig). However, the assessed values of the 32 combinations often show deviations to the predictions of the sum model (Fig 4). The measured luminescence values were smaller than the values predicted by the sum model after transfection of all combinations in which neither guide A nor B were involved. As soon as either A or B was present in the combinations,

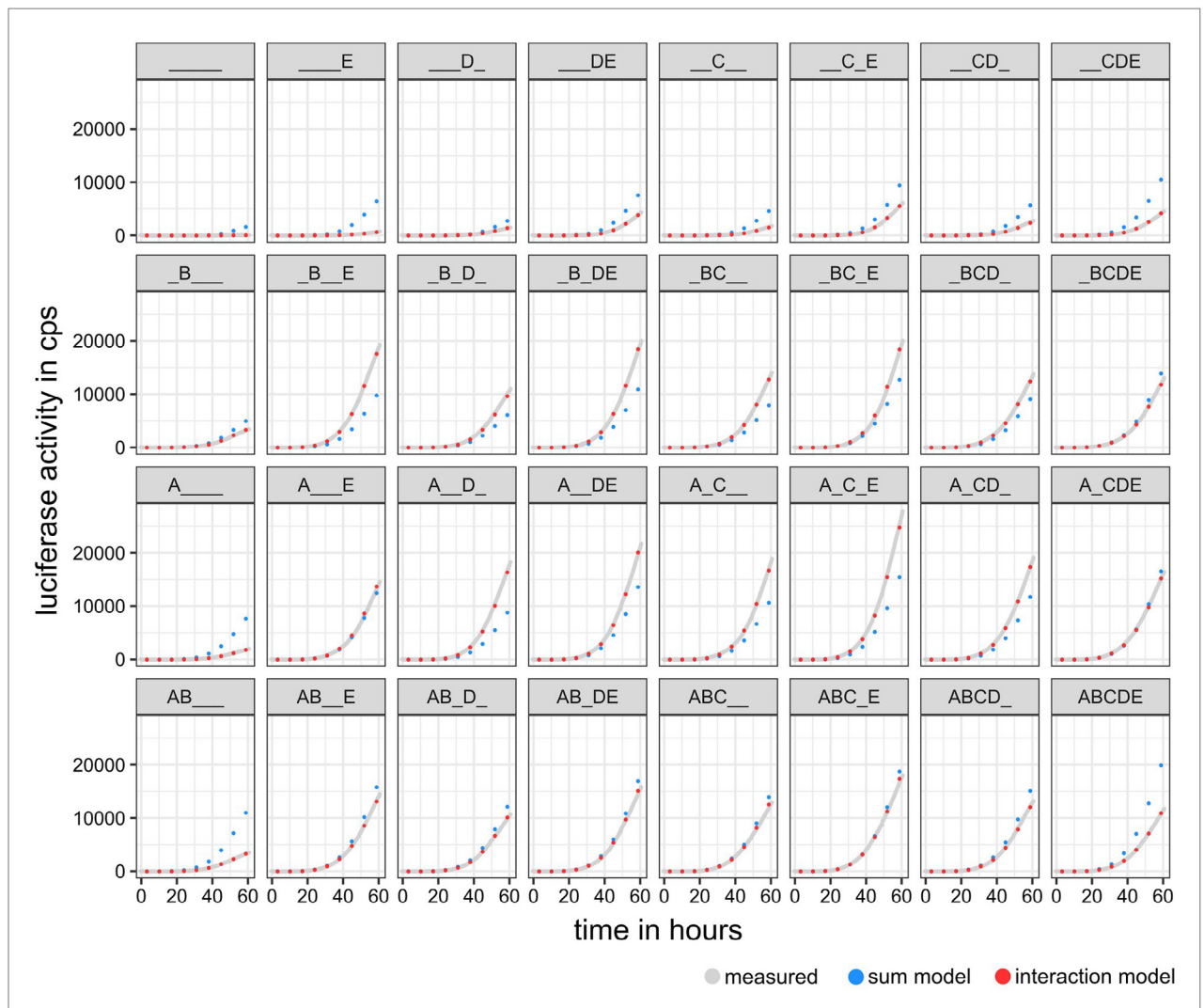


Fig 4. Luciferase measurements vs. predicted values of the models. The measured values (grey) of each combination are plotted with the respective prediction of the **sum model** (blue) and the **interaction model** (red). The measured values show deviations from the predicted values of the **sum model**. The prediction of the **interaction model** better matches the measured expression levels of the 32 different combinations. The luciferase activity is expressed in counts per second (cps). The measured values are presented by their medians. The curves of the measured values result from >3,000 measuring points per combination (32 combinations, each n = 6).

<https://doi.org/10.1371/journal.pone.0235530.g004>

the luciferase activities were higher or at least at the level of the prediction. Exceptions were combinations in which A or B stimulated the expression of the *REN* in each case by solely driving the promoter sequence. Interestingly, when guides A and B were both present in the combinations, values in the range of the prediction or below were measured.

The second investigated model, the interaction model, allows interactions existing between the individual promoter regions in order to explain the measured data. The deviations of the predicted values of the interaction model are smaller than the predicted values of the sum model (Fig 4). Thus, the interaction model seems to better describe the measured expression levels of the 32 different guide combinations.

In the next step we compared the two models. By calculating differences of the measured values against the predicted values of the two models, we were able to visualise the degree of

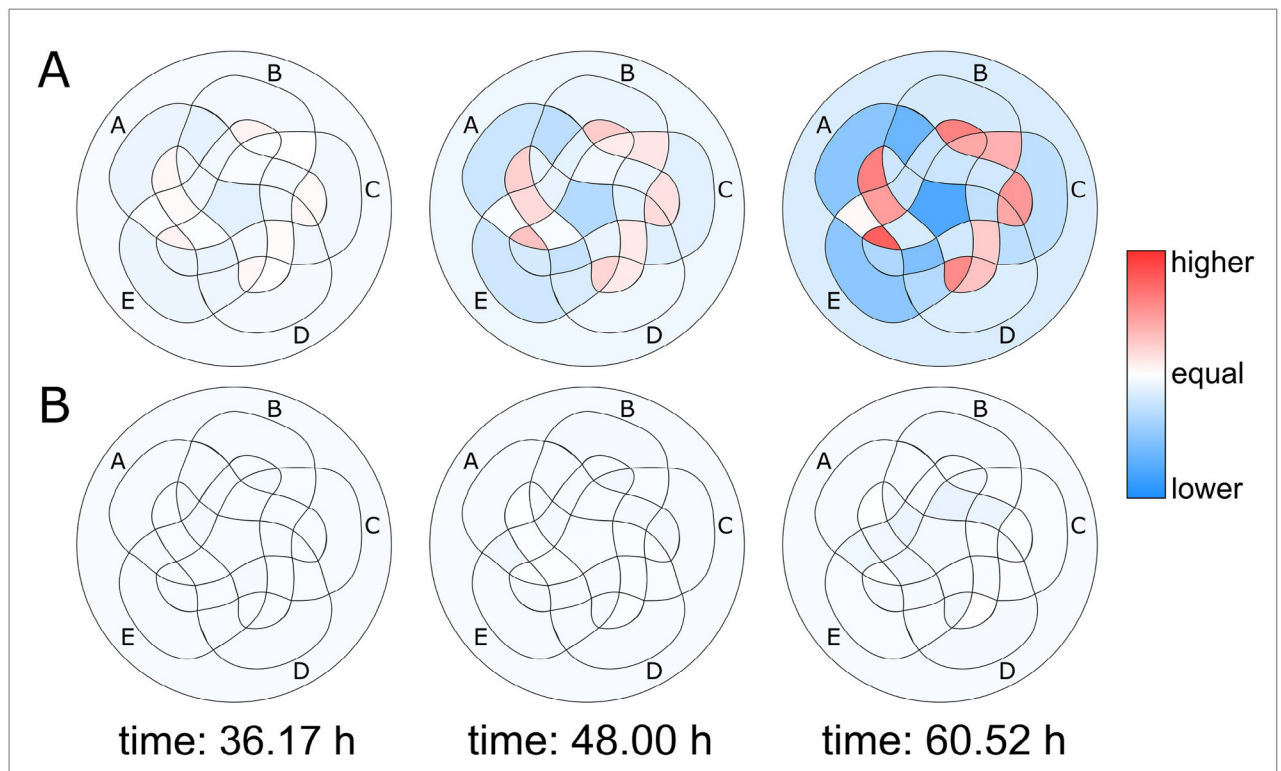


Fig 5. Differences between measured values and predicted values of the models at three different time points. The differences of measured values and predicted values of the 32 possible combinations assuming the two generated models are shown here for three different time points after transfection. Areas turn red when the differences between the measured luciferase activity and the predicted values of the respective model are positive and turn blue when the differences are negative. (A) In the sum model over time, the colouration of the combinations increases. (B) For the interaction model almost no colouration can be recognised over time. The values of the differences are linearly scaled (each area $n = 6$).

<https://doi.org/10.1371/journal.pone.0235530.g005>

similarity at three chosen time points at Fig 5. It becomes obvious that an independent relationship of the promoter sequences as assumed in the sum model cannot adequately explain *REN* expression.

The prediction of the sum model and the measured values differ, as seen at Figs 4 and 5. This motivated us to further analyse the better fitting interaction model. The influences of the linear single factors A to E are no longer significant (with exception of B) when interaction of the sequences is allowed (Fig 6). The activities of the combinations A—C, A—D, A—E, B—C, B—D, B—E significantly outweighed those linear single factors. These results underline that interactions within the promoter regions are essential for the extent of the assessed gene activity.

Further, for both models, the absolute error (absolute difference between measured and predicted model values) was calculated. The error was significantly smaller for the interaction model than for the sum model ($p < 2.2 \cdot 10^{-16}$ according to 2-group Mann-Whitney U Test; Fig 7). For further comparison, the AIC was calculated for both models. This AIC calculation returned a value of 3881.4 for the sum model and a value of 3609.6 for the interaction model, indicating the interaction model to be statistically more adequate. In order to check whether these results could have been a coincidence, all measurement data of the individual promoter sequences were randomised. After applying the generated models to this randomised data, no significant effects on the expression of *REN* were found for the selected promoter sequences (S1 Script).

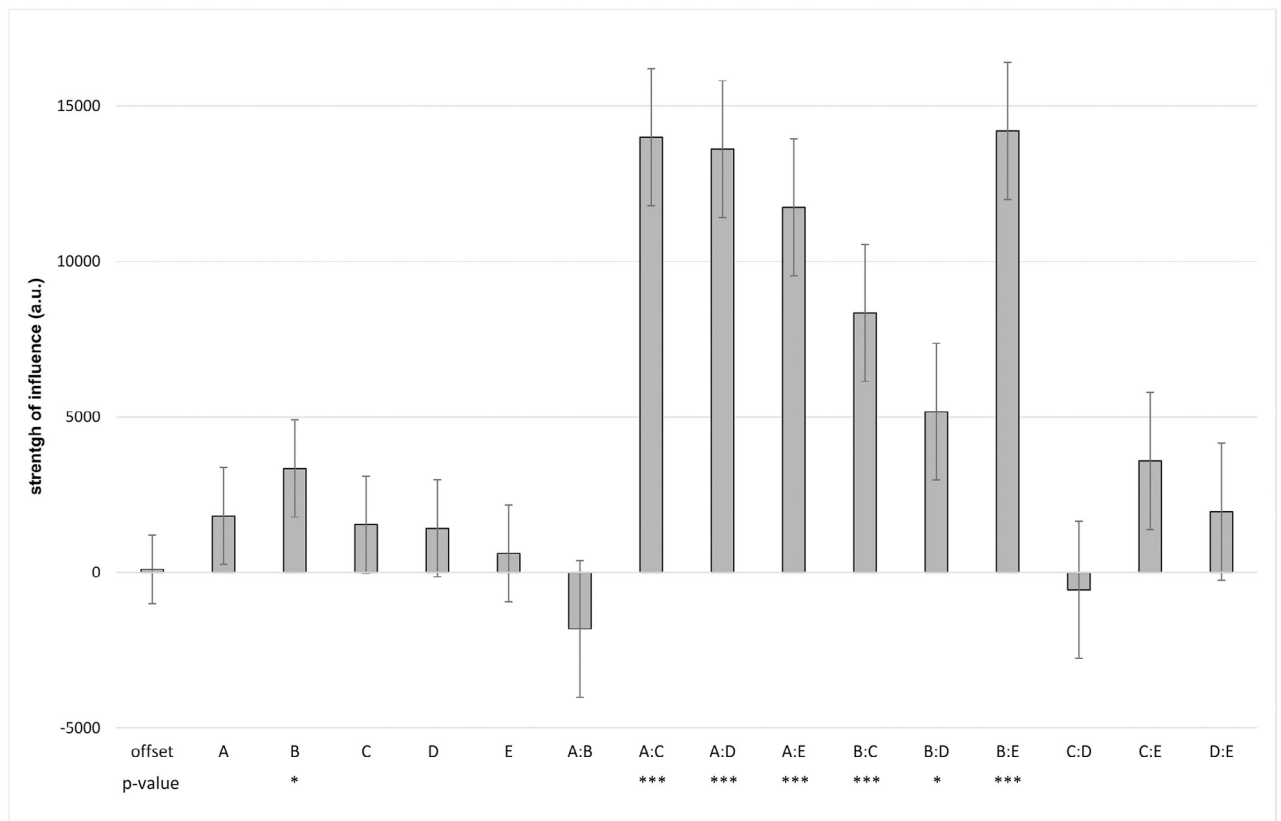


Fig 6. Influence of the promoter regions (A to E) and their potential interactions on gene activity of REN. The strength of influence of certain combinations of promoter sequences outweighs the influences of the singular linear individual factors. Influences of linear single factors A-E are no longer significant, with exception of B. Significances resulted from the multiple linear regression modelling with the lm function of the build in stats package in R. Strength of influence is represented as arbitrary unit (a.u.). (mean and SD, n = 6, p-values: * < 0.05, *** < 0.0005).

<https://doi.org/10.1371/journal.pone.0235530.g006>

Discussion

In a combined approach, consisting of combinatorial transfections and computational modelling of the measured data we analysed the role of *cis*-regulatory elements in the proximal promoter of human *REN*. According to research on murine *Ren-1C* via reporter assays, the proximal *renin* promoter is described to have low activation potential for the *REN* [11, 14]. However, compared to the murine renal enhancer the gene activating capacity of human renal enhancer is lower [20], which could enhance the influence of the proximal promoter in the *REN*. Furthermore, a 99% decrease in the transcriptional activity of renin was also found when the region of the proximal promoter was deleted in reporter assays performed in As4.1 cells [26].

However, in our experimental setup the *REN* expression could be triggered by the targeted translocation of the SAM complexes to all five promoter sequences (Fig 3, S2 Fig).

The combinatorial transfections were performed with a constant amount of guide-DNA per guide. In a previously unpublished study, we have determined renin levels after activation of this promoter region. We were able to show that the catalytic activity of renin was present in the supernatant of the HEK cells. This shows that we operate in a biological sensible range of renin expression. Further it shows that secreted renin (biologically relevant for blood pressure regulation) was produced in our cells.

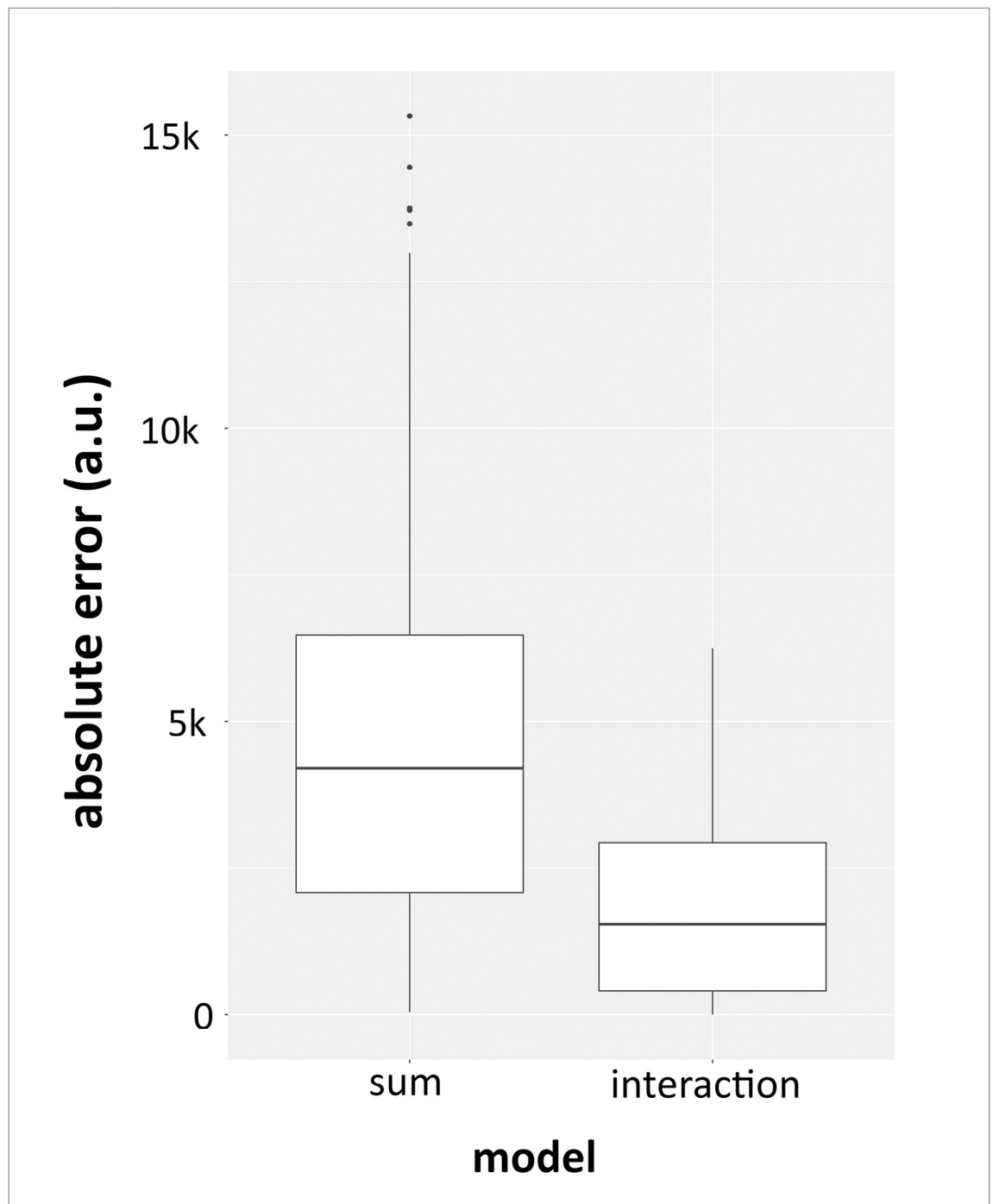


Fig 7. Distribution of the respective absolute error of the sum model vs. the interaction model. Here, the two models are compared by calculation of their respective absolute prediction errors. The distribution of the absolute errors is shown as a box plot. The absolute prediction error of the interaction model is significantly smaller than that of the sum model ($p < 2.2 \cdot 10^{-16}$ according to 2-group Mann-Whitney U Test). Absolute error is represented as arbitrary unit (a.u.).

<https://doi.org/10.1371/journal.pone.0235530.g007>

Our new approach includes computational modelling of the measured data besides the combinatorial transfections. This enabled deeper insights into *cis*-regulatory interactions. Independent linear models are a method to describe multifactorial influences on an output variable (in this case renin transcription). In the simplest case this would be a two-factor setup, e.g. whether a drug (factor 1) acts dependent on sex (factor 2). An example of a linear sum model would be the calculation of the melting temperature (T_m) of the DNA: $T_m = 4(G + C) + 2(A + T)$. A more advanced and complex formula that considers the neighbouring relationships of the base pairs was described by Hooyberghs et al. in 2009 [51].

Considering the independent **sum model**, guide A reached the most potent effect. The homologous region coincides with a well conserved sequence in mice and humans and is considered to be an important sequence for responding to cAMP stimulation within the proximal *renin* promoter [21]. cAMP is known to be an important transcription factor for stimulating *REN* transcription [5, 9].

So far, we can state that the examined promoter regions seem to have different degrees of influence on gene expression of *REN*. The different measured expression levels of the individual transfections can depend not only on the possible different activation ability but also on other factors that were not considered in this work. For example, the regions examined can be occupied by other TFs. A special secondary structure of the DNA or the arrangement of the nucleosomes might also make the access to the SAM complex more difficult [52]. However, this gain in knowledge would probably also have been possible with classical reporter assays. Most of the knowledge about the importance of *cis*-regulatory areas of the *REN* is based on this type of promoter studies [8, 9].

However, whether there are interactions within *cis*-regulatory regions should be a central question in the research of the complex transcriptional network [44]. In particular, if one looks at the proximal promoter of *REN*, in our opinion, the question about the existence of complex interactions and whether these may have significance for the interpretation of *cis*-regulatory elements was given too little attention [9, 26]. Due to the experimental setup, the ability to explore these dynamics is a limitation for classical promoter assays. With our approach, however, this question could be examined.

To address this issue, we enabled more complexity to explain the measured expression levels. The next level of complexity would include—apart from the linear scaling factors—those factors describing statistical interaction. Therefore, an **interaction model** was subsequently computed and fitted to the measured data. Looking at the interaction model we could see the importance of interaction (Fig 6). By allowing interactions of the promoter regions the influences of the linear single factors A to E were no longer significant (with exception of B). Certain sequence interactions significantly outweighed the independent relationship which is assumed by the sum model. This gain of knowledge is one of the major benefits of this combined approach. When interpreting the importance of certain *cis*-regulatory elements, dependencies and interactions between neighbouring *cis*-regulatory elements should be considered. For example, according to our results region A seems to be important for *REN* activation especially in combination with the neighbouring sequences C, D or E. This would mean that if one analyses the role of a native transcription factor targeting region A the other regions need to be considered in this analysis as well.

By tagging the native *REN* with the reporter *firefly luciferase*, the *REN* expression could be analysed in its natural position in the human genome. This is another great advantage over classical reporter studies, as they do not allow investigation in an endogenous context [10–15]. To our knowledge such a combined approach of experimental research of *cis*-regulatory importance and possible statistical interaction by means of a modified Cas9-system and subsequent computational modelling has not been performed before. The only study known to us

that attempted to elucidate the complexity of a *cis*-regulatory element of renin using a modeling approach has been published in 2007 by Mrowka et. al. However, the endogenous context was not considered in that study [42]. Another advantage of this approach is the induction of gene expression by the used trans-activation complex SAM [34]. This is because the TFs of this complex are always the same. Therefore, any variations in the diversity of transcription result from the *cis*-regions. For instance, if one would use two different transcription factors with two different *cis*-regulatory regions it would not be possible to make a statement about *cis*-regulatory interaction of the region in question.

Of course, each of the generated models have advantages and disadvantages. The sum model is based on fewer parameters than the interaction model. More parameters increase the risk of overfitting of a selected linear model, which is why the sum model has advantages here. When looking at the calculated prediction error, the interaction model shows a smaller value. This fact can also be attributed to the higher number of parameters which the interaction model is based on. Another possibility of comparing the models is the AIC. The AIC is a common criterion for the evaluation of different linear models of the same dataset. On the one hand, it rewards the goodness of fit (likelihood function), but it also contains a penalty term, which penalises too high model complexity. This corresponds to an evaluation based on the principle of maximum parsimony. With an increase in model complexity, the goodness of the fit usually gains as well (risk of overfitting). But the calculated AIC was also smaller for the interaction model. Another big advantage of this approach is that the entire statistical evaluation in the form of model fitting and model evaluation by 2-group Mann-Whitney U Test of the respective prediction errors and calculation of the AIC did not need any further parameters. This does not apply to the experimental parameters such as incubation temperature or the determination of the timepoint of approximately 60 hours after transfection for the modeling approach.

In conclusion, we found that the interaction model explains the measured expression levels of *REN* better. Thus, interactions between the individual sequences seem to be necessary to explain the measured activity levels of *REN*. This in turn would mean that interactions between the individual promoter sections are necessary in order to describe the transcription of *REN* via the proximal *renin* promoter. In this study showed that complex interactions occur within the selected *cis* regions of the endogenous proximal promoter of *REN*, which could be relevant for a better understanding of its transcriptional regulation. This novel combined approach of combinatorial transfections and mathematical modelling and the use of a modern Cas9-based trans-activating complex expands the possibility to study the *cis*-regulatory importance of non-coding DNA in an endogenous context. Further, this approach might be potentially useful to examine other genomic regions.

Supporting information

S1 Fig. Cassette stably integrated into the genome of the HEK dCas9-SAM_Renin-luciferase cells. The cassette contains the Firefly luciferase, which was used as a reporter for *REN* expression. Puromycin was used to select the cells. The homology arms were required for the correct in frame insertion of the cassette. The elements G418-resistance and thymidine-kinase also contained in the cassette were not used in this study.

(TIF)

S2 Fig. Strength of influence of the investigated promoter regions based on the sum model. All investigated sequences significantly affect the activation of *REN*. Considering the sum model, guide A reached the most potent effect. Significances resulted from the

linear regression modelling with the *lm* function of the *stats* package in R. (p-values: * <0,05, *** <0,0005).

(TIF)

S1 Table. Luciferase activity for all 32 guide combinations at three different time points.

(DOCX)

S1 Script. R script source code.

(PDF)

S1 Data.

(CSV)

Acknowledgments

We thank Silke Nossmann, Gesine Hauschild, and Nicole Kaiser for technical assistance in cell culture.

Author Contributions

Conceptualization: Steven Kirchner, Stefanie Reuter, Ralf Mrowka.

Data curation: Steven Kirchner.

Formal analysis: Steven Kirchner, Ralf Mrowka.

Investigation: Steven Kirchner.

Methodology: Anika Westphal, Ralf Mrowka.

Resources: Stefanie Reuter, Anika Westphal, Ralf Mrowka.

Software: Ralf Mrowka.

Supervision: Stefanie Reuter, Ralf Mrowka.

Validation: Steven Kirchner, Ralf Mrowka.

Visualization: Steven Kirchner.

Writing – original draft: Steven Kirchner.

Writing – review & editing: Steven Kirchner, Stefanie Reuter, Anika Westphal, Ralf Mrowka.

References

1. Yuh CH, Bolouri H, Davidson EH. Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science*. 1998; 279(5358):1896–902. <https://doi.org/10.1126/science.279.5358.1896> PMID: 9506933
2. Soufi M, Ruppert V, Rinné S, Mueller T, Kurt B, Pilz G, et al. Increased KCNJ18 promoter activity as a mechanism in atypical normokalemic periodic paralysis. *Neurol Genet*. 2018; 4(5):e274–e. <https://doi.org/10.1212/NXG.0000000000000274> PMID: 30338294
3. Patel S, Rauf A, Khan H, Abu-Izneid T. Renin-angiotensin-aldosterone (RAAS): The ubiquitous system for homeostasis and pathologies. *Biomed Pharmacother*. 2017; 94:317–25. <https://doi.org/10.1016/j.biopha.2017.07.091> PMID: 28772209
4. Sparks MA, Crowley SD, Gurley SB, Mirotso M, Coffman TM. Classical Renin-Angiotensin System in Kidney Physiology. *Compr Physiol*. 2014; 4(3):1201–28. <https://doi.org/10.1002/cphy.c130040> PMID: 24944035
5. Kurtz A. Control of Renin Synthesis and Secretion. *Am J Hypertens*. 2012; 25(8):839–47. <https://doi.org/10.1038/ajh.2011.246> PMID: 22237158

6. World Health Organization. Hypertension 2019, september 13 [<https://www.who.int/news-room/fact-sheets/detail/hypertension>].
7. Schweda F, Friis U, Wagner C, Skott O, Kurtz A. Renin release. *Physiology*. 2007; 22(5):310–9.
8. Castrop H, Hocherl K, Kurtz A, Schweda F, Todorov V, Wagner C. Physiology of Kidney Renin. *Physiol Rev*. 2010; 90(2):607–73. <https://doi.org/10.1152/physrev.00011.2009> PMID: 20393195
9. Glenn ST, Jones CA, Gross KW, Pan L. Control of rene gene expression. *Pflugers Arch*. 2013; 465(1):13–21. <https://doi.org/10.1007/s00424-012-1110-2> PMID: 22576577
10. Borensztein P, Germain S, Fuchs S, Philippe J, Corvol P, Pinet F. CIS-REGULATORY ELEMENTS AND TRANS-ACTING FACTORS DIRECTING BASAL AND CAMP-STIMULATED HUMAN RENIN GENE-EXPRESSION IN CHORIONIC CELLS. *CircRes*. 1994; 74(5):764–73.
11. Petrovic N, Black TA, Fabian JR, Kane C, Jones CA, Loudon JA, et al. Role of proximal promoter elements in regulation of renin gene transcription. *J Biol Chem*. 1996; 271(37):22499–505. <https://doi.org/10.1074/jbc.271.37.22499> PMID: 8798416
12. Shi Q, Black TA, Gross KW, Sigmund CD. Species-specific differences in positive and negative regulatory elements in the renin gene enhancer. *CircRes*. 1999; 85(6):479–88.
13. Voigtlander T, Ganten D, Bader M. Transcriptional regulation of the rat renin gene by regulatory elements in intron I. *Hypertension*. 1999; 33(1):303–11.
14. Pan L, Black TA, Shi Q, Jones CA, Petrovic N, Loudon J, et al. Critical roles of a cyclic AMP responsive element and an E-box in regulation of mouse renin gene expression. *J Biol Chem*. 2001; 276(49):45530–8. <https://doi.org/10.1074/jbc.M103010200> PMID: 11564732
15. Pan L, Jones CA, Glenn ST, Gross KW. Identification of a novel region in the proximal promoter of the mouse renin gene critical for expression. *Am J Physiol-Renal Physiol*. 2004; 286(6):F1107–F15. <https://doi.org/10.1152/ajprenal.00319.2003> PMID: 14761860
16. Mrowka R, Steinhage K, Patzak A, Persson PB. An evolutionary approach for identifying potential transcription factor binding sites: the renin gene as an example. *Am J Physiol-Regul Integr Comp Physiol*. 2003; 284(4):R1147–R50. <https://doi.org/10.1152/ajpregu.00448.2002> PMID: 12626372
17. Zhou XY, Davis DR, Sigmund CD. The human renin kidney enhancer is required to maintain base-line renin expression but is dispensable for tissue-specific, cell-specific, and regulated expression. *J Biol Chem*. 2006; 281(46):35296–304. <https://doi.org/10.1074/jbc.M608055200> PMID: 16990260
18. Fuchs S, Philippe J, Germain S, Mathieu F, Jeunemaitre X, Corvol P, et al. Functionality of two new polymorphisms in the human renin gene enhancer region. *J Hypertens*. 2002; 20(12):2391–8. <https://doi.org/10.1097/00004872-200212000-00018> PMID: 12473863
19. Glenn ST, Jones CA, Pan L, Gross KW. In vivo analysis of key elements within the renin regulatory region. *Physiol Genomics*. 2008; 35(3):243–53. <https://doi.org/10.1152/physiolgenomics.00017.2008> PMID: 18780761
20. Itani HA, Liu XB, Pratt JH, Sigmund CD. Functional characterization of polymorphisms in the kidney enhancer of the human renin gene. *Endocrinology*. 2007; 148(3):1424–30. <https://doi.org/10.1210/en.2006-1381> PMID: 17158202
21. Germain S, Konoshita T, Philippe J, Corvol P, Pinet F. Transcriptional induction of the human renin gene by cyclic AMP requires cyclic AMP response element-binding protein (CREB) and a factor binding a pituitary-specific trans-acting factor (Pit-1) motif. *Biochem J*. 1996; 316:107–13. <https://doi.org/10.1042/bj3160107> PMID: 8645191
22. Tamura K, Umemura S, Yamaguchi S, Iwamoto T, Kobayashi S, Fukamizu A, et al. MECHANISM OF CAMP REGULATION OF RENIN GENE-TRANSCRIPTION BY PROXIMAL PROMOTER. *J Clin Invest*. 1994; 94(5):1959–67. <https://doi.org/10.1172/JCI117547> PMID: 7962542
23. Ying LH, Morris BJ, Sigmund CD. Transactivation of the human renin promoter by the cyclic AMP/protein kinase a pathway is mediated by both cAMP-responsive element binding protein-1 (CREB)-dependent and CREB-independent mechanisms in Calu-6 cells. *J Biol Chem*. 1997; 272(4):2412–20. <https://doi.org/10.1074/jbc.272.4.2412> PMID: 8999953
24. Todorov VT, Desch M, Schmitt-Nilson N, Todorova A, Kurtz A. Peroxisome proliferator-activated receptor-gamma is involved in the control of renin gene expression. *Hypertension*. 2007; 50(5):939–44. <https://doi.org/10.1161/HYPERTENSIONAHA.107.092817> PMID: 17785633
25. Todorov VT, Desch M, Schubert T, Kurtz A. The Pal3 promoter sequence is critical for the regulation of human renin gene transcription by peroxisome proliferator-activated receptor-gamma. *Endocrinology*. 2008; 149(9):4647–57. <https://doi.org/10.1210/en.2008-0127> PMID: 18483152
26. Pan L, Gross KW. Transcriptional regulation of renin—An update. *Hypertension*. 2005; 45(1):3–8. <https://doi.org/10.1161/01.HYP.0000149717.55920.45> PMID: 15545507
27. Fietze S, Farnham PJ. Transcription factor effector domains. *Subcell Biochem*. 2011; 52:261–77. https://doi.org/10.1007/978-90-481-9069-0_12 PMID: 21557087

28. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin YM, Albu M, et al. The Human Transcription Factors. *Cell*. 2018; 172(4):650–65. <https://doi.org/10.1016/j.cell.2018.01.029> PMID: 29425488
29. Reese JC. Basal transcription factors. *Curr Opin Genet Dev*. 2003; 13(2):114–8. [https://doi.org/10.1016/s0959-437x\(03\)00013-3](https://doi.org/10.1016/s0959-437x(03)00013-3) PMID: 12672487
30. Shilatifard A, Conaway RC, Conaway JW. The RNA polymerase II elongation complex. *Annu Rev Biochem*. 2003; 72:693–715. <https://doi.org/10.1146/annurev.biochem.72.121801.161551> PMID: 12676794
31. Lobe CG. TRANSCRIPTION FACTORS AND MAMMALIAN DEVELOPMENT. *Curr Top Dev Biol*. 1992; 27:351–83. [https://doi.org/10.1016/s0070-2153\(08\)60539-6](https://doi.org/10.1016/s0070-2153(08)60539-6) PMID: 1424766
32. Pawson T. SIGNAL-TRANSDUCTION—A CONSERVED PATHWAY FROM THE MEMBRANE TO THE NUCLEUS. *Dev Genet*. 1993; 14(5):333–8. <https://doi.org/10.1002/dvg.1020140502> PMID: 8293575
33. Korge S, Grudziecki A, Kramer A. Highly Efficient Genome Editing via CRISPR/Cas9 to Create Clock Gene Knockout Cells. *J Biol Rhythms*. 2015; 30(5):389–95. <https://doi.org/10.1177/0748730415597519> PMID: 26243628
34. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015; 517(7536):583–U332. <https://doi.org/10.1038/nature14136> PMID: 25494202
35. Maeder ML, Linder SJ, Cascio VM, Fu YF, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*. 2013; 10(10):977+. <https://doi.org/10.1038/nmeth.2598> PMID: 23892898
36. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013; 500(7463):472+. <https://doi.org/10.1038/nature12466> PMID: 23877069
37. Gilbert LA, Larson MH, Morsut L, Liu ZR, Brar GA, Torres SE, et al. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*. 2013; 154(2):442–51. <https://doi.org/10.1016/j.cell.2013.06.044> PMID: 23849981
38. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013; 10(10):973+. <https://doi.org/10.1038/nmeth.2600> PMID: 23892895
39. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012; 337(6096):816–21. <https://doi.org/10.1126/science.1225829> PMID: 22745249
40. Beerli RR, Segal DJ, Dreier B, Barbas CF. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc Natl Acad Sci U S A*. 1998; 95(25):14628–33. <https://doi.org/10.1073/pnas.95.25.14628> PMID: 9843940
41. Tamura K, Tanimoto K, Murakami K, Fukamizu A. A COMBINATION OF UPSTREAM AND PROXIMAL ELEMENTS IS REQUIRED FOR EFFICIENT EXPRESSION OF THE MOUSE RENIN PROMOTER IN CULTURED-CELLS. *Nucleic Acids Res*. 1992; 20(14):3617–23. <https://doi.org/10.1093/nar/20.14.3617> PMID: 1641328
42. Mrowka R, Steege A, Kaps C, Herzelt H, Thiele BJ, Persson PB, et al. Dissecting the action of an evolutionary conserved non-coding region on renin promoter activity. *Nucleic Acids Res*. 2007; 35(15):5120–9. <https://doi.org/10.1093/nar/gkm535> PMID: 17660193
43. Setty Y, Mayo AE, Surette MG, Alon U. Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A*. 2003; 100(13):7702–7. <https://doi.org/10.1073/pnas.1230759100> PMID: 12805558
44. Nieuwenhuis S, Forstmann BU, Wagenmakers EJ. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nat Neurosci*. 2011; 14(9):1105–7. <https://doi.org/10.1038/nn.2886> PMID: 21878926
45. Fisher RA. The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients. *Journal of the Royal Statistical Society*. 1922; 85(4):597–612.
46. Akaike H. CITATION CLASSIC—A NEW LOOK AT THE STATISTICAL-MODEL IDENTIFICATION. *Current Contents/Engineering Technology & Applied Sciences*. 1981(51):22-.
47. Dull T, Zufferey R, Kelly M, Mandel RJ, Nguyen M, Trono D, et al. A third-generation lentivirus vector with a conditional packaging system. *J Virol*. 1998; 72(11):8463–71. PMID: 9765382
48. Zhang Lab. sgRNA design tool 2020 [<http://sam.genome-engineering.org/database/>].
49. Hsu PD, Lander ES, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*. 2014; 157(6):1262–78. <https://doi.org/10.1016/j.cell.2014.05.010> PMID: 24906146

50. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2017.
51. Hooyberghs J, Van Hummelen P, Carlon E. The effects of mismatches on hybridization in DNA microarrays: determination of nearest neighbor parameters. *Nucleic Acids Res.* 2009; 37(7):11.
52. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Bio.* 2015; 16(3):178–89.

5 Diskussion

CRISPR mit seinen assoziierten Cas-Enzymen hat sich seit seiner Entdeckung zu der potentesten und am weitesten verbreiteten Methode des Gene-Editing entwickelt. In einem neuartigen kombinierten Ansatz wird mit dieser Arbeit eine Erweiterung der Einsatzmöglichkeiten der ursprünglichen Endonuklease Cas9 vorgestellt. Durch kombinatorische Transfektionen spezifischer sgRNAs in Zellen, die stabil einen modifizierten Cas9-Komplex exprimieren und mathematische Modellierung der gemessenen experimentellen Daten soll aufgezeigt werden, dass CRISPR/Cas9 effektiv zur Promotorforschung im endogenen genomischen Kontext eingesetzt werden kann.

5.1 Gen-Aktivierung durch einen modifizierten Cas9-Komplex

Die für die experimentellen Arbeiten verwendeten HEK-Zellen weisen keine basale Expression von REN auf. Die Zellen wurden über einen lentiviralen Ansatz stabil mit allen notwendigen Genen zur Expression eines modifizierten und aktivierenden Cas9-Komplexes (SAM-Komplex) transfiziert. Durch Transfektion spezifischer sgRNA, die homolog zu gewählten Sequenzen des proximalen Renin-Promotors sind, konnte eine gezielte Translokation des SAM-Komplexes zu den Promotorregionen erwirkt werden.

Dem proximalen Renin-Promotor wird nach Experimenten am murinen Ren-1C-Gen generell ein geringes Aktivierungspotential zugesprochen (Petrovic et al. 1996, Pan et al. 2001). In unserer Arbeit konnte jedoch eine signifikante REN-Expression durch gezielte Translokation der SAM-Komplexe zu allen fünf Promotorsequenzen induziert werden. Diese Ergebnisse verdeutlichen, wie effizient Gene mithilfe von Weiterentwicklungen der Cas9 aktiviert werden können.

In experimentellen Vorversuchen wurde der Reninspiegel nach Aktivierung dieser Promotorregion bestimmt. Im Überstand der HEK-Zellen konnte eine katalytische Aktivität des Renins nachgewiesen werden. Dies zeigt, dass die Experimente in einem physiologischen Bereich der Reninexpression erfolgten. Weiterhin zeigt es, dass in unseren Zellen sekretiertes Renin produziert wurde.

5.2 Erforschung *cis*-regulatorischer Interaktionen durch einen kombinierten experimentellen und modellierenden Ansatz

Neben den kombinatorischen Transfektionen beinhaltet unser kombinierter Ansatz auch die mathematische Modellierung der Messdaten. Dies ermöglicht tiefere Einblicke in *cis*-regulatorische Wechselwirkungen. Unabhängige lineare Modelle sind eine Methode, um multifaktorielle Einflüsse auf eine Ausgabevariable (in diesem Fall die Renin-Transkription) zu beschreiben. Im einfachsten Fall wäre dies ein Zwei-Faktor-Setup, z.B. ob ein Medikament (Faktor 1) geschlechtsabhängig wirkt (Faktor 2). Ein Beispiel für ein lineares Summenmodell wäre die vereinfachte Berechnung der Schmelztemperatur (T_m) der DNA: $T_m = 4(G+C) + 2(A+T)$.

Unserer Meinung nach sollte es bei der Erforschung des komplexen transkriptionellen Netzwerkes von zentralem Interesse sein, ob statistisch relevante Interaktionen untersuchter *cis*-regulatorischer Regionen existieren oder nicht (Nieuwenhuis et al. 2011). Bezogen auf die in der vorliegenden Arbeit gewählte Region des proximalen Renin-Promotors wurde dieser Fragestellung trotz umfangreicher Forschung zu wenig Beachtung geschenkt. Eine Begründung liegt hierfür im experimentellen Setup der bislang erfolgten Promotorforschung, deren Erkenntnisse maßgeblich auf Ergebnissen klassischer Plasmid basierter Promotorstudien beruhen. Zum einen ist eine Untersuchung möglicher Interaktionen *cis*-regulatorischer Elemente im endogenen Kontext in klassischen Promotorstudien nur eingeschränkt möglich. Zum anderen zeigen Promotoren, die aus ihrem nativen Kontext herausgelöst werden, oft veränderte Expressionsmuster (Hernandez-Garcia und Finer 2014). Mit dem für diese Arbeit entwickelten kombinierten Ansatz können wir dieser Fragestellung nach möglichen statistischen Interaktionen im endogenen genomischen Kontext nachgehen.

Mit Hilfe der entwickelten methodischen Kombination gelang es, signifikante Erkenntnisse in Bezug auf die Wechselwirkungen der untersuchten genomischen Regionen zu erlangen. Die kombinatorischen Transfektionen der spezifischen sgRNAs bedingten unterschiedliche Expressionsmuster von REN. Alle fünf sgRNAs bewirkten eine signifikante Aktivierung von REN über ihre spezifische Promotorregion. Durch ein Zulassen von Interaktionen in der computergestützten Modellierung und statistischen Auswertung der gemessenen Expressionsmuster

verloren vier der fünf untersuchten Regionen ihren signifikanten Einfluss zur Erklärung der Expressionsmuster, sofern man diese als lineare Einzelfaktoren betrachtete. Deren Einflussstärke der REN-Aktivierung wurde durch bestimmte Kombinationen der untersuchten Promotorregionen deutlich überwogen. Diese Ergebnisse unterstreichen, dass Interaktionen innerhalb der Promotorregionen für das Ausmaß der bewerteten Genaktivität wesentlich sind. Das bedeutet, dass bei der Interpretation der Bedeutung bestimmter *cis*-regulatorischer Elemente Abhängigkeiten und Wechselwirkungen zwischen benachbarten *cis*-Elementen berücksichtigt werden sollten. Wenn man also die Rolle eines nativen Transkriptionsfaktors analysiert, der auf eine untersuchte regulatorische Region abzielt, dann sollten benachbarte relevante Regionen ebenfalls bei dieser Analyse berücksichtigt werden.

5.3 Promotorforschung im endogenen genomischen Kontext

Um eine potenzielle Genexpression eines Zielgens quantifizieren zu können ist die Markierung mit einem Reporter gen eine etablierte molekulargenetische Methode. In der vorliegenden Arbeit wurden die verwendeten HEK-Zellen stabil mit der kodierenden Sequenz des Reporters Luziferase markiert. Dies gelang durch den Einsatz einer katalytisch aktiven Cas9 (Kirchner et al. 2020). Nach gezielter Einführung eines Doppelstrangbruchs und Bereitstellen der genetischen Sequenz der Luziferase erfolgt unter Ausnutzen des Reparaturmechanismus der homologen Rekombination die stabile Insertion des Reporters an das native REN. Eine Kodierung der Luziferase resultiert somit nur bei einer erfolgreichen Aktivierung von REN. Die REN-Expression kann damit in ihrer natürlichen Position im menschlichen Genom analysiert werden. Dies ist ein weiterer großer Vorteil gegenüber klassischen Reporterstudien, die keine Untersuchung im endogenen Kontext zulassen (Borensztein et al. 1994, Petrovic et al. 1996, Shi et al. 1999, Voigtlander et al. 1999, Pan et al. 2001, Pan et al. 2004). Nach unserem Kenntnisstand wurde ein solcher kombinierter Ansatz aus experimenteller Forschung zu *cis*-regulatorischer Interaktion mittels eines modifizierten Cas9-Systems und anschließender computergestützter Modellierung noch nicht durchgeführt. Die einzige uns bekannte Studie, die versucht hat, die Komplexität eines *cis*-regulatorischen Elements von REN mit Hilfe eines Modellierungsansatzes aufzuklären, wurde 2007 von Mrowka et al. vorgestellt. Der endogene Kontext wurde jedoch auch in dieser Studie nicht berücksichtigt (Mrowka et al. 2007).

5.4 Methodendiskussion

5.4.1 Einsatz der Cas9 zur Promotorforschung

Für effizientes Gene-Editing stehen heute verschiedene etablierte molekulargenetische Werkzeuge zur Verfügung. Den breitesten Einsatz erfährt inzwischen die Familie der CRISPR assoziierten Proteine und deren Weiterentwicklungen. Mit dem Einsatz der Cas9 erhält man durch ein relativ simples Design einer sgRNA ein spezifisches Werkzeug, das je nach verwendeter Cas9-Variante zum gezielten Gen-Knockout, zur effektiven Aktivierung oder Suppression von Genen, zur epigenetischen Modifikation, zur stabilen Gen-Insertion, zur Durchführung multiplexer Genmodulation oder genomweiten Screenings eingesetzt werden kann (Jinek et al. 2013, Sakuma et al. 2014, Alerasool et al. 2020, Nuñez et al. 2021, Komor et al. 2017). Zudem ist die Umprogrammierung der Cas9 auf neue genomische Zielsequenzen durch simples Design einer neuen sgRNA möglich (Ran et al. 2013a). Verglichen mit etablierten Verfahren wie ZFNs oder TALENs wird besonders bei den ZFNs für neu gewählte genomische Zielsequenzen ggf. das Design, sowie Synthese und Validierung eines neuen DNA-Bindeproteins erforderlich, was deutlich zeit- und kostenintensiver ist (Komor et al. 2017, Ran et al. 2013a).

ZFNs und TALENs sind mit der Endonuklease *FokI* verknüpft, welche nur als Dimer einen Doppelstrangbruch der DNA bewirkt, was zudem im Unterschied zur Cas9 das Design von zwei TALEN- bzw. ZFN-Proteinen an gegenüberliegenden Strängen und in unmittelbarer Nähe der genomischen Zielsequenz erforderlich macht (Kim et al. 1996, Christian et al. 2010).

Ein relevantes Problem beim Einsatz aller Designer-Nukleasen stellen potenzielle off-target Effekte dar, die durch unerwünschte DNA-Spaltung außerhalb des gewünschten Locus zur Toxizität und Absterben der Zellen führen können. Für alle bestehenden Designer-Nukleasen ermöglichten eingeführte Weiterentwicklungen eine Minimierung möglicher off-target Effekte (Ramalingam et al. 2011, Montague et al. 2014). Im Falle der ZFNs und TALENs werden diese bereits durch die Bedingung der Dimerisierung der *FokI*-Nuklease auf natürliche Weise begrenzt.

Das Wirksamwerden der Cas9 außerhalb des gewünschten genomischen Locus beruht hauptsächlich auf potenziell mehreren komplementären Genomstellen der sgRNAs (Zhang et al. 2015). Inzwischen sind eine Reihe an Cas9-Varianten mit niedrigerer off-target Spezifität etabliert (Kleinstiver et al. 2016, Slaymaker et al. 2016). Für das experimentelle Setup der vorliegenden Arbeit stellen potenzielle off-target Effekte keine relevante Limitation dar. Zum einen wurden die kombinatorischen Transfektionen nicht mit einer katalytisch aktiven Cas9 durchgeführt, was zytotoxische Nebeneffekte durch fehlerhafte DNA-Spaltung unwahrscheinlich macht. Zum anderen ist nur das untersuchte Zielgen mit einem Reporter gen stabil markiert worden, wodurch falsch positive Ergebnisse bei potenzieller Aktivierung anderer Gene nicht zu erwarten sind.

Ein wesentlicher Vorteil des Cas9-Systems besteht zudem darin, durch das Design und Bereitstellen von mehreren sgRNAs effizient multiplexes Gene-Editing bewältigen zu können (McCarty et al. 2020). Bezogen auf unser experimentelles Setup war es möglich, denselben Cas9-Komplex im Rahmen kombinatorischer Transfektionen von verschiedenen sgRNAs simultan zu mehreren genomischen Zielsequenzen zu leiten. Mit aktivierenden ZF- oder TALE-Domänen ist ein solcher Versuchsaufbau nur unter erheblichen Mehraufwand umsetzbar.

Als eine Limitation des Einsatzes der Cas9 zum gezielten Gene-Editing wird in der Literatur oft die notwendige PAM-Sequenz erwähnt, die für die volle Funktionsfähigkeit der Cas-Proteine notwendig ist (Shah et al. 2013). Für unser experimentelles Setup konnten jedoch innerhalb der gewählten genomischen Zielsequenz fünf qualitativ vergleichbare, sich nicht überlappende Guide-Sequenzen designt werden. Die PAM-Sequenz der verwendeten Cas9 hat die Basenfolge -NGG-. Da diese Sequenz durchschnittlich alle 8 bis 12 bp im menschlichen Genom vorkommt (Ran et al. 2013a), stellte das Design für unser experimentelles Setup keine relevante Limitation dar.

Aufgrund der Untersuchung im endogenen Kontext könnten die gemessenen Expressionsniveaus der einzelnen Transfektionen nicht nur von der möglichen unterschiedlichen Aktivierungsfähigkeit der *cis*-regulatorischen Elemente abhängen, sondern auch von anderen Faktoren, die in dieser Arbeit nicht berücksichtigt wurden. Beispielsweise könnten die Regionen von anderen TFs besetzt sein.

Auch eine spezielle Sekundärstruktur der DNA oder die Anordnung der Nukleosomen könnte den Zugang zum SAM-Komplex erschweren (Venkatesh und Workman 2015, Hernandez-Garcia und Finer 2014).

Zudem könnte ein Ansteuern multipler Cas9-Komplexe an eng benachbarte Zielsequenzen zu gegenseitigen Behinderungen und damit verbundenen Funktionseinschränkungen führen. Zur Adressierung dieses Problems stehen inzwischen alternative Cas9-Proteine, beispielsweise aus *Staphylococcus aureus*, mit deutlich geringerer Molekülgröße zur Verfügung (Komor et al. 2017). Diese benötigen jedoch oft eine komplexere PAM-Sequenz, was den frei programmierbaren Einsatz speziell für unser gewähltes experimentelles Setup deutlich einschränken würde.

Die Verwendung des SAM-Komplexes zur Induktion der Genexpression stellt einen wesentlichen Vorteil der Studie dar. Aufgrund dessen, dass die TFs dieses Komplexes immer gleich sind, resultieren alle Variationen in der Diversität der Transkription aus den *cis*-Regionen. Würde man beispielsweise zwei unterschiedliche Transkriptionsfaktoren zur Untersuchung zweier oder mehrerer *cis*-regulatorischen Regionen verwenden, wäre keine valide Aussage über die *cis*-regulatorische Interaktion der jeweiligen Region möglich.

5.4.2 Experimentelles Setup

Um bestmögliche Validität der Ergebnisse zu erzielen, wurden in Vorbereitung auf die kombinatorischen Transfektionen nicht publizierte experimentelle Vorarbeiten durchgeführt. Hieraus ergaben sich die optimale Transfektionsmenge von 30 ng Guide-DNA pro Well einer 96 Well Platte, sowie die Ermittlung der optimalen Inkubationsdauer nach erfolgter Transfektion. Im Zeitverlauf konnte ab ca. 70 Stunden nach Transfektion eine Abflachung des Anstiegs der Luziferase-Aktivität beobachtet werden. Mehrere Ursachen können dies bedingen. Eine Abnahme der Qualität des Nährmediums, ein gedrosselter Stoffwechsel der Zellen aufgrund der zu diesem Zeitpunkt bereits sehr hohen Zelldichte, regulatorische Effekte und Feedback-Mechanismen innerhalb der Expression des Renin-Gens könnten zu einer solchen Abflachung führen. Um diesen systematischen Fehlerquellen auszuweichen, wurden für die erfolgten kombinatorischen Experimente die Auswertungen nach ca. 60 Stunden vorgenommen. Außerdem wurde in den durchgeführten Experimenten nicht der Einfluss von unterschiedlichen Konformationen und Zuständen der Plasmid-DNA auf die Effizienz der Transfektionen untersucht.

Zur Auswertung der Messdaten wurden zwei mathematische Modelle verwendet. Natürlich hat jedes der generierten Modelle Vor- und Nachteile. Das Summenmodell basiert auf weniger Parametern als das Interaktionsmodell. Mehr Parameter erhöhen das Risiko einer Überanpassung eines ausgewählten linearen Modells, weshalb das Summenmodell hier bevorteilt ist. Bei Betrachtung des berechneten Vorhersagefehlers zeigt das Interaktionsmodell einen kleineren Wert. Diese Tatsache ist auch auf die höhere Anzahl von Parametern zurückzuführen, auf denen das Interaktionsmodell basiert. Eine weitere Möglichkeit zum Vergleich der Modelle ist das AIC. Das AIC ist ein Kriterium für die Bewertung verschiedener linearer Modelle desselben Datensatzes. Es wird einerseits die Anpassungsgüte bewertet. Daneben enthält es aber auch einen Strafterm, der eine zu hohe Modellkomplexität bestraft. Dies entspricht einer Bewertung nach dem Prinzip maximaler Parsimonie bzw. Sparsamkeit. Mit steigender Modellkomplexität nimmt in der Regel auch die Güte der Anpassung zu (Gefahr der Überanpassung). Aber auch das berechnete AIC war für das Interaktionsmodell kleiner.

6 Schlussfolgerungen

Die Entdeckung von CRISPR und seinen assoziierten Proteinen hat das Gene-Editing revolutioniert. Die vorliegende Arbeit unterstreicht die Vielfältigkeit und Effizienz, mit der Gene-Editing heute unter Verwendung von Cas9-Varianten möglich ist.

So konnte in dieser Arbeit unter Verwendung einer enzymatisch aktiven Cas9 erfolgreich eine stabile Markierung des endogenen REN mit einer kodierenden Sequenz des Reportergens Luziferase erfolgen. Durch gezieltes Ansteuern eines modifizierten, aktivierenden Cas9-Komplexes auf eine gewählte genomische Promotorregion gelang anschließend erfolgreich die Aktivierung des markierten Zielgens. Mit Hilfe unseres kombinierten experimentellen Ansatzes konnten zudem signifikante statistische Interaktionen der untersuchten *cis*-regulatorischen Elemente festgestellt werden.

Die Ergebnisse dieser Arbeit zeigen, dass das Cas9-System auch effizient zur Promotorforschung im endogenen genomischen Kontext Verwendung finden kann.

Die erlangten Erkenntnisse könnten zudem dazu beitragen, die transkriptionelle Regulation des humanen Renins, welches als Schlüsselprotein innerhalb des lebensnotwendigen Renin-Angiotensin-Aldosteron-Systems zählt, besser zu verstehen.

7 Literatur- und Quellenverzeichnis

- Akaike H. 1981. CITATION CLASSIC - A NEW LOOK AT THE STATISTICAL-MODEL IDENTIFICATION. *Current Contents/Engineering Technology & Applied Sciences*, (51):22-22.
- Alerasool N, Segal D, Lee H, Taipale M. 2020. An efficient KRAB domain for CRISPRi applications in human cells. *Nature Methods*, 17 (11):1093-1096.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315 (5819):1709-1712.
- Bedell VM, Wang Y, Campbell JM, Poshusta TL, Starker CG, Krug RG, Tan WF, Penheiter SG, Ma AC, Leung AYH, Fahrenkrug SC, Carlson DF, Voytas DF, Clark KJ, Essner JJ, Ekker SC. 2012. In vivo genome editing using a high-efficiency TALEN system. *Nature*, 491 (7422):114-U133.
- Bednarski C, Cathomen T. 2015. Maßgeschneidertes Genom — Designer-Nukleasen im Einsatz. *BIOspektrum*, 21 (1):22-24.
- Berli RR, Dreier B, Barbas CF. 2000. Positive and negative regulation of endogenous genes by designed transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (4):1495-1500.
- Bibikova M, Beumer K, Trautman JK, Carroll D. 2003. Enhancing gene targeting with designed zinc finger nucleases. *Science*, 300 (5620):764-764.
- Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326 (5959):1509-1512.
- Borensztein P, Germain S, Fuchs S, Philippe J, Corvol P, Pinet F. 1994. CIS-REGULATORY ELEMENTS AND TRANS-ACTING FACTORS DIRECTING BASAL AND CAMP-STIMULATED HUMAN RENIN GENE-EXPRESSION IN CHORIONIC CELLS. *Circulation Research*, 74 (5):764-773.
- Carroll D. 2011. Zinc-Finger Nucleases: A Panoramic View. *Current Gene Therapy*, 11 (1):2-10.
- Castrop H, Hoehrl K, Kurtz A, Schweda F, Todorov V, Wagner C. 2010. Physiology of Kidney Renin. *Physiological Reviews*, 90 (2):607-673.
- Chen BH, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. 2013. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, 155 (7):1479-1491.
- Chirgwin JM, Schaefer IM, Rotwein PS, Piccini N, Gross KW, Naylor SL. 1984. HUMAN RENIN GENE IS ON CHROMOSOME-1. *Somatic Cell and Molecular Genetics*, 10 (4):415-421.
- Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. 2010. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*, 186 (2):757-U476.
- Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, 34 (2):184-191.
- Fisher RA. 1922. The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients. *Journal of the Royal Statistical Society*, 85 (4):597-612.
- Frangoul H, Altshuler D, Cappellini MD, Chen YS, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, Ho TW, Kattamis A, Kernytsky A, Lekstrom-Himes J, Li AM, Locatelli F, Mapara MY, de Montalembert M, Rondelli D, Sharma A, Sheth S, Soni S, Steinberg MH, Wall D, Yen A, Corbacioglu S. 2021. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and beta-Thalassemia. *New England Journal of Medicine*, 384 (3):252-260.
- Fuchs S, Philippe J, Germain S, Mathieu F, Jeunemaitre X, Corvol P, Pinet F. 2002. Functionality of two new polymorphisms in the human renin gene enhancer region. *Journal of Hypertension*, 20 (12):2391-2398.
- Gaj T, Gersbach CA, Barbas CF. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, 31 (7):397-405.

- Germain S, Konoshita T, Philippe J, Corvol P, Pinet F. 1996. Transcriptional induction of the human renin gene by cyclic AMP requires cyclic AMP response element-binding protein (CREB) and a factor binding a pituitary-specific trans-acting factor (Pit-1) motif. *Biochemical Journal*, 316:107-113.
- Gersbach CA, Perez-Pinera P. 2014. Activating human genes with zinc finger proteins, transcription activator-like effectors and CRISPR/Cas9 for gene therapy and regenerative medicine. *Expert Opinion on Therapeutic Targets*, 18 (8):835-839.
- Gilbert LA, Larson MH, Morsut L, Liu ZR, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell*, 154 (2):442-451.
- Glenn ST, Jones CA, Pan L, Gross KW. 2008. In vivo analysis of key elements within the renin regulatory region. *Physiological Genomics*, 35 (3):243-253.
- Glenn ST, Jones CA, Gross KW, Pan L. 2013. Control of renin gene expression. *Pflügers Archiv-European Journal of Physiology*, 465 (1):13-21.
- Greiner A. 2015. CRISPR/Cas9 und Zinkfinger-Nukleasen für die gezielte Genstilllegung in *Chlamydomonas reinhardtii*.
- Hernandez-Garcia CM, Finer JJ. 2014. Identification and validation of promoters and cis-acting regulatory elements. *Plant Science*, 217:109-119.
- Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. NUCLEOTIDE-SEQUENCE OF THE IAP GENE, RESPONSIBLE FOR ALKALINE-PHOSPHATASE ISOZYME CONVERSION IN *ESCHERICHIA-COLI*, AND IDENTIFICATION OF THE GENE-PRODUCT. *Journal of Bacteriology*, 169 (12):5429-5433.
- Itani HA, Liu XB, Pratt JH, Sigmund CD. 2007. Functional characterization of polymorphisms in the kidney enhancer of the human renin gene. *Endocrinology*, 148 (3):1424-1430.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337 (6096):816-821.
- Jinek M, East A, Cheng A, Lin S, Ma EB, Doudna J. 2013. RNA-programmed genome editing in human cells. *Elife*, 2:9.
- Joung JK, Sander JD. 2013. INNOVATION TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, 14 (1):49-55.
- Kim YG, Cha J, Chandrasegaran S. 1996. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93 (3):1156-1160.
- Kirchner S, Reuter S, Westphal A, Mrowka R. 2020. Decipher the complexity of cis-regulatory regions by a modified Cas9. *Plos One*, 15 (7):21.
- Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK. 2016. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, 529 (7587):490-495.
- Komor AC, Badran AH, Liu DR. 2017. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell*, 168 (1-2):20-36.
- Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. 2013. Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, 500 (7463):472-+.
- Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, 517 (7536):583-U332.
- Kurtz A. 2012. Control of Renin Synthesis and Secretion. *American Journal of Hypertension*, 25 (8):839-847.
- Lambert SA, Jolma A, Campitelli LF, Das PK, Yin YM, Albu M, Chen XT, Taipale J, Hughes TR, Weirauch MT. 2018. The Human Transcription Factors. *Cell*, 172 (4):650-665.
- Lobe CG. 1992. TRANSCRIPTION FACTORS AND MAMMALIAN DEVELOPMENT. *Current Topics in Developmental Biology*, 27:351-383.

- Maeder ML, Linder SJ, Cascio VM, Fu YF, Ho QH, Joung JK. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature Methods*, 10 (10):977-+.
- McCarty NS, Graham AE, Studena L, Ledesma-Amaro R. 2020. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat Commun*, 11 (1):1281.
- Molina C, Grotewold E. 2005. Genome wide analysis of Arabidopsis core promoters. *Bmc Genomics*, 6:12.
- Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Research*, 42 (W1):W401-W407.
- Mrowka R, Steinhage K, Patzak A, Persson PB. 2003. An evolutionary approach for identifying potential transcription factor binding sites: the renin gene as an example. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 284 (4):R1147-R1150.
- Mrowka R, Steege A, Kaps C, Herzel H, Thiele BJ, Persson PB, Bluthgen N. 2007. Dissecting the action of an evolutionary conserved non-coding region on renin promoter activity. *Nucleic Acids Research*, 35 (15):5120-5129.
- Nieuwenhuis S, Forstmann BU, Wagenmakers EJ. 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, 14 (9):1105-1107.
- Nuñez JK, Chen J, Pommier GC, Cogan JZ, Replogle JM, Adriaens C, Ramadoss GN, Shi Q, Hung KL, Samelson AJ, Pogson AN, Kim JYS, Chung A, Leonetti MD, Chang HY, Kampmann M, Bernstein BE, Hovestadt V, Gilbert LA, Weissman JS. 2021. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell*, 184 (9):2503-2519.e2517.
- Pan L, Gross KW. 2005. Transcriptional regulation of renin - An update. *Hypertension*, 45 (1):3-8.
- Pan L, Jones CA, Glenn ST, Gross KW. 2004. Identification of a novel region in the proximal promoter of the mouse renin gene critical for expression. *American Journal of Physiology-Renal Physiology*, 286 (6):F1107-F1115.
- Pan L, Black TA, Shi Q, Jones CA, Petrovic N, Loudon J, Kane C, Sigmund CD, Gross KW. 2001. Critical roles of a cyclic AMP responsive element and an E-box in regulation of mouse renin gene expression. *Journal of Biological Chemistry*, 276 (49):45530-45538.
- Patel S, Rauf A, Khan H, Abu-Izneid T. 2017. Renin-angiotensin-aldosterone (RAAS): The ubiquitous system for homeostasis and pathologies. *Biomedicine & Pharmacotherapy*, 94:317-325.
- Pawson T. 1993. SIGNAL-TRANSDUCTION - A CONSERVED PATHWAY FROM THE MEMBRANE TO THE NUCLEUS. *Developmental Genetics*, 14 (5):333-338.
- Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods*, 10 (10):973-+.
- Petrovic N, Black TA, Fabian JR, Kane C, Jones CA, Loudon JA, Abonia JP, Sigmund CD, Gross KW. 1996. Role of proximal promoter elements in regulation of renin gene transcription. *Journal of Biological Chemistry*, 271 (37):22499-22505.
- Ramalingam S, Kandavelou K, Rajenderan R, Chandrasegaran S. 2011. Creating Designed Zinc-Finger Nucleases with Minimal Cytotoxicity. *Journal of Molecular Biology*, 405 (3):630-641.
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013a. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8 (11):2281-2308.
- Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y, Zhang F. 2013b. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*, 154 (6):1380-1389.
- Reardon S. 2015. Leukaemia success heralds wave of gene-editing therapies. *Nature*, 527 (7577):146-147.
- Reese JC. 2003. Basal transcription factors. *Current Opinion in Genetics & Development*, 13 (2):114-118.
- Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. 2016. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature Biotechnology*, 34 (3):339-+.

- Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T. 2014. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep*, 4:5400.
- Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 32 (4):347-355.
- Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F, Doench JG. 2018. Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications*, 9:15.
- Schweda F, Friis U, Wagner C, Skott O, Kurtz A. 2007. Renin release. *Physiology*, 22 (5):310-319.
- Shah SA, Erdmann S, Mojica FJM, Garrett RA. 2013. Protospacer recognition motifs: Mixed identities and functional diversity. *Rna Biology*, 10 (5):891-899.
- Shi Q, Black TA, Gross KW, Sigmund CD. 1999. Species-specific differences in positive and negative regulatory elements in the renin gene enhancer. *Circulation Research*, 85 (6):479-488.
- Shilatifard A, Conaway RC, Conaway JW. 2003. The RNA polymerase II elongation complex. *Annual Review of Biochemistry*, 72:693-715.
- Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. 2016. Rationally engineered Cas9 nucleases with improved specificity. *Science*, 351 (6268):84-88.
- Soufi M, Ruppert V, Rinné S, Mueller T, Kurt B, Pilz G, Maieron A, Dodel R, Decher N, Schaefer JR. 2018. Increased KCNJ18 promoter activity as a mechanism in atypical normokalemic periodic paralysis. *Neurology Genetics*, 4 (5):e274-e274.
- Sparks MA, Crowley SD, Gurley SB, Mirotsoy M, Coffman TM. 2014. Classical Renin-Angiotensin System in Kidney Physiology. *Comprehensive Physiology*, 4 (3):1201-1228.
- Tamura K, Umemura S, Yamaguchi S, Iwamoto T, Kobayashi S, Fukamizu A, Murakami K, Ishii M. 1994. MECHANISM OF CAMP REGULATION OF RENIN GENE-TRANSCRIPTION BY PROXIMAL PROMOTER. *Journal of Clinical Investigation*, 94 (5):1959-1967.
- Todorov VT, Desch M, Schubert T, Kurtz A. 2008. The Pal3 promoter sequence is critical for the regulation of human renin gene transcription by peroxisome proliferator-activated receptor-gamma. *Endocrinology*, 149 (9):4647-4657.
- Todorov VT, Desch M, Schmitt-Nilson N, Todorova A, Kurtz A. 2007. Peroxisome proliferator-activated receptor-gamma is involved in the control of renin gene expression. *Hypertension*, 50 (5):939-944.
- Turchiano G, Andrieux G, Klermund J, Blattner G, Pennucci V, El Gaz M, Monaco G, Poddar S, Mussolino C, Cornu TI, Boerries M, Cathomen T. 2021. Quantitative evaluation of chromosomal rearrangements in gene-edited human stem cells by CAST-Seq. *Cell Stem Cell*, 28 (6):1136-+.
- Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. 2010. Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11 (9):636-646.
- Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews Molecular Cell Biology*, 16 (3):178-189.
- Voigtlander T, Ganten D, Bader M. 1999. Transcriptional regulation of the rat renin gene by regulatory elements in intron I. *Hypertension*, 33 (1):303-311.
- Wang Y, Li Z, Xu J, Zeng B, Ling L, You L, Chen Y, Huang Y, Tan A. 2013. The CRISPR/Cas System mediates efficient genome engineering in *Bombyx mori*. *Cell Research*, 23 (12):1414-1416.
- World Health Organization Hypertension <https://www.who.int/news-room/fact-sheets/detail/hypertension>.
- Wu J, Kandavelou K, Chandrasegaran S. 2007. Custom-designed zinc finger nucleases: What is next? *Cellular and Molecular Life Sciences*, 64 (22):2933-2944.
- Wyman C, Kanaar R. 2006. DNA double-strand break repair: All's well that ends well. *Annual Review of Genetics*, 40:363-383.
- Yang H, Wang HY, Shivalila CS, Cheng AW, Shi LY, Jaenisch R. 2013. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*, 154 (6):1370-1379.

- Ying LH, Morris BJ, Sigmund CD. 1997. Transactivation of the human renin promoter by the cyclic AMP/protein kinase a pathway is mediated by both cAMP-responsive element binding protein-1 (CREB)-dependent and CREB-independent mechanisms in Calu-6 cells. *Journal of Biological Chemistry*, 272 (4):2412-2420.
- Yuh CH, Bolouri H, Davidson EH. 1998. Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science*, 279 (5358):1896-1902.
- Zhang Lab sgRNA design tool <http://sam.genome-engineering.org/database/>.
- Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. 2015. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy-Nucleic Acids*, 4:8.
- Zhou XY, Davis DR, Sigmund CD. 2006. The human renin kidney enhancer is required to maintain base-line renin expression but is dispensable for tissue-specific, cell-specific, and regulated expression. *Journal of Biological Chemistry*, 281 (46):35296-35304.
- Zu Y, Tong XJ, Wang ZX, Liu D, Pan RC, Li Z, Hu YY, Luo Z, Huang P, Wu Q, Zhu ZY, Zhang B, Lin S. 2013. TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nature Methods*, 10 (4):329-+.

8 Anhang

8.1 Danksagung

An dieser Stelle möchte ich allen Menschen meinen großen Dank aussprechen, die mich bei der Anfertigung meiner Dissertation unterstützt haben.

Besonders danken möchte ich Prof. Ralf Mrowka, Frau Dr. Stefanie Reuter und Frau Dr. Anika Westphal für die ausgezeichnete Betreuung und Unterstützung bei der Umsetzung der gesamten Arbeit.

Außerdem möchte ich mich bei Frau Silke Noßmann bedanken, die mich zu Beginn meiner wissenschaftlichen Tätigkeiten im Labor geduldig eingearbeitet hat und zu jeder Zeit mit Rat und Tat zur Seite stand.

Zudem gilt mein Dank der Friedrich-Schiller-Universität Jena, die durch die Etablierung von JENOS und der möglichen Schwerpunktlegung FoM die notwendigen Grundlagen geschaffen haben, bereits während des Studiums effizient wissenschaftlich arbeiten zu können.

Ein besonderer Dank gilt außerdem meiner Lebenspartnerin für ihre Ermutigungen und den Freiraum, den Sie mir trotz vieler Verpflichtungen für die Anfertigung dieser Dissertation geschaffen hat.

8.2 Abbildungsverzeichnis

- Abbildung 1: Designer-Nukleasen im Vergleich.** ZFN: üblicherweise 3 bis 4 Zinkfingerdomänen (ZF) bilden eine DNA-Bindedomäne; eine ZF bindet jeweils ein Basentriplett; durch Dimerisierung der gekoppelten Nuklease *FokI* erfolgt die spezifische Einführung eines Doppelstrangbruches. 8
- Abbildung 2: Gene-Editing mit CRISPR/Cas9.** Dargestellt ist eine vereinfachte, schematische Übersicht der Möglichkeiten des Gene-Editing mit CRISPR/Cas9. Durch gezielte Mutationen und Deaktivierung der Endonuklease-Domänen wurde ein spezifisches DNA-Bindeprotein (dCas9) generiert, wodurch sich die Möglichkeiten des programmierbaren Gene-Editing erweitern. Erstellt mit *Biorender.com*. 11
- Abbildung 3: Humanes Renin-Gen.** Das humane Renin-Gen verfügt über mehrere 5'-flankierende cis-regulatorisch wirksame Elemente. Dazu zählen der bei ca. -12 kbp gelegene Renin-Enhancer, der bei ca. -5,5 kbp gelegene Chorion-Enhancer und die proximale Promotorregion. Erstellt mit *Biorender.com*. 13
- Abbildung 4: Programmierbare spezifische Aktivierung des endogenen REN.** Nach der Assemblierung des stabil transfizierten SAM-Komplexes wird dieser über 5 spezifische sgRNAs an komplementäre Promotorregionen des endogenen REN transloziert. An der Promotorregion bewirkt der modifizierte Cas9-Komplex eine ortsspezifische Genaktivierung. Das resultierende Expressionslevel kann durch das stabil integrierte Reporter-gen Luziferase gemessen werden. Erstellt mit *Biorender.com*. 17

8.3 Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die Promotionsordnung der Medizinischen Fakultät der Friedrich-Schiller-Universität bekannt ist,

ich die Dissertation selbst angefertigt habe und alle von mir benutzten Hilfsmittel, persönlichen Mitteilungen und Quellen in meiner Arbeit angegeben sind,

mich folgende Personen bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt haben:

Dr. Stefanie Reuter

Dr. Anika Westphal

Prof. Ralf Mrowka,

die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,

dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe, Teile jedoch in einer Abschlussarbeit der FoM-Linie des JENOS-Systems Verwendung fanden, welche im Rahmen des Curriculums des Universitätsklinikums Jena anzufertigen war, und

dass ich die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung nicht bei einer anderen Hochschule als Dissertation eingereicht habe.

Magdeburg, den